Article

User Behavior Adaptive AR Guidance for Wayfinding and Tasks Completion

Camille Truong-Allié [1],*, Alexis Paljic [1] , Alexis Roux [1] and Martin Herbeth [2]

1   Centre de Robotique, MINES ParisTech, Université PSL, 75006 Paris, France;
    alexis.paljic@mines-paristech.fr (A.P.); alexis.roux@mines-paristech.fr (A.R.)
2   Spectral TMS, 91000 Évry-Courcouronnes, France; martin.herbeth@spectraltms.com
*   Correspondence: camille.truong-allie@mines-paristech.fr

Abstract: Augmented reality (AR) is widely used to guide users when performing complex tasks, for example, in education or industry. Sometimes, these tasks are a succession of subtasks, possibly distant from each other. This can happen, for instance, in inspection operations, where AR devices can give instructions about subtasks to perform in several rooms. In this case, AR guidance is both needed to indicate where to head to perform the subtasks and to instruct the user about how to perform these subtasks. In this paper, we propose an approach based on user activity detection. An AR device displays the guidance for wayfinding when current user activity suggests it is needed. We designed the first prototype on a head-mounted display using a neural network for user activity detection and compared it with two other guidance temporality strategies, in terms of efficiency and user preferences. Our results show that the most efficient guidance temporality depends on user familiarity with the AR display. While our proposed guidance has not proven to be more efficient than the other two, our experiment hints toward several improvements of our prototype, which is a first step in the direction of efficient guidance for both wayfinding and complex task completion.

Keywords: augmented reality; wayfinding; multitask; large space; head-mounted display

1. Introduction

Augmented reality (AR) devices have the potential to help people learn or perform complex tasks by displaying virtual content spatially registered over real objects or points of interest (POIs). This AR guidance for complex task completion has been widely studied in the literature. For example, head-mounted display (HMD) device is a convenient method to keep users' hands free. The user can then perform tasks in the real world while following virtual instructions given by the device.

Sometimes, several spatially distant tasks are combined to form a sequence of operations in a large space. This is what happens when operators need to perform assembly, disassembly or maintenance of massive machines, such as plane motors, trucks, etc. This also happens when operators need to inspect large spaces, such as in powerplants. In these cases, users need both help from the AR device to realize their local tasks and guidance from the AR device to locate these tasks in space.

This leads to a new problem: how to combine AR guidance for complex tasks and AR guidance for wayfinding. In particular, we do not want the wayfinding guidance to disturb the user (for example, by occlusion or visual clutter) when he or she is performing a complex task or reading important information. In this paper, we choose to approach this problem by adding temporality to the wayfinding guidance. We use a neural network model to determine user behavior in real-time and we propose a strategy to determine when to display the wayfinding guidance depending on the current behavior. Our focus is not on task completion guidance; hence, our proposed wayfinding guidance can supplement any kind of AR-assisted task completion guidance.

First, we review existing guidance techniques for wayfinding and analyze whether they are adapted to our setting of interest—that is, adapted to the case when a task should be performed at each location indicated by the wayfinding guidance. We then propose a prototype for an adaptive guidance temporality based on user behavior; this behavior is inferred with a neural network model. Finally, we compare this adaptive guidance temporality to two other temporal strategies for the activation/deactivation of the wayfinding guidance.

2. Related Work

2.1. Guidance Format

Wayfinding is how people orient themselves in their environment and how they navigate through it. Wayfinding can be broken down into four stages: Orientation (how to determine one's location relative to the nearby objects and the target location), Route decision (selection of the route to take to arrive at the target location), Route monitoring (checking that the selected route is indeed heading towards the target location) and Destination recognition (when the target location is recognized) [1]. Orientation may be assimilated to the visual search of the target location—a visual scan of the environment aiming at the target identification.

2.1.1. Visual Guidance

Many 3D visual guidances have been proposed. For example, Bork et al. compared six different guidances [2]. The guidance 3D Arrows [3–5] points towards a POI. In AroundPlot [6], the POIs positions are mapped to the outside of a rectangle using 2D fisheye coordinates; the rectangle indicates user focus area. The guidance 3D Radar [2] displays on a plane a top view of the user and POIs, and a vertical arrow at the proxy positions of POIs indicates the POI height relative to user eye level. EyeSee360 [7] is composed of a rectangle in the middle of the screen representing the user field of view, and of a larger ellipse around this rectangle representing out-of-view space. POIs proxies are positioned relative to a horizontal line representing user eye level and a vertical line representing a null distance to the user. In SidebARs [8], two vertical bars at the right and left of user field of view are used to display proxies representing out-of-view POIs). Finally, Mirror Ball [2] proposes a reflective sphere displaying the distorted reflection of the POIs).

The existing guidances can roughly be grouped by design objectives. These groups are search time minimization (3D Arrow [3,4], 3D Radar, Attention Funnel, [9]), intuitivity and low workload (sidebARs, AroundPlot, [10], 3D Arrow [4,5]), user representation in 3D space of the virtual information (EyeSee360, 3D Arrow [4,5]) and occlusion management and visual clutter (AroundPlot, 3D Halo [11]).

When integrating these designs in a setting where the local tasks as well as the wayfinding task need to be AR-assisted, the criterion of interest is mostly occlusion management and visual cutter. We do not want the visual guidance to impede the local task completion or to hide important virtual information about the task. However, the work of Bork et al. (see Figure 1 to visualize the guidances) shows that guidances occluding the least user field of view (sidebARs, AroundPlot, 3D Arrow) have lower usability and speed scores compared with more imposing guidances (EyeSee360, 3D Radar) [2].

2.1.2. Non-Visual Cues

To limit visual clutter and cognitive load, researchers have investigated non-visual cues [12]. The main possibilities are auditory and vibrotactile cues, and combinations of the two. For example, HapticHead [13] uses a vibrotactile grid around the user's head to indicate the target direction by activating the corresponding vibrotactile actuators. Marcquart et al. used vibrotactile actuators to encode target relative longitude; depth is encoded with the vibration pulse and an audio feedback encodes user viewing angle and target relative elevation level with its pitch and volume [12].

When using vibrotactile cues, the less important visual clutter can come at the cost of a higher completion time and more head movements for the user, as shown in Haptic-Head [13]. The authors compare a vibrotactile grid around user head with the attention funnel visual guidance, a visual display of a curved funnel from user location to his or her target. In addition, the vibrotactile guidance can only provide focus on one target at a time, which can be an important limitation.
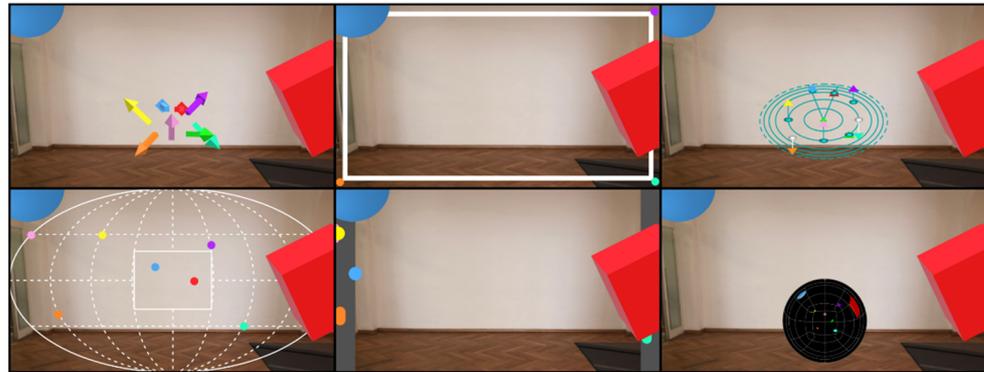

Figure 1. The 6 visual guidances compared in the work of Bork et al. Guidances 3D Arrows (top-left), AroundPlot (top-center), 3D Radar (top-right), EyeSee360 (bottom-left), sidebARs (bottom-center) and Mirror Ball (bottom-right). Reprinted with permission from ref. [2]. Copyright 2018 IEEE.

When using auditory cues, the less important visual clutter comes at the cost of a higher completion time [12] and a possible overload of the auditory channel. This may be a problem, depending on the context in which the wayfinding guidance is used. In industrial settings for instance, where sequence of tasks in large spaces are likely to happen, the auditory channel may already be overloaded, or even unusable because of noise protection. For example, noise protection is a common requirement in the three sites described by Lorenz et al. [14].

2.1.3. Guidance in a Multitask Setting

There exist some works on guidance techniques combining wayfinding task with other tasks (which do not require wayfinding guidance). However, in these works, the other tasks are only tools to evaluate the guidance techniques performances and do not constitute a meaningful objective by themselves.

They are used to measure situational awareness [12], visual working memory [15], or to analyze how specific guidances perform when workload increases [16]. In these works, the visual search task, where the wayfinding guidance technique is needed, is the main task. None of these works consider our problem, where the visual search can be seen as an auxiliary task and where wayfinding is only needed to guide the user from one main task location to another. Additionally, as seen in Sections 2.1.1 and 2.1.2, new challenges arise from this problem: occlusion or efficiency when performing local tasks for visual guidance, efficiency for non-visual guidance.

2.2. Guidance Adaptivity
2.2.1. Adaptivity Justification

Section 2.1 highlights that the minimization of visual clutter and occlusion comes at the cost of efficiency (for visual cues and non-visual cues) and reduce the application possibilities (for non-visual cues) of the guidance.

However, the visual clutter and occlusion minimization is wanted mostly during the real-world task completion phase; during the wayfinding task phase, visual clutter and occlusion are less important and other criteria may be more relevant for the choice of the guidance technique. For example, when navigating in hazardous environments, guidance designs facilitating situational awareness (for example, 3D Halo, or auditory cues) may be

preferred, while guidance designs which have been proven to be efficient in terms of speed (as EyeSee360) may be used in safer environments.

Because the wayfinding and the local tasks do not happen at the same time, AR assistance for local task completion and AR guidance for wayfinding can be decoupled. If done, the guidance design for wayfinding will not have to handle both visual clutter minimization and any other criterion (situational awareness, user speed . . .) relevant for the current context.

Separating local tasks and wayfinding can be performed in several ways. The device can adapt to the user location, for instance by displaying the visual guidance depending on how far the user is from his or her target. The device can also adapt to the user behavior: display the visual guidance only when the user is performing some specific tasks, or stop displaying the visual guidance when he or she has seen the target location.

Many context-aware approaches have been proposed to adapt the AR display to the context at hand. For example, the environment luminosity can be considered to decide the most appropriate virtual text color [17,18]. AR content can be filtered depending on user expertise [17,19] or location [20], etc. To the best of our knowledge, none of the existing context-aware approaches consider wayfinding guidance. In the following sections, we will review two context-aware approaches that seem relevant for the combination of AR-assisted local tasks completion and wayfinding guidance.

2.2.2. Task-Based Adaptivity

Task-related content is extremely common in AR. For example, many industrial AR devices propose to guide users in a step-by-step manner, by displaying information relative to the task at hand. Often, the progression in the different steps requires conscious commands from the user. This can be carried out with voice command [21–24], a tap on the headset [22,25], a click on connected devices [26], a gestural command [21] or click/selection on virtual menu with hands or gaze [27–29].

Only a few works automatically adapt the HMD display based on user activity. Tsai and Huang propose a device detecting the user behavior when he or she is walking around in an exploration task [30]. These behaviors are generic ("Stationary and Staring (SS), Standing Still but Turning Around (ST), Looking Around While Moving (LM), Staring While Moving (SM), and Turning Around (TA)" [30]), and associated with a type of information display (Typical mode, Stationary Focus mode, Aggregation mode, Moving Focus mode and Transparent mode). These modes give information about all the POIs in the area or about the POI the user is currently looking at. User behaviors are detected using the Perceptive GPS (PGPS) algorithm based on a Hidden Markov Model [31].

In this work, the user activity is the exploration of large space. The user does not interact with the environment or with the AR headset, as it is the case in our setting of interest—AR guidance for complex task completion and AR guidance for wayfinding. Therefore, in our case, we would need to detect other user behaviors (see Section 3.2) than the ones detected in Tsai and Huang work, and to display different types of information (a visual guidance, or no visual guidance).

2.3. Summary

Many AR guidances for wayfinding have been proposed. Visual guidances raise the issues of visual clutter and occlusion, while non-visual guidances raise the issue of efficiency in terms of speed. Depending on the context in which wayfinding guidance is needed, these issues can be critical. To limit visual clutter and occlusion, a solution would be to display the wayfinding guidance only when necessary.

The need of wayfinding depends on the current context the user is in: for example, distance to the target task location, user current behavior. The two context-aware approaches we found the most suitable to our setting are spatial context and task-based context. To the best of our knowledge, no context-aware device has taken interest in the need of wayfinding.

3. Prototype

3.1. Design Justification

We seek to propose a prototype combining visual guidance for navigation between tasks locations (wayfinding) and AR assistance for local tasks completion. In this work, we focus on the wayfinding assistance display and use an industrial AR software solution [29] for the local task completion assistance.

From the literature, three different approaches emerge regarding when to display visual guidance: a manual approach, where the user indicates to the device whether he or she wants the guidance to be displayed or not; a spatial approach, where the guidance appearance depends on the user location related to the target; and a task-based approach, depending on what the user is doing.

The manual approach is the simplest one, but has many drawbacks: operators are likely to perform tasks requiring both hands—for example, carrying objects and performing gestures, making click and menu navigation commands ill-adapted. Voice command is also not suited for noisy environments [28]. Click on gaze is currently limited because of tracking issues [32] and requires a virtual button which permanently occludes user field of view.

A spatial approach could consist of displaying the guidance only when the user is far from the target location. However, this approach does not cover many possible scenarios. For instance, a user can be looking at a machine from a distant point of view to detect possible defects, and then come closer to repair it. Visual guidance is likely to disturb him or her in the defect detection task. Another scenario where a spatial approach is not fitted is a scenario in hazardous environments. In this case, on his or her way to a local task, the user may need to also be looking at information display (real or virtual) about the environment— for example, safety warnings. As they are not task-related, these warnings may appear far from the target location, where wayfinding guidance is activated. The guidance could then distract the user from the warnings, which could be potentially dangerous.

This is why the task-based approach seems to be the most generic one. We could take inspiration from the work of Tsai and Huang and associate specific user states to relevant virtual display content. In our prototype, we propose a new range of possible user behaviors and associate them with two possible displays: with wayfinding guidance and without wayfinding guidance. In the following sections, we will describe how the user behaviors to be detected were chosen and how we associate them with the need (or not) of wayfinding guidance.

3.2. User States Detection

User behavior classification was made based on the taxonomy of user tasks in an industrial context from Dunston et al. [33] (see Figure 2). They classified industrial tasks into five categories. Application domain is a promising area where mixed reality (MR) technology could be applied. Application-specific operation is a more specific operation within the application domain. Operation-specific activity is a specific unit of work; multiple operation-specific activities form an application-specific operation. Composite task (such as measure, connect, align, etc.) is a fundamental building block of an operation-specific activity; a series of composite tasks form an operation-specific activity. Finally, primitive task "refer[s] to elemental motion such as reaching, grasping, moving, eye travel, etc." [33] (p. 437).

According to Dunston et al., the composite and primitive tasks levels are the two levels relevant for AR applications. Since task recognition algorithms at the level of primitive tasks are highly domain-specific (every single scenario needs specific data), in our prototype, the granularity of user behavior classification is at the level of the composite task. The above-mentioned taxonomy does not consider the user interaction with the augmented reality device, which is yet part of the industrial activity. We therefore made the choice to include interactions with the AR device to the composite tasks category.
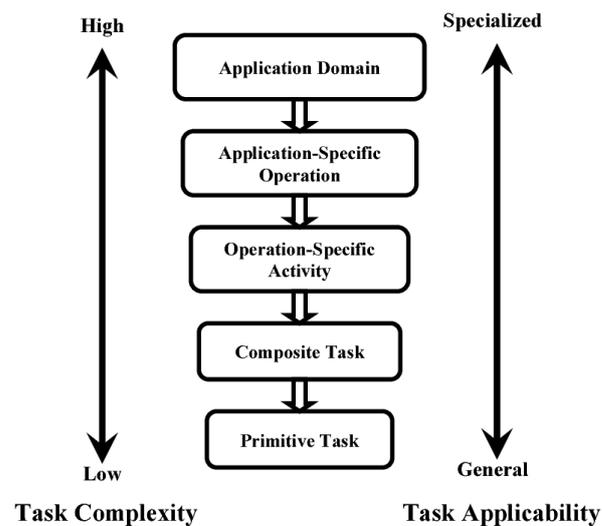
Figure 2. Taxonomy of tasks in an industrial context. Reprinted from [33].

We chose the resulting tasks for our prototype design:

- *Local task*: in an industrial context, this is any composite task (outside of *walk* and *visual search*) which would have been performed by the operator without the display.
- *Compare real with virtual*: in practice, this is the same as *local task* but with the user frequently rotating his or her head to look at a virtual content when realizing the *local task*. They are separated in two different categories since they imply different user behaviors.
- *Walk*: this happens when the user goes from one point of interest to another, for example.
- *Menu interaction*: these are the operator's direct interactions with the augmented reality device, for example, when he or she needs to indicate that he or she has finished a step and needs information to perform the next step, or when he or she needs to move a virtual object to see it better.
- *Virtual information assimilation*: this is the state the user is in when visualizing virtual content displayed by the device (i.e., to learn what to perform in the next step) without realizing any other task.
- *Real-world information assimilation*: this is the state the user is in when visualizing real-world content (i.e., a component, a security panel, etc.) without realizing any other task.
- *Visual search*: this is the task a user performs while looking for specific object or information, real or virtual. In our setting, the visual search target is mostly the next local task location, but it can also be any other object. These tasks were chosen after a careful observation of users performing several AR-assisted local tasks in large space. We made sure that these tasks were associated with specific user behaviors: we could know the task the user was performing by looking at his or her behavior. For example, we could know the difference between *menu interaction* and *virtual information assimilation* by the use of the hands in the first case, but not in the latter. We set up a neural network model to detect those tasks. See Section 3.3 for its implementation.

3.3. Visual Guidance Appearance

The chosen visual guidance is a blinking 3D arrow. We choose it because of its simplicity and intuitivity for users.

User behaviors for which guidance is useful were identified as *walk* (when walking, users may need information about their surroundings and where they should be going) and *visual search*. This identification was made from a careful observation of people using an AR headset to realize local tasks in large space. When these two behaviors are detected by the neural network, guidance is displayed, and hidden otherwise.

To smooth out high frequency changes in the guidance apparition, instead of simply appearing when needed and disappearing when not needed, the visual guidance had its opacity increased (when guidance is needed) and decreased (when guidance is not needed) at each timeframe. The guidance takes about 1 s to fully appear or disappear.

3.4. Materials

3.4.1. User Behavior Inference Algorithm

The neural network used to infer user behaviors is the the open-source network proposed by Fawaz et al. [34]. The input data are time series data tracked from the Hololens2 HMD [35]. It consists of the head linear and angular velocity, the hands velocity, the angular gaze velocity, if the user is looking at a virtual object, if the user is pointing at a virtual object and if the user is clicking at a virtual object. These data were acquired from twelve people and in three different environments. These environments were: a large room with obstacles limiting user visibility and walk, two neighboring large rooms with no obstacles and three neighboring large rooms. The data recordings were from 3 to 10 min long. After acquisition, the data were manually labeled frame by frame using a first-person view video, which was synchronized with the data acquisition (every 0.2 s). The labels corresponded to the behaviors defined in Section 3.2. The neural network was trained during 1500 epochs with a learning rate of 0.001. Its input format was a sequence of 5 consecutive data points.

Verification heuristics were used to correct possible neural network errors. The prediction walk was verified by a speed threshold, and the compare real with virtual task and virtual information assimilation task predictions were disregarded if the user was not looking at virtual content within the 5 past input data points. When a prediction is disregarded, the last valid prediction is returned.

3.4.2. AR Device and Display

The HMD used in this study is a Hololens2 [35], and the software application used to display the virtual objects is adapted from Spectral TMS software [29].

This solution enables the user to switch from one task instruction to the other by clicking on virtual buttons located at these task locations. See Figure 3. for a user view of the application and of the guidance display.
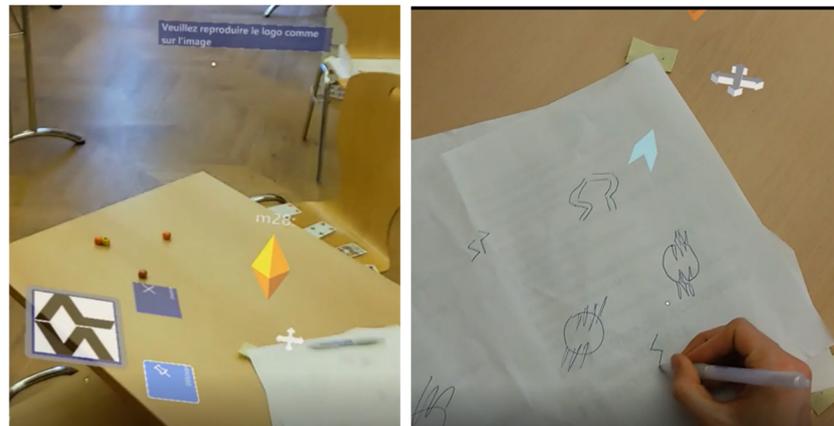

Figure 3. User view of the task at hand. The left image shows the task setup, with the instruction and explicative media displayed at the task location which is indicated by an orange gem. The right image shows the user performing the task with the permanent 3D arrow blinking in his or her field of view.

3.4.3. Real-Time Inference

The user behavior inference algorithm ran on a computer GPU. The input data of the neural network and the opacity to display were sent from the Hololens2 to the computer and conversely via a UDP protocol at a frequency of 5 Hz. This low frequency enables the

input to be sent to the computer, the neural network to compute the output, and finally the output to be sent back to the headset. In a more developed version (for example, using cloud processing), a higher frequency could be used.

4. User Study

4.1. Hypotheses

Our user study setting mimics an industrial scenario assisted with an AR device. In this setting, a scenario consists of a series of steps associated with a task to perform (i.e., remove the top cover of a machine) at a specific location (i.e., the machine emplacement). Users are helped by virtual content when performing the task (the description of the task is displayed in the headset, and virtual medias can be displayed near the location of the task), and wayfinding to the task location is assisted with a visual guidance.

In this work, we compare an adaptive guidance (our prototype) with two other guidances: a minimal guidance and a permanent guidance.

- The adaptive guidance appears according to the algorithm described in Sections 3.2 and 3.3.
- The minimal guidance appears when the next step of a task begins and disappears when the target location (where the real-world task should be performed) is in the field of view of the user.
- The permanent guidance is always displayed.

We wanted to verify the following hypotheses:

Hypothesis 1 (H1). *Participants are more efficient (in terms of speed) with the adaptive guidance than with the minimal guidance.*

Hypothesis 2 (H2). *Participants are more efficient (in terms of speed) with the permanent guidance than with the adaptive guidance.*

Hypothesis 3 (H3). *Participants find the adaptive guidance more comfortable than the permanent guidance.*

4.2. Methods

4.2.1. Subjects

The experiment was conducted on 28 persons, involving students and teachers, from Mines ParisTech and developers and commercials from the Spectral TMS entreprise. Participants were aged from 20 to 58 years old, with 21 men and 7 women. The participants overall felt quite familiar with augmented reality (familiarity was rated 3.0/5 in average), and extremely familiar with technological devices (familiarity was rated 4.5/5 in average). Participants were classified as either "expert" or "novice" depending on the grade they gave and their professional background. All participants calibrated the headset to their view. They all gave their consent to have their data recorded anonymously and to take part in the experiment. They were given the choice to stop the experiment at any point without having their data registered or to refuse to take part in the experiment.

No particular risk linked to the AR experiment was noticed. The headset was disinfected after each participant, the room was aerated and both participants and supervisors wore masks during the experiment.

4.2.2. Task and Procedure

We employed a within-subjects design with the independent variable being the guidance type to examine the effect of guidance temporality on participants' performances and comfort.

Performance was measured with task completion time, and comfort was evaluated by grades given by the participants on a visual analog scale. These grades evaluated: the general feeling about the guidance, the feeling of comfort towards the guidance (did the

participant accept the visual guidance without feeling disturbed?), the feeling of efficiency (did the participant feel that he or she was faster with the help?) and the feeling of relevance towards the presence or the absence of guidance (did the participant need the help when it was not displayed and was the help displayed when needed?). These times and grades are the dependent variables to measure.

During the evaluation phase of the experiment, participants were asked to perform three scenarios, each associated with a different guidance. A scenario consisted of 10 steps to be executed in a precise order. At each step, a location was given to the participant, materialized by a flying orange gem. Half of the locations were associated with a task to perform in the real world, and half of them were associated with no task, and the participant were only asked to move towards the gem. The possible tasks were for example sorting cards by color or drawing a specific logo displayed through the AR headset. The locations were randomly chosen in advance, so that for each scenario, between two consecutive target locations, there is a random distance of 2, 3 or 5 m and the angle between them is a random multiple of $\pi/4$. There were a total of 10 possible scenarios. Every participant had the headset calibrated to his or her view and the curtains were closed, ensuring that all participants were in similar settings regarding the visualization of the virtual objects.

For each participant, the three scenarios were randomly chosen between the 10 possible scenarios. The order of the guidances shown to the participant was also random to minimize the impact of the learning effect on the speed performance for each type of guidance.

After each scenario of the evaluation, the participant was asked for his or her feedback on the guidance by rating the four above-mentioned criteria (general feeling, feeling of comfort, feeling of efficiency, feeling of relevance) and to freely add a comment. Then, at the end of the whole evaluation part, the participant was given the opportunity to choose his or her favorite guidance type for each criterion and to leave general comments. To limit any bias from the participants, when explaining the experiment, participants were only told that they were going to take part in a 30-min-long experiment involving an AR headset, and that this experiment aimed at evaluating different guidance temporality types. They did not have more details about the three guidance behaviors.

Before the evaluation phase was a training phase in which the participants completed a scenario without any guidance. The aim was to help the participant learn how the application works and get comfortable with the clicking mechanics required by the AR device. This would minimize any learning effect between the three scenarios of the evaluation phase.

4.2.3. Results Analysis

We realized a within-subject study with the three types of guidance as independent parameters, and the scenario total completion time and participants rating as dependent variables. For every dependent parameter, a 1×3 repeated measures ANOVA was performed when normality assumption was verified by a Shapiro test, and Friedman test was used as non-parametric analysis otherwise. Pairwise comparisons between guidances were further made if the ANOVA or Friedman test indicated significant differences between the guidances. The Wilcoxon signed-rank test was conducted after the ANOVA test, and the Nemenyi test was conducted after the Friedman test.

5. Results

5.1. User Behavior Inference Results

During the experiment, the resulting balanced accuracy (from the formula described in the review of Grandini et al. [36]) was 67%. From the confusion matrix (Table 1), one can see that some misclassifications are likely to happen due to unbalanced data during training (see, for example, the high rate of visual search samples which are classified as local task). Overfitting over the training data is another possible reason for misclassification: training and validating sets were quite similar, which prevented us from detecting it beforehand.

Finally, smooth transitions occurred between user behaviors; however, only one behavior was captured by timestamp. This may lead to small uncertainty when users are in between two behaviors (for example, from visual search to walk).

Table 1. Confusion matrix of the user behavior inference algorithm with the data obtained from the experiment of Section 4. LT refers to *local task*, CRV to *compare real with virtual*, W to *walk*, MI to *menu interaction*, RIA to *real-world information assimilation*, VIA to *virtual information assimilation*, VS to *visual search*. Values highlighted in green are the values where the decision to not display the guidance is well computed, values highlighted in blue are the values where the decision to display the guidance is well computed, and values highlighted in orange are the values where the decision to display or not the guidance is wrong.

		Computed Classes							
		LT	CRV	W	MI	RIA	VIA	VS	Total
	LT	1378	512	3	26	62	242	2	2225
	CRV	588	1216	2	8	13	2	221	2050
	W	303	274	2301	12	19	0	30	2939
Actual	MI	381	325	1	2036	5	75	1	2824
classes	RIA	281	103	4	25	457	1	1	872
	VIA	127	19	1	18	1	585	24	775
	VS	201	112	20	2	2	19	361	717
	Total	3259	2561	2561	2332	559	924	640	12402

A total of 7 classes are used in the user behavior inference algorithm. However, only 2 states are output: guidance display (which triggers guidance opacity increase) and no guidance display (which triggers guidance opacity decrease). When computing this confusion matrix (Table 2), a balanced accuracy of 86% is obtained. The guidance display state accuracy is 74%, while the no guidance display state accuracy is 97%.

Table 2. Confusion matrix of the guidance display decision algorithm. NGD means *no guidance display* and GD means *guidance display*.

		Computed Classes		
		NGD	GD	Total
	NGD	8486	260	8746
Actual classes	GD	968	2712	3680
	Total	9454	2972	12402

5.2. Efficiency

Efficiency was measured in terms of completion time. We measured the time users took to visualize the target location, the time they took to reach it and the time they took to perform the whole scenario. The Friedman p-values of these different metrics for the three guidances are reported in Table 3.

Contrary to our assumptions,

Result 1(R1): *Participants were significantly faster to perform the whole scenario with minimal guidance than with adaptive guidance (p-value of 0.036 from Table 4).*

Result 1bis (R1bis): *Novices were also significantly faster to perform the whole scenario with minimal guidance than with adaptive guidance (p-value of 0.002 from Table 4).*

Result 2 (R2): *Novices were significantly faster to perform the whole scenario with minimal guidance than with permanent guidance (p-value of 0.036 from Table 5).*

Result 3 (R3): *However, experts were, on the contrary, significantly slower to perform the whole scenario with minimal guidance than with permanent guidance (p-value of 0.036 from Table 4).*

Table 3. Friedman *p*-values of the different efficiency metrics comparison between guidances.

	Novices	Experts	Total
Time to visualize target location	0.116	0.050	0.023
Time to reach target location	0.735	0.174	0.039
Whole scenario duration	0.002	0.039	0.035

Table 4. *p*-values of pairwise comparisons using the Nemenyi post hoc test for the time participants took to perform the whole scenario.

	Minimal			Adaptive			Permanent		
	Novices	Experts	Total	Novices	Experts	Total	Novices	Experts	Total
Minimal	1.000	1.000	1.000	0.002	0.740	0.027	0.0291	0.036	0.555
Adaptive	0.002	0.740	0.027	1.000	1.000	1.000	0.695	0.181	0.271
Permanent	0.0291	0.036	0.555	0.695	0.181	0.271	1.000	1.000	1.000

Table 5. Pairwise comparisons using the Nemenyi post hoc test for the time participants took to visualize the target location.

	Minimal			Adaptive			Permanent		
	Novices	Experts	Total	Novices	Experts	Total	Novices	Experts	Total
Minimal	1.000	1.000	1.000	0.900	0.739	0.357	0.260	0.036	0.357
Adaptive	0.900	0.739	0.357	1.000	1.000	1.000	0.122	0.181	0.017
Permanent	0.256	0.036	0.357	0.122	0.181	0.017	1.000	1.000	1.000

R1 and R1bis contradict H1, which is therefore rejected.

R2 and R3 (see Figure 4) may suggest that novices performances are hindered by the permanent guidance when it is displayed when not needed, when, on the contrary, experts perform better with the permanent guidance because it is always there when they need it, and they are able ignore it when they do not need it.
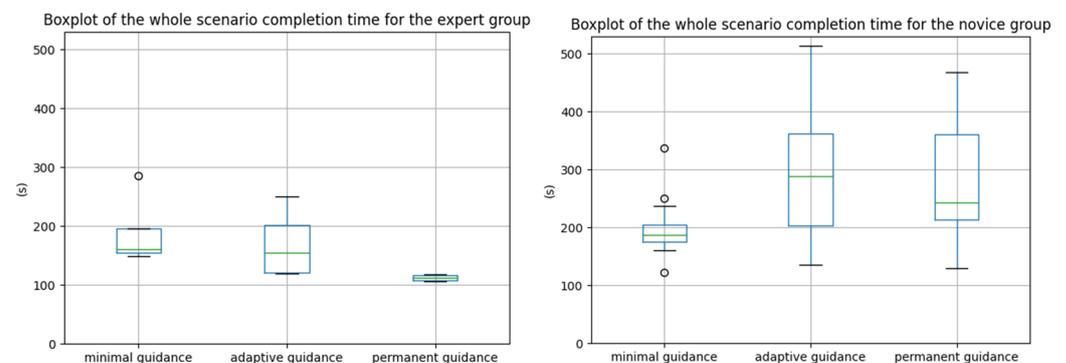

Figure 4. Boxplots of task completion by guidance type. On the **left** are the boxplots of experts' task completion time, and on the **right** are the boxplots of novices' task completion time. While experts are significantly faster with permanent guidance, novices are significantly faster with minimal guidance. To explain why H1 was rejected, we looked into the time participants took to visualize their target. We noticed that participants were significantly slower to notice the target with the adaptive guidance than with the permanent guidance (see Figure 5 and Table 5). The reason behind this was extremely noticeable during the experiments: participants tended to wait for the adaptive guidance to appear.
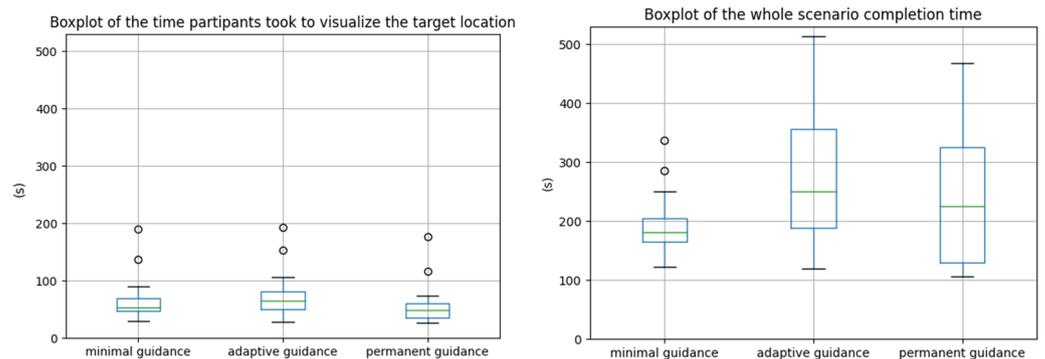
Figure 5. Boxplots by guidance type. On the **left** is the boxplot of the time participants took to visualize the target location and on the **right** is the boxplot of the time participants took to complete the whole scenario. Participants were significantly slower to find the target location with the adaptive guidance than with the permanent guidance. Participants were significantly faster to complete the whole task with the minimal guidance than with the adaptive and permanent guidances. H2 is not accepted nor rejected, and the boxplot of task duration in Figure 5 shows similar task duration between the adaptive and permanent guidances. This suggests that participants wasted time (compared to the minimal guidance) in an equivalent manner for these two guidances. A hypothesis is that the time wasted when participants were waiting for the adaptive guidance is compensated by faster local tasks completion compared with the permanent guidance. This faster task completion may be because they were not disturbed by the wayfinding guidance with the adaptive guidance when performing local tasks, contrary to the permanent guidance. We cannot verify that participants were indeed faster when performing tasks with the adaptive guidance or with the minimal guidance compared to the permanent guidance, because the variety of tasks to perform was high, and the different tasks needed different completion time. This made time completion between different tasks impossible to compare. A future work would be to ensure similar local task difficulties for the three guidance temporalities evaluation.

5.3. Users Preferences

To our surprise, no clear preference was set for any of the guidances. H3 is then neither accepted nor rejected. This means that the three guidances have advantages and drawbacks for all the participants (see Table 6 for the details of participants' grades).

Table 6. Participants grades for the four criteria described in Section 4.2.2. Task and procedure.

Guidance Type	Comfort Grade		Efficiency Grade		Relevance Grade		General Feeling Grade	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Minimal	7.15	1.72	6.63	1.90	5.96	2.40	7.04	1.57
Adaptive	5.98	1.86	7.37	1.88	6.16	2.36	7.38	1.59
Permanent	7.12	1.70	7.85	1.66	6.78	2.30	8.00	1.31

However, two other limitations due to the experiment setting can explain the lack of difference between the participants' grades. Firstly, the 3D Arrow does not occlude much user field of view, and therefore the differences between the three guidances were not as noticeable as they could have been with other guidances. Some participants mentioned that they did not really notice any difference between the three guidances. Secondly, a learning effect seems to have appeared. Despite a long training scenario when participants learned how to use the AR headset, novices were performing better with time with the Hololens2 commands and getting accustomed to how the 3D Arrow behaves. Therefore, they felt more efficient and generally preferred the last guidance they tried because they felt more at ease. This effect is partly smoothed with the randomness of the guidance order, but it increases variability and may have participated to the lack of general trend for participants preferences.

Some general observations can still be made from participants comments on the different guidances.

When participants preferred the minimal guidance, they said that it disappeared at the right moment to enable them to visualize the target location. However, some participants complained that the minimal guidance disappeared too early or did not appear at all. The latter happened when the gem indicating the target location was already in their field of view, but far from them, or at their field of view periphery.

Participants generally said that the adaptive guidance appeared and disappeared globally at the right moments, but the few times when it did not appear or disappear when needed was enough to disturb them and make them feel uncomfortable. An improvement of the user behavior detection may fix this issue.

Some participants said that the permanent guidance was the smoothest to follow. However, some of them complained that it was tiring to always have to obey to it.

A common complaint for the adaptive and permanent guidances is that participants tend to focus on the guidance for too long instead of focusing on their surroundings: some participants noticed that they would have seen the target location earlier if they were not too concentrated on following the visual guidance indications.

6. Discussion

6.1. User Behavior Inference Algorithm Improvement

Improving the user behavior inference algorithm accuracy can be carried out by using more training data and by adapting the neural network architecture so that it can output several labels for a single output. This will be performed in a future work by using the data collected during the current experiment. Achieving a multi-label prediction can be performed by replacing the initial softmax activation function of the output layer by a sigmoid activation function and using a binary-crossentropy loss.

The improvement of the user behavior inference algorithm accuracy would of course consequently improve the guidance display decision algorithm, in particular if more effort is put in the *visual search* behavior detection. However, other measures could also be taken, as discussed in Section 6.3.

6.2. User Preferences

The results of this experiment show that the three proposed guidances have advantages and drawbacks and no conclusion on participants preferences could be made. On average, participants were the fastest with the minimal guidance; however, in some cases they were lost for a long time looking for the target location after the minimal guidance disappearance, causing them to dislike this guidance. It is also noticeable that experts' and novices' highest efficiency is not obtained with the same guidance.

The fact that experts were faster with the permanent guidance while the novices are faster with the minimal guidance can be explained by the following hypotheses: Experts are able to ignore the visual guidance when they find it useless and therefore, they can look at the guidance direction rather than at the guidance itself, making them find the target location easily. Novices tend to focus on the virtual guidance when displayed, preventing them from looking around in the direction pointed by the guidance and finding it. These hypotheses would require further gaze tracking analysis for verification. This would imply that: 1. Users would prefer a path or global direction to indicate the target rather than step-by-step instructions preventing them from looking around. 2. Because the too-long display time of 3D Arrow slows down users, it is more complex to define when guidance is needed and when it is not needed. 3. The definition of the moments when the guidance is useful may depend on the guidance ability to let people concentrate on other parts of their environment (that is, on people's situational awareness while following the guidance).

6.3. Prototype Improvement

A future work would test once again an adaptive guidance based on user behavior, but with a guidance design more focused on situational awareness.

This adaptive guidance could also be further improved by adding information in addition to current user behavior. Regarding the guidance appearance, we chose to adapt the guidance to what the user is doing without taking in consideration the fact that the user would adapt to the guidance in return. That is, once he or she notices that a guidance will help him or her navigate, the user may tend to wait for the guidance to appear by itself—and then, to not understand why the guidance fails to appear. To consider user adaptivity to the guidance and to integrate it into the guidance temporality design, we could indicate to the user that the guidance is always here, at his or her disposal, only waiting for the user to show his or her need of a guidance. This could be carried out by always displaying the guidance, but with an extremely low opacity value, for example. Regarding the guidance disappearance, we did not take user knowledge into account; in particular, user detection of the target location was ignored. Thus, when the user has found the target location, the guidance did not disappear, and possibly disturbed his or her navigation in the environment. A more refined design of an adaptive guidance could take this knowledge into account.

7. Conclusions

To the best of our knowledge, the present work is the first to explore how to guide users when they perform multiple AR real-world tasks in large space. For this exploration, several approaches are possible. For example, a guidance for wayfinding could be carefully designed so that it does not disturb the user when he or she is performing real-world tasks. Another approach would be to determine when the user needs guidance to navigate in the environment, and when the user need guidance to perform a task in the real world.

This work is a first step in this direction. Making the assumptions that users may need guidance only when walking or performing a visual search, and that the guidance display design does not influence when the guidance is needed, we realized a prototype of a user behavior-based adaptive guidance, with a 3D Arrow as guidance display. We compared this adaptive guidance to two other guidance appearance/disappearance temporalities. This showed us the potential of an adaptive guidance while highlighting the imperfections of our first prototype. In particular, the two above-mentioned assumptions seem to be inaccurate. First, user behavior alone is not enough to determine whether a guidance is needed or not; a more refined adaptive guidance could include user adaptivity to the guidance and user knowledge in its guidance appearance/disappearance decision algorithm. Second, the design of the guidance display seems to have an influence on whether the guidance is needed or not: for example, a visual guidance providing step-by-step instructions may force the user to focus on it all the time and prevent him or her from focusing on areas outside of the guidance.

An interesting future work would then be to refine the presently proposed adaptive guidance and to couple it to a visual display adapted to user perception of the environment. We would like to explore two propositions: 1. Trajectory (for example, a path from the user to the target location) against step-by-step direction instruction (for example, 3D arrow), and 2. Obstacle awareness facilitation (for example, by highlighting obstacles while indicating the target location). In this future work, it is planned to test the prototype with a larger panel of users' backgrounds; in particular, it would be interesting to know user preferences among people not particularly familiar with AR.

Author Contributions: Conceptualization, C.T.-A. and M.H.; Formal analysis, C.T.-A.; Investigation, C.T.-A.; Methodology, C.T.-A. and A.P.; Project administration, A.P.; Software, C.T.-A.; Supervision, A.P. and M.H.; Writing—original draft, C.T.-A.; Writing—review and editing, A.P. and A.R. All authors have read and agreed to the published version of the manuscript.

References

1. Guastello, S.J. *Human Factors Engineering and Ergonomics*; Taylor & Francis Group: Boca Raton, FL, USA, 2014.
2. Bork, F.; Schnelzer, C.; Eck, U.; Navab, N. Towards efficient visual guidance in limited field-of-view head-mounted displays. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 2983–2992. [CrossRef] [PubMed]
3. Burigat, S.; Chittaro, L. Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids. *Int. J. Hum.-Comput. Stud.* **2007**, *65*, 945–958. [CrossRef]
4. Tonnis, M.; Klinker, G. Effective control of a car driver's attention for visual and acoustic guidance towards the direction of imminent dangers. In Proceedings of the 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality, Santa Barbara, CA, USA, 22–25 October 2006; IEEE: Barbara, CA, USA, 2006; pp. 13–22.
5. Schinke, T.; Henze, N.; Boll, S. Visualization of off-screen objects in mobile augmented reality. In Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, Lisbon, Portugal, 7–10 September 2010; pp. 313–316.
6. Jo, H.; Hwang, S.; Park, H.; Ryu, J.H. Aroundplot: Focus+ context interface for off-screen objects in 3D environments. *Comput. Graph.* **2011**, *35*, 841–853. [CrossRef]
7. Gruenefeld, U.; Ennenga, D.; Ali, A.E.; Heuten, W.; Boll, S. Eyesee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. In Proceedings of the 5th Symposium on Spatial User Interaction, Brighton, UK, 16–17 October 2017; pp. 109–118.
8. Siu, T.; Herskovic, V. SidebARs: Improving awareness of off-screen elements in mobile augmented reality. In Proceedings of the 2013 Chilean Conference on Human-Computer Interaction, Temuco, Chile, 11–15 November 2013; pp. 36–41.
9. Biocca, F.; Tang, A.; Owen, C.; Xiao, F. Attention funnel: Omnidirectional 3D cursor for mobile augmented reality platforms. In Proceedings of the SIGCHI Conference on Human Factors in computing systems, Montréal, QC, Canada, 22–27 April 2006; pp. 1115–1122.
10. Matsuzoe, S.; Jiang, S.; Ueki, M.; Okabayashi, K. Intuitive visualization method for locating off-screen objects inspired by motion perception in peripheral vision. In Proceedings of the 8th Augmented Human International Conference, Silicon Valley, CA, USA, 16–18 March 2017; pp. 1–4.
11. Trapp, M.; Schneider, L.; Lehmann, C.; Holz, N.; Dllner, J. Strategies for visualising 3D points-of-interest on mobile devices. *J. Locat. Based Serv.* **2011**, *5*, 79–99. [CrossRef]
12. Marquardt, A.; Trepkowski, C.; Eibich, T.D.; Maiero, J.; Kruijff, E.; Schöning, J. Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 3389–3401. [CrossRef] [PubMed]
13. Kaul, O.B.; Rohs, M. Haptichead: A spherical vibrotactile grid around the head for 3d guidance in virtual and augmented reality. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 3729–3740.
14. Lorenz, M.; Knopp, S.; Klimant, P. Industrial augmented reality: Requirements for an augmented reality maintenance worker support system. In Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Munich, Germany, 16–20 October 2018; IEEE: Munich, Germany, 2018; pp. 151–153.
15. Woodman, G.F.; Vogel, E.K.; Luck, S.J. Visual search remains efficient when visual working memory is full. *Psychol. Sci.* **2001**, *12*, 219–224. [CrossRef] [PubMed]
16. Hopkins, K.; Kass, S.J.; Blalock, L.D.; Brill, J.C. Effectiveness of auditory and tactile crossmodal cues in a dual-task visual and auditory scenario. *Ergonomics* **2017**, *60*, 692–700. [CrossRef] [PubMed]
17. Gattullo, M.; Dalena, V.; Evangelista, A.; Uva, A.E.; Fiorentino, M.; Boccaccio, A.; Gabbard, J.L. A context-aware technical information manager for presentation in augmented reality. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; IEEE: Osaka, Japan, 2019; pp. 939–940.
18. Gabbard, J.L.; Swan, J.E.; Hix, D. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence* **2006**, *15*, 16–32. [CrossRef]
19. Mourtzis, D.; Xanthi, F.; Zogopoulos, V. An adaptive framework for augmented reality instructions considering workforce skill. *Procedia CIRP* **2019**, *81*, 363–368. [CrossRef]

20. Anandapadmanaban, E.; Tannady, J.; Norheim, J.; Newman, D.; Hoffman, J. Holo-SEXTANT: An augmented reality planetary EVA navigation interface. In Proceedings of the 48th International Conference on Environmental Systems, Albuquerque, NM, USA, 8–12 July 2018.

21. Schlagowski, R.; Merkel, L.; Meitinger, C. Design of an assistant system for industrial maintenance tasks and implementation of a prototype using augmented reality. In Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 10–13 December 2017; IEEE: Singapore, 2017; pp. 294–298.

22. Ramakrishna, P.; Hassan, E.; Hebbalaguppe, R.; Sharma, M.; Gupta, G.; Vig, L.; Shroff, G. An ar inspection framework: Feasibility study with multiple ar devices. In Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Merida, Mexico, 19–23 September 2016; IEEE: Merida, Mexico, 2016; pp. 221–226.

23. Upskill. Available online: https://upskill.io/ (accessed on 7 October 2021).

24. Taqtile. Available online: https://taqtile.com/ (accessed on 7 October 2021).

25. Mira. Available online: https://www.mirareality.com/ (accessed on 7 October 2021).

26. Muhammad, A.S.; Chandran, K.; Albuquerque, G.; Steinicke, F.; Gerndt, A. A Suggestion-Based Interaction System for Spacecraft Design in Augmented Reality. In Proceedings of the 2021 IEEE Aerospace Conference (50100), Big Sky, MT, USA, 6–13 March 2021; pp. 1–10.

27. Fiorentino, M.; Radkowski, R.; Boccaccio, A.; Uva, A. EMagic mirror interface for augmented reality maintenance: An automotive case study. In Proceedings of the International Working Conference on Advanced Visual Interfaces, Bari, Italy, 7–10 June 2016; pp. 160–167.

28. Nilsson, S.; Gustafsson, T.; Carleberg, P. Hands free interaction with virtual information in a real environment. In Proceedings of the 3rd Conference on Communication by Gaze Interaction—COGAIN, Leicester, UK, 3–4 September 2007; pp. 53–57.

29. Spectral TMS. Augmented Reality Applied to Your Industrial Site & Approved by the Field. Available online: https://www.spectraltms.com/en/ (accessed on 7 October 2021).

30. Tsai, C.H.; Huang, J.Y. Augmented reality display based on user behavior. *Comput. Stand. Interfaces* **2018**, *55*, 171–181. [CrossRef]

31. Tsai, C.H.; Huang, J.Y. PGPS: A Context-Aware Technique to Perceive Carrier Behavior from GPS Data. *J. Internet Technol.* **2018**, *19*, 1413–1423.

32. Blattgerste, J.; Renner, P.; Pfeiffer, T. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In Proceedings of the Workshop on Communication by Gaze Interaction, Warsaw, Poland, 15 June 2018; pp. 1–9.

33. Dunston, P.S.; Wang, X. A hierarchical taxonomy of AEC operations for mixed reality applications. *J. Inf. Technol. Constr. ITcon* **2011**, *16*, 433–444.

34. Fawaz, H.I.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.A. Deep learning for time series classification: A review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963. [CrossRef]

35. Hololens2. Available online: https://www.microsoft.com/en-us/hololens/buy (accessed on 7 October 2021).

36. Grandini, M.; Bagli, E.; Visani, G. Metrics for multi-class classification: An overview. *arXiv* **2020**, arXiv:2008.05756.