International Journal of Advances in Production Research

Industrial Augmented Reality: A Framework for Defining Requirements

Jan Luca Siewert[a,1], Detlef Gerhard[a,2*]

[a] *Digital Engineering Chair, Ruhr University Bochum, Germany*

[1]Jan.Siewert@ruhr-uni-bochum.de [2]Detlef.Gerhard@ruhr-uni-bochum.de

Abstract— **Augmented Reality (AR) boasts a wide array of applications throughout the entire product lifecycle; however, its adoption in industrial settings is often impeded by factors such as high setup costs, poor system integration, and limited modality. Despite these challenges, current Industrial Augmented Reality (IAR) applications exhibit a significant overlap in components used for information gathering and visualization. This paper describes the current state of modern IAR architectures and introduces a novel framework for the definition of requirements specific to IAR ap-plications. This draws upon the principles of human-cantered design to provide a structured approach for integrating IAR more effectively in industrial contexts. Finally, a case-study shows how the framework can be used to help imple-ment an assistance system for assembly of a simple product.**

Keywords— **Augmented Reality, Industrial Augmented Reality, Requirements Engineering, Human Machine Interaction, System Design**

I. Introduction

Industrial Augmented Reality (IAR) has many applications across the whole product lifecycle, including as-sembly support, service and maintenance, or marketing applications [1]. Various studies show that the use of IAR can be advantageous, for example in reducing training times or the number errors when performing complex tasks [2], [3], [4]. Despite more capable AR hardware [5] and recent progress in presentation and tracking quality [6], industrial adaption for IAR remains low [7]. Future research should therefore focus more on organizational issues instead of technology alone, to make implementing AR more cost-effective and reducing training times [7].

The biggest issues relate to available software as well as required implementation and modification time to integrate them into existing processes. This is especially true for small and medium-sized companies (SMEs). One strategy might be introducing more modular solution that reduce these required implementation times and could help boaster adoption for SMEs in particular [5].

Even when working with those modular architectures, the afore mentioned limitations and trade-offs makes the requirement definition for these applications a special challenge in and itself. In a modular system, one has to decide if existing modules can be reused or new ones have to be developed, and if so, how functionality should be bundled into independent modules. This contribution therefore presents a methodology for implementing a mod-ular IAR system out of reusable components. To describe potential applications for such systems, IAR use-cases are described and classified in more detail. Then, potential approaches and architectures for modular IAR systems are described. Based on an overview of the current literature regarding requirement definition for AR and IAR systems, the mentioned methodology is developed. This includes both a proposed process based on DIN EN ISO 9241-210 [8] and some considerations and essential inquiries one should ask when defining requirements as a starting point. Finally, the proposed methodology is applied to a case study.

II. State of the Art and Prior Research

To be able to describe a framework for requirement definition, first, the current state of the art and research gaps are described. To be able to generalize and break down monolithic IAR applications into reusable components, a common terminology to describe the usage context has to be established. For that, classification schemes for IAR use-cases are presented. Then, different approaches for modular architectures are presented. Finally, com-mon requirements and frameworks for requirement engineering for IAR are described.

A. State of the Art

Today's AR hardware is already adequate to support even large-scale AR applications [5]. A lot of progress also has been made in presentation quality, as well as tracking and registration [6]. Context-aware AR applications have been proposed numerous times, for example with the introduction of deep-learning methods. The process made in display technologies, presentation quality, as well as tracking and registration is highlighted in [6]. While content can be presented using 3D models, texts, and symbols, the most appropriate representation used is highly dependent on the operator's preference [6], [9], [10].

*Corresponding Author

Today, there are four predominant presentation technologies used in AR: video-see-through (VST) headsets, where the operator sees both a camera feed and virtual information through displays; optical see through (OST), where the operators directly sees the environment; mobile AR, where smart devices are used to render the camera feed and overlays; and projection based AR, where information is projected directly in top of the environment. All of these have different advantages and disadvantages, as [11] show with expert interviews. VST allows hands-free operation and high-quality overlays, while raising ergonomic problems. Furthermore, because the environment is only observed through screens and cameras, it raises both safety and security issues. On the other hand, OST allows direct perception of the environment, with the downside of decreased quality of virtual overlays. Ergonomic issues prevalent in these kinds of head mounted devices (HMDs), as well. Mobile AR allows known interactions. However, hands-free operation is an issue, as the operator constantly has to place and pick up the device. Projection based AR offers good compromises, allows hands-free operation, undisrupted overview of the environment, and no additional weight or devices to wear. Issues are its inflexibility and possible occlusions [11].

The different trade-offs can be seen on the example of maintenance specifically. There, mobile AR has a lot of advantages, like the ease-of-use and reliability. HMDs are therefore most useful when hands-free activity is required. That is usually during after an initial information phase during the actual execution [9].

While a lot of research has been done to improve model recognition and tracking, even recent AR applications mostly rely on a marker-based approach because of better reliability. Besides this and the afore mentioned ergonomic issues, organizational challenges, like data integration and content authoring are major challenges for the adoption of AR in the industry [12].

When the accuracy requirements are higher than today's tracking and registration methods allow, combining external sensors becomes a viable alternative [9]. This is also often used in industrial applications [13]. Industrial experiences also highlight that HMDs are mostly not suitable for industrial use while that mobile and projection-based AR coupled with external tracking becomes a viable solution [13]. Other external sensors that might be coupled with IAR systems might be used for data acquisition, to monitor the environment, but also depth cameras for motion capture [9]. Additionally, industrial experiences shows that workers prefer and require different level of support or else may feel hindered by an assistant system [13], [9].

B. Industrial AR Use-Cases

Röltgen and Dumitrescu present a systematic literature review about the different use-cases of IAR. They define a total of 26, like marketing, product design, assembly, or maintenance support. They also present a classification scheme that can be used to describe a use-case. To describe the context for the system, they define four types of actions a system can support: inform, execute, plan, and control. It is influenced by the field of application in the product lifecycle the system is used. For the lifecycle phase they propose to use (1) procurement, (2) engineering, (3) production, (4) logistics, (5) maintenance, (6) decommissioning, and (7) training. Additionally, they differentiate whether the system effects the virtual or real world. Technological factors of the systems focus on the aim of the augmentation (performance enhancement, qualification, or enhanced perception). This influences the spatial location of information (user, object, or environment), the temporal context of the augmentation (whether it refers to information from the past, present, future, or is fiction). Finally, they differentiate if it is desirable to separate virtual information from reality or not [14].This classification scheme spans eight dimensions with a lot of possible options, which is in contrast to the 16 observed applications. Furthermore, the classification introduces both the context and the technological implementation, which makes it suitable to describe existing applications, but is not as useful when trying to describe to goal of a future application, alone.

Therefore, we have proposed an alternative classification based on this [15]. Focus is the description of the use-cases alone. This classification scheme keeps both the supported action and the supported lifecycle phase. These are usually well-defined when implementing an assistant system. As third and final dimension, the authors propose to use the desired level of support. As discussed above, using a suitable level of support for the specific workers is an important factor to consider. The degree of support might range from low, which includes visualizing information while the operator is responsible for decision-making, to high where specific actions to perform are recommended by the system. A special case in this category is collaboration with a remote expert. This is considered part of this category because the support is not offered by the system, which is only responsible for facilitating communication, but by using the expert's expertise.

C. Modular IAR Architectures

Modular architectures for AR applications have various advantages for the different roles involved. As the implementation work can be broken down into the different domains, specialists can effectively work on them individually. Functionality can be evaluated more effectively and generalized into reusable components. This allows for quicker prototypes and testing, as well as for increased flexibility to address the needs of specific users [16]. The contribution proposes to create reusable components along the traditional three-layer architecture of software systems. The view layer is separated into interaction modules, presentation modules, and tracking modules that incorporate their data into a universal world module, that controls placement of virtual information. This approach allows the combination of multiple sources for tracking information, like AR tracking and GPS. The application layer is responsible for integrating the data, for example by providing the necessary algorithms for sensor fusion. A task flow engine models a finite-state-machines, where each state is associated with necessary support documents that are shown to the user while the state is active. Context aware services abstract external services that might be added to the AR system, like printing. The user interface is modelled on top of a specific markup language and is also associated with a state machine, comparable to the task-flow engine [16].

For mobile devices specifically, the three core components are responsible for tracking, rendering a camera image, and rendering 3D components [17]. In their work for mobile devices, they decouple these functionalities to allow an easier implementation and adoption for content. Both works show the advantages such a modular architecture can have for adaptability and flexibility. However, the fundamental ways AR systems are developed has shifted in recent years. The described capabilities are provided by software development kits (SDKs). This includes both mobile devices, for which various available SDKs bundle tracking and rendering functionality, but even more so for HMDs, where this core functionality is provided directly by the operating systems.

Recently proposed architectures focus on IAR applications and shifted away from the technical implementation to data integration and communication between different applications. Instead of dividing the technical implementation on a device into the three layers, one can separate between the visualization and interaction layer, a data transport layer, and a data acquisition layer as a viable alternative for IAR applications [18]. The data acquisition layer could connect to existing PLM, ERP, and IoT systems, while the data transport layer could exist as edge computing and be responsible for gateways and data proxies to Cache dynamically created IAR information. Finally, the data visualization layer is responsible for interactions with the operator [18].

To facilitate communication between different services in IAR applications, a distributed and service-oriented architecture can be used [19]. Different services, like object recognition, barcode decoding, or knowledge management, register availability on a service registry. Data from these services can be accessed through a companion device, that then passes the data on to the actual AR device. The architecture focuses on service and maintenance tasks across multiple companies. The work plan is described using Business Process Model and Notation (BPMN) [18]. The proposed architecture shows how modularity can facilitate shared implementation work across multiple companies, but focuses on data flows and communication, not on the technical implementation of the AR content.

Content for AR applications can be shown in different ways. The different presentation devices, VST, OST, mobile AR, and projection, all have different advantages and limits, so that the most suitable devices for a given use case must choose on a case-by-case basis [11]. Studies also show performance differences between content representations. For example, presenting 3D models are especially useful when highlighting blind spots where the operator's view is obstructed. It decreased the completion time compared to 2D renderings and other content representation times [20].

A comprehensive literature review for content representation describes IAR content as a combination of a feature, the asset or content representation used, and the anchor that describes its position in space [21]. The content types that can be used can be text, signs, photography, video, drawings, technical drawing, a product model, or auxiliary models, like arrows [21]. Similar to the output device, the most appropriate representation to used depends on many factors. The most common ones are text, symbols as a combination of 2D signs and 3D auxiliary models, and product models. The fact that the same data can be represented differently allows for more flexibility in the development of AR systems. As not all output devices support all data representations, e.g. a projector not supporting full 3D models, alternatives may be providing a 2D drawing or rendering instead.

Based on this works and assumptions, we have developed a dynamic architecture based on reusable components [22]. Based on the three layers established in [16] and [18], the architectures devices a data, application, and view layer. The data layer is responsible for data persistency and interfaces to other enterprise systems. It may implement a proxy to cache dynamic content and provide services that may be used to convert data representations by changing data formats, or by transforming one representation to another, e.g. by rendering a 3D model into an image. These services registry their availability into a service registry, as established in [19]. The application layer is responsible for executing the work plan. It is modelled as a state-machine. Each state describes the desired AR annotation that should be shown to the operator, while state transitions are triggered by occurring events. Furthermore, the application checks if available components are in fact capable of executing the work plan with its required data representations, and schedules data conversions on the data layer when necessary. The view layer is divided into components responsible for interaction and presentation. While interaction systems emit events, e.g. by user interaction or through automated systems or components, like buttons, voice interactions, or external sensors, presentation components show AR annotations to the operator. An AR annotation combines an anchor point with multiple data representations. When an annotation contains more than one representation, the most suitable one is used in the current environment. The view layer communicates with the application layer through MQTT. Presentation components subscribe to a combination of anchor point and used data representation. By utilizing MQTT topics, data is only forwarded to components that support it.

This architecture differs from other proposed solution in two ways:

1. it explicitly supports not only full AR on HMDs or mobile devices, but spans from these AR implementations to already established pick-to-light systems. Because the required data representations and anchor points can be fulfilled through multiple components, a pick-to-light system may provide the object anchor, while 2D data representations are presented on a monitor.
2. the proposed components make use of anchor points and data representation for typical IAR applications, while leaving the technical implementation of those as a black-box. This middle-ground maintains a high flexibility in the developed systems, while establishing a common ground for defining and describing the individual components.

To describe data representation, the architecture focuses on the technical aspects. In contrast to [21], photography, drawings, and technical drawings are combined into an image representation type while both 2D signs and 3D auxiliary models are combined into a symbol representation. The main motivation behind the change is that the representation type is used to define capabilities of components instead of the effect it has on the worker. As shown in figure 1, the anchor points used are adapted to IAR applications as well. Detached content, or content without an anchor, is shown in screen-space, for example in a head-up-display style in an HMD or at another, ease-to-read position chosen by the component. Content can be anchored spatially, usually at a specific product, or by anchoring it indirectly on an object using a given offset. Content may also only reference a given location. Additionally, content can be anchored to either a specific instance of an object or a given object type. The latter is especially important when indicating the location of a storage container.

Before implementing an IAR system, the desired functionalities and requirements have to be defined. To support this process, various frameworks exist for systems in human-machine-interactions in general and IAR in particular. These are presented in the next section.

D. Requirement Definition for Industrial AR

When designing IAR systems, various factors need to be considered to ensure the project's success. Aspects include technical considerations for the output hardware, how information is presented, and which information is required, how the operator can interact with the system, authoring and availability of content and data integration and processing. For that, a human-centred design should be followed [23]. Additional, common requirements for IAR applications include cost efficiency, data security, established regulations and laws, including ergonomics. For the first aspect, short setup times, the overall reliability of the system, as well as the accuracy of shown information, which are often gathered in near-real-time, are crucial [24].

Based on expert workshops, some general requirements for a maintenance support application are described in [25]. First, one of the biggest advantages of the technology is portable access to relevant information. This includes information about required tools, materials, and spare parts. The Spatial information in IAR system can for example be used to assist orientation in an assembly. Workflow guidance can be given more effectively using 3D animation, while the hands-free nature of most HMDs makes taking notes and pictures easier. Access to live telemetry data or cross-referencing existing cases are other advantages of permanent access to information. Finally, video calls with experts allow more effective work procedures. Other often requested features are an offline mode or recording of statistical data [21].

[26] describes factors that affect industrial adaption in four categories: task, workforce, context, and technology. Considerations regarding the tasks, it should be sufficiently complex to justify the overhead of an AR solution. Tasks that benefit from remote work and off-site experts benefit greatly from AR. Information that is presented by an IAR system needs to be established and codified and required information for defining and describing the tasks needs to be available in the first place. Additionally, the skills of the workforce need to be considered. Providing simple instructions to expert technicians does not benefit them, neither does an instruction that requires some expertise for novice workers. The right balance between technical and practical abilities and shown information needs to be found. Additionally, digital skills and the technology acceptance level of the workforce needs to be taken into account. For the work context, leadership and organizational processes to be respected so that the solution does not only integrate into established software systems but processes as well. Additional considerations are accessibility, connectivity, comfort, but also tool availability and ease of use as technological success factors. [26]

When data and instructions are only available on paper, [27] proposed a simple process for converting those into interactive AR work plans. First, the existing manual is analysed. Then, the work is divided in atomic actions. These can either be concepts or references, that should describe information using simple imagery and text, or actionable tasks. These should use annotated images, descriptive graphical symbols, and descriptions of simple texts. Finally, the actions should be grouped and organized with a single message per entity, consistent texts, and recurring symbols [27].

Based on the general requirements and [8], a process for defining requirements for IAR systems is developed in [28]. First, the user context and task in analysed and the desired hardware is selected. Then, user requirements are collected based upon first tests with the hardware, and suitable interaction schemes are selected. The designed solution is developed and evaluated. For evaluation, two tests in laboratory environments and one in the field of application are proposed [28].

For evaluation, [29] presents an overview of different usability studies for AR in general and IAR in particular. For those, most studies are recruiting young university studies and are mostly lab-based tests. The NASA TLX score is the de-facto standard for testing usability, while time and error rate and accuracy might be evaluated as secondary metrics [29].

As in [28], [30] puts a lot of focus on the hardware-selection and developed criteria for evaluation HMDs in particular. Relevant factors are the cost, weight, and technical aspects, like the field-of-view battery power, camera resolutions, as well processor speed, available storage, and memory. Other important factors are available programming interfaces, availability, and connectivity to external sensors or audio devices, as well as the used OS. Finally, ratings for dust and water resistance might be critical depending on the use-case [30].

The most comprehensive framework for requirement definition for IAR system is presented based on a literature review and expert interviews in [31]. First, they define 21 task keywords for atomic actions in maintenance and assembly task, for example remove, push, or measure, in the categories locate, check, operate, or other tasks. The main information type that can be indicated to the user might me operation, indication of a goal, or showing a sample. This is used to define a usage context.
They define the types for coordinate systems or anchor points for virtual information: the coordinate system of the HMD in the form of a head-up-display; the world; a body part; an object; or a combination of the HMD and the object or of two objects. For data representations, they follow [21]. Then, they outline 18 conditions that can be used to select an appropriate combination of data representation and anchor point based on the usage context. Some examples are to use only texts and images for simple instructions, to not use an object anchor when the system requires accuracy below 1 cm, or to not show product models that do not fit into the field-of-view of the headsets.
Main limitations of the study are its focus on HMDs, the fact that preferences and experiences of the user are not taken into account, and that it only describes requirements for the overall system.

To build on the advantages for modular architectures outlined in the previous works, for example by using the architecture described in [22], the requirements must not only be defined for the overall system, but there needs to be a process for decomposing the system and defining the individual components.

III. Proposed Framework

As described in the previous section, splitting monolithic IAR systems into reusable components has several advantages. In previous works, the authors have shown the advantages of a three-layer architecture with separate interaction and presentation modules in the view layer [22]. The interaction events are responsible for triggering events that advance a work plan, while presentation modules show annotations to the worker. Annotations are a combination of a data representation, like images, videos, or 3D models, and anchor points. An annotation can be shown in a single representation component or split between them. The anchor point, e.g. a specific object or other position in space, is the spatial information attached to the information.

When designing and implementing such a modular system, a developer must select the required events, anchor points, and suitable representations. As described earlier, there is no single best system for a given use-case. Rather, the solution depends on available data, expertise, existing systems and experiences, and the preferences and requirements the worker that uses the IAR system. The presented methodology for defining and implementing an IAR system out of reusable components is based upon best practices and experiences from the literature, adapted to the component-based architecture. It follows the standard for human centred design for interactive systems [8] to describe the context and requirements for the overall system, define required events, anchor points, and data representation the system should be capable of.

The basic process and its relation to [8] is presented in figure 2.

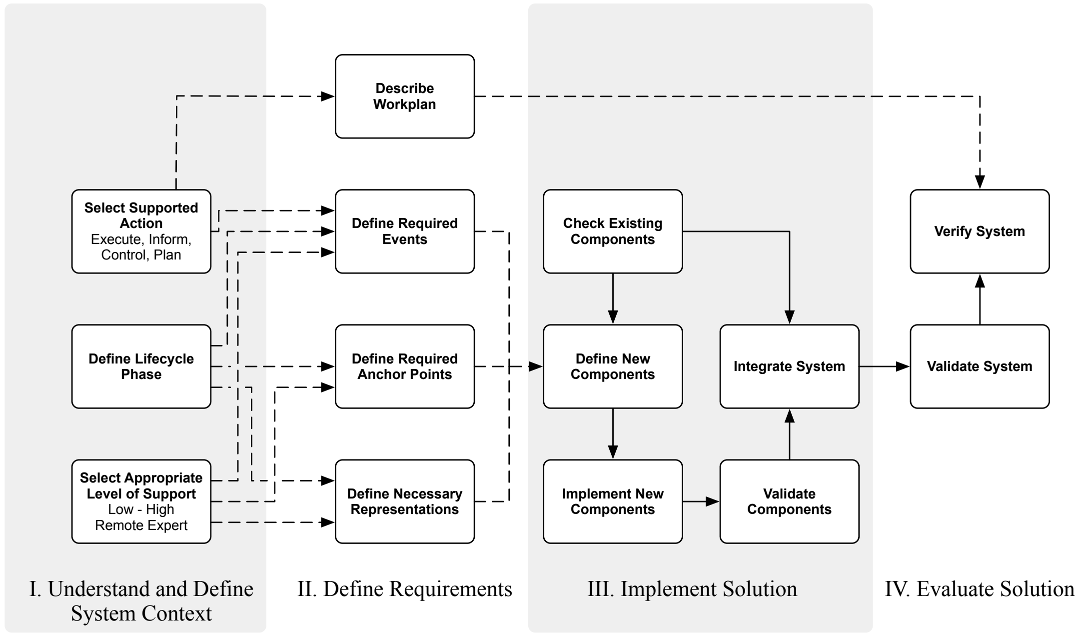

FİGURE 1. Process for Defining and Implementing a Modular IAR System

A. Understanding and Defining Context

As already stated, the usage context of an IAR application can be classified into various categories. Because the context in which the IAR system should be described without predetermining the technical implementation, the classification presented in [15] is used. Therefore, the system context is described by the supported action, the life-cycle phase, and the desired level of support.

The classification in these categories is the first indicator for the requirement definition in the second phase. The supported action can be one of four types: 1. an *execution task*, where a specific goal is to be accomplished; 2. a *control task* where the status quo is compared to a target state; 3. *planning tasks* in which the operator has to choose from an array of alternatives to establish the most effective course of action, given a set of predefined parameters; 4. and *inform tasks* where relevant data or insights are communicated to the operator, not with the immediate aim of achieving a specific result within the system, but to provide essential information that might be beneficial for future actions.

These actions effect the requirements of the system in two ways: the structure of the work plan as well as the type of events typically required to support them. First, the action has a fundamental influence on the structure of the work plan. As described, the work plan is a finite-state machine where the active state describes the presented information, while events form interaction systems are mapped to state transitions. In execution tasks, a state usually correlates with an atomic action that should be performed. When an event indicates the completion of the current task, the work plan moves to the next. Additionally, these tasks might be grouped together as to not distract experienced workers. This results in a tree-like structure of the work plan. Furthermore, the required events for such an execution task are typically limited to navigating this tree structure. Typical events are going to the next or previous step or viewing more or less details on one to navigate deeper inside the tree.

Controlling tasks are similar in that the properties to be checked and validated are often predefined in a task list. Therefore, the structure of the work plan and the required events are very similar. The main differences are that control tasks generate output. Therefore, the events that a system supporting control tasks supports must include a payload, e.g. by recording positions or measurements. Similar, planning tasks are performed to select from a choice of alternatives. The choices, for example selected products, configurations, or positions, must be recorded, as well. Therefore, the system also requires event that contain additional data. In contrast to control tasks, the structure of the work plan is not as predefined. It may contain valid configurations that must be converted into a suitable finite-state-machine before using it in a system.

Informing tasks do not generate output data directly. Rather, they are supporting the user in its decision-making. Similar to planning tasks, the required state-machine must be created beforehand out of possible states of the application, e.g. by including a state for each object an inform application should present data to.

The lifecycle phase that the system supports has the greatest impact on the specific data that can be presented, as some data is only available in later phases. While 3D models are often created early during product development, telemetry data is only available in the usage phase. It should be noted that the context might be described using multiple phases. When a product is in production in an assembly line, the product is in the phase “production” while the assembly line itself is in the “usage” phase. This makes it viable to show telemetry data, like manufacturing machine parameters or robot paths, while the actual product is assembled.

Defining the necessary level of support is crucial for defining and developing successful and helpful IAR applications. While implemented IAR prototypes are frequently presenting very detailed instructions [15], for example to support students in laboratory environments [29], research continues to show that this high level of support reduces performance of expert workers [3]. This is furthermore surprising considering preparing a complex task into a detailed and interactive AR instructions requires more work than using a limited scope for AR, like only using it to localize objects. A special case in the level of support is the remote expert. Here, the IAR application is only responsible for relaying information between the operator and the remote expert instead of giving support by itself.

B. Defining Requirements

While the previous section presents some general guidance broad requirements to describe the system context, the actual requirements depend on various factors and are mostly unique for any given system. This section presents influence factors that affect the actual requirements together with some sample questions that might be asked when detailing the requirements, including potential influences on the designed system.

Table1 1
Aspects and Guidelines that are relevant when defining requirements for IAR systems

ID	Category	Aspect	Guidelines
Q11	Task	Supported Action	Does the task require hands-free usage?
Q12		Accuracy	High Accuracy required? Consider external tracking systems
Q13		Task Complexity	How complex is the task? Which information needs to be provided to the worker while performing an action?
Q14		Spatial Information	Is orientation helpful? To which points in space should the operator be guided?
Q21	User	Experience	What are prior experiences of the user? Does he need continuous support or just a training phase? What information is relevant, what might be known to him?
Q22		Preferences	Can you offer choices, like what and how much information is needed. Consider alternatives to body-worn devices, like HMDs
Q23		Technology Level	How experienced is the user with technology, in general? Use simple and familiar interactions, where possible
Q31	Environment	Lightning	Is the environment well lid? If not, camera-based tracking systems might be unreliable
Q32		Location	Obscured or dark? Use spatial hints to guide user to the location
Q33		Spatial Stability	Is the environment changing or fixed? For fixed setups, spatial or projection-based AR might be a good alternative
Q41	Business Context	Existing Systems	Are there existing AR or worker support systems, for example pick-by-light systems that can be integrated?

Q42		Data Interfaces	With which interfaces should the system interact, for example to record data?

The first aspect that needs to be examined is the supported task. The overall supported action was already explained in the previous section. One central question when analysing the specific task is whether it requires hands-free usage. While execution tasks often greatly benefit from hands-free operation [25], it is always a trade-off with safety, ergonomic, and cost efficiency, which are also important requirements to consider [24]. The second aspect to consider is the required accuracy. Today's camera-based tracking is not capable of precise tracking. In [31], an accuracy requirement of 1 cm is mentioned as a rule-of-thumb at which precise AR hints should not be used. While [31] recommend using images or videos in such cases, external tracking systems can also be considered, as demonstrated in [13]. As AR is fundamentally a spatial form of presenting information, the most natural type of information to present is localization and orientation. AR can be efficiently used to guide workers to specific locations or draw attention to parts of a machine. Other aspects, like presenting 3D content, always comes with an added complexity during content generation and consumption. Symbols and texts are the most common representation of data in IAR applications, while 3D models and animations are only relevant to specific applications [16], like marketing or product design.

The second aspect to consider is the users and workers that will use the assistant system. Central questions are their general experience with the specific process, they individual preferences, but also the technology level they are familiar with. First, experienced workers need different support and information than novice workers or employees with cognitive impairments. Showing to many or to detailed instructions might increase the cognitive load for experienced workers, while novice workers want reduced information as they get more proficient with the task at hand [3]. For novice workers, the usage of AR has shown to reduce the training time and number of training cycles when performing complex tasks at the first time [32]. Therefore, one should consider using AR for training of workers not experienced with the specific tasks, unless variations or other requirements require continuous access to information for experts as well. On the other hand, workers with cognitive impairments might benefit from detailed instructions that are visible at all times [33]. Overall, one should always question whether data representation is actually necessary in this form for the worker at a specific task. Instead of using sophisticated hardware to display 3D animations, the necessary spatial information might be provided using a traditional pick-by-light setup, instead.

As different information can be presented in different ways using AR [34], one can also consider the preference of the users when designing an IAR system. This might include placement of information, filtering and deciding on the level of support wanted, to basic accessibility settings like text sizes. Similarly, the interactions with such a system should be designed multimodal [6], for example by adding an alternative input method to voice and dictation. Similarly, the technology level the users are familiar with need to be considered. Studies show that HMDs in particular tend to have higher mental load than other AR and non-AR devices, like projection-based or mobile AR or paper manuals [35]. Mobile AR, for example, uses smartphones and tablets with interactions the users might be more familiar than novel HMDs. This also underlines that the most sophisticated IAR application might be unsuitable for a given task if it does not take into considerations the needs and requirements of the user.

The physical environment in which the system should be used limits the technologies that can be implemented successfully. In dark environments, camera-based AR tracking is often limited. At the same time, users benefit from spatial information and directions in these circumstances [31]. When the environment and setup is not changing, fixed AR setups offer better accuracy and reliability, for example by using external tracking systems. Data can also be made available through projection-based AR, which reduces mental load and improves economy, which are important aspects to consider when defining the requirements.

Finally, the business context needs to be taken into account. Important requirements are the existing business processes in order to define the data that the system needs to collect and the data format used. This could, for example, be used to create a maintenance report after an execution task or saving results of quality control. Then existing interfaces to other enterprise tools, like product lifecycle management systems for product data and models, enterprise resource planning for tasks and orders, as well as internet of things platforms for live telemetry. For these interfaces, data availability and formats, potential conversions as well as access rights and caching potentials.

Overall, the given aspects and considerations can be used as a framework when defining requirements for IAR systems. They are developed based on experience and best practices from the literature. These general considerations can offer a good baseline from which to specify more detailed requirements that fit the use case to implement.

C. Implement Solution

Based on the requirements that have been defined in the previous step with help from the provided guidelines, the next step is to implement the IAR system. Then, suitable components have to be defined or selected by breaking down and decoupling the functionality required by the overall system. These include the required interactions and events, the data that should be presented, and the spatial information that is necessary for the user.

As reusability is one of the main advantages of the proposed architecture, checking and evaluating already existing components is a first step. As each component is defined by supported events, data representations, and anchor points, these can be evaluated against the requirements defined before. When deciding on reusing components, one can consider transforming data into a simpler representation that is supported by the existing one, for example by rendering a 3D model into a 2D rendering or converting it into supported formats. In general, the overall requirements can be fulfilled by combining components. Providing both assembly instructions and supporting the localization of parts to be used during an assembly step might be implemented by a pick-to-light component for the parts and a monitor component that displays the instructions. Alternatively, both features can be integrated into a single component, for example in an HMD or using projection-based AR.

When the requirements cannot be satisfied using existing components, new components need to be implemented. First, the required events, anchors, and representations need to be split into one or more components to implement. In general, one should separate interaction components and their events from presentation components that support anchors and representations. Interaction components often support multiple events. A voice recognition system might enable event to navigate through a work plan by proceeding to the next or previous event or asking for help. Exceptions are events that hold additional data. A component where the operator enters information, for example the results of a conducted measurement, typically only support a single event type that holds the entered information. For presentation components, the separation should be done primarily by their supported anchors. These describe which objects or other features the component can recognize and use to display information near them. Typically, a presentation component only supports a limited number of different objects. They are often only available after the component is calibrated. This could include providing reference images or 3D models that are used in the tracking algorithms.

In the next phase, the components are implemented individually based on their defined interfaces. This decomposition enables the work to be split across teams or companies. For example, a team could implement a model tracking algorithm and provide it as a component without being concerned with how this functionality can be used in an IAR system. This separation of concern greatly increases the flexibility and efficiency in the implementation phase The component is then described by its capabilities. For an interaction component, these are the events that the component emits, while presentation components are described by supported anchor points and data representations. Additionally, each component is assigned a unique identifier. In this phase, the implemented component is verified against their requirements to ensure that it is able to fulfil the advertised capabilities.

Both existing and newly implemented components are then integrated into the overall system. The application planning module collects the individual capabilities and compares them to the required capabilities from the work plan that should be executed. From there, it can be determined if the overall system can support all features required. As supported anchors might require additional calibrated, for example by providing tracking information for additional objects, at this stage, the planning module might prompt the operator to configure missing capabilities. Only after all capabilities are fulfilled can the system be used with the provided workplan.

D. Evaluate Solution

Finally, the overall system needs to be evaluated. Common approaches are validation first in a laboratory environment and then in the final environment by end users [28]. The NASA TLX index is often used to evaluate the performance of IAR systems, sometimes in combination with performance indicators, like learning time, task execution time, and error rate [29]. Overall, it is important to evaluate a solution in the specific workplace where it will later be used to identify problems with the environment, such as camera-based tracking issues or speech recognition. Additionally, workers who will later use the system, need to test the implementation in this phase, as well. During this phase, issued surfaced, and general feedback can be used to further improve individual components and the overall system. Finally, it is important to check the reusability of the newly introduced components to ensure that they can also be used in future systems.

IV. CASE STUDY

To demonstrate the proposed methodology, a case study with an assembly support system is presented in this section. After describing the context of use, the requirements are defined on the basis of the key questions outlined above. Finally, it is described how the system is decomposed and implemented from individual components.

A. Understanding and Defining Context

The system to be implemented is an assembly support system for the assembly of toy robots. The robots are about 20 cm long and are assembled using plates and screws. The required parts are in storage containers at the assembly station, while a simple fixture is used during the process. The assembly stand is used in a laboratory environment mostly by students and guests without deeper experiences with the assembly. The desired degree of support can be considered medium to high, as detailed and ongoing descriptions of tasks to be performed should be shown to the user. However, different users will have different skills and experiences. The required level of support implies a structured work plan. This can be generated based on the assembly instruction, for example by following the process outlined in [23]. The resulting tree-like data structure describes an assembly step in each state, while similar steps are grouped together to account for the expected varying skill levels of the users.

In the lifecycle phase production, full 3D models of the overall assembly as well of the assembly steps can be provided. Textual descriptions for the steps can be provided. The necessary parts required in each assembly step can be extracted from the assembly procedure, as well. Due to the unknown assembly state, the support system should provide guidance so that the operators can quickly pick the correct part for each assembly step.

B. Defining Requirements

Following the outlined guidelines in the previous chapter, the requirements for the overall system can be defined. The supported action (Q11) is an execution task, making hands-free usage an important consideration. In fact, this use-cases benefits a lot from it, because the operator will regularly use both hands, e.g. when screwing the plates together. Hands free usage is therefore desired. However, it is not complex enough to warrant 3D animations (Q13).

As only instructions and 3D models are used, there is no need for especially high tracking accuracy (Q12). Spatial information can be used to guide the user to the correct storage container and highlight the points on the assembly, e.g. for showing screws, as operators are not familiar with the assembly stand (Q14)

As the support system will be used mainly as a demonstrator there will be no dedicated training phase. Instead, users will require ongoing support, but this will vary depending on their previous experience (Q21). As this is purely a demonstrator, the system does not need to adapt to preferences of the user (Q22). The technology level the users are used to is expected to be high (Q23), so a full-fledged IAR application including advanced features, like voice or gesture recognition, can be used. The environment is a laboratory setting, so lightning (Q31) and obstructions (Q32) will not be a concern. Because of the dedicated assembly stand, including fixtures, there is a high spatial stability (Q33). The setup and surroundings will not change often. This makes a fixed setup, e.g. by using a projector, a viable option. Additionally, the setup can be calibrated manually, for example by manually configuring spatial hints. This allows the usage of object anchors without dedicated object tracking capabilities. The locations of the storage containers or screws, for example, can be calibrated and saved once or after changes in the setup. While data and information is available in enterprise systems, like 3D models in a PLM environment (Q41), manually preparing and converting data for the demonstrator is a viable alternative to implementing data interfaces (Q42). Based on these requirements and concerns, the system can be decomposed into individual components.

C. Implement Solution

The main requirements for the system are presentation of textual instructions, a 3D model of the current assembly step, symbols or indicators at objects or parts of objects, interactions for navigating through the state machine, and recognition of the parts the user has taken out of storage containers. Therefore, the system requires text, symbol, and 3D model data representations, as well as interactions for navigation (next, previous, and help), and the selection of specific parts. This can be broken into individual components as shown in figure 3. The components are implemented in two dedicated modules. The first is responsible for rendering the instructions and the 3D model, together with the navigation interactions. The second is capable of indicating the position of storage container and can register when the user grabs a part out of one of them.

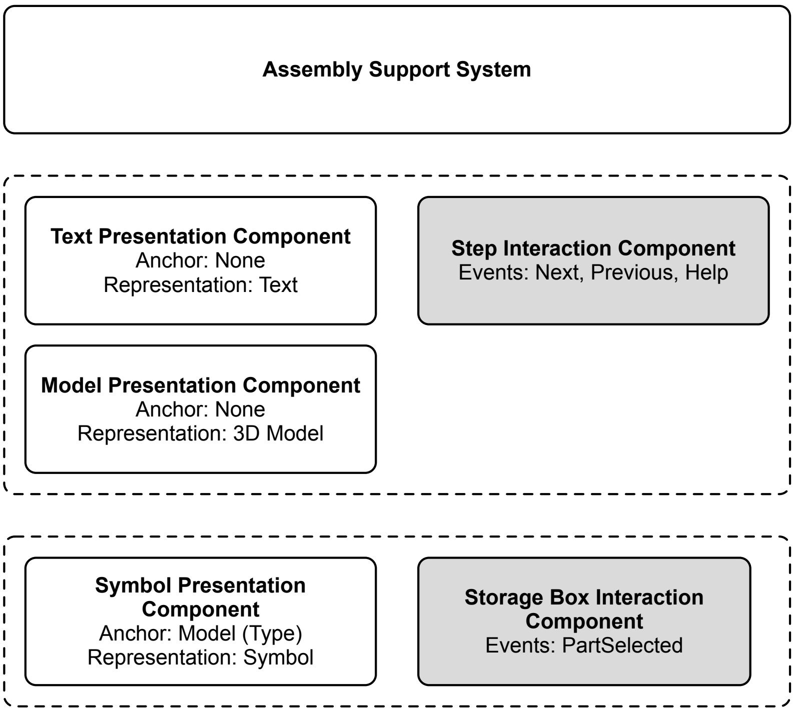

FİGURE 2. Implemented Components for the Assembly Support System

Overall, the assembly station is equipped with a projector and a depth sensor. Instructions and images are presented directly on the work surface. The position of the storage container is set manually. The depth sensor recognizes gestures as well as the grabbing of objects out of those. The implemented system can be seen in figure 4.

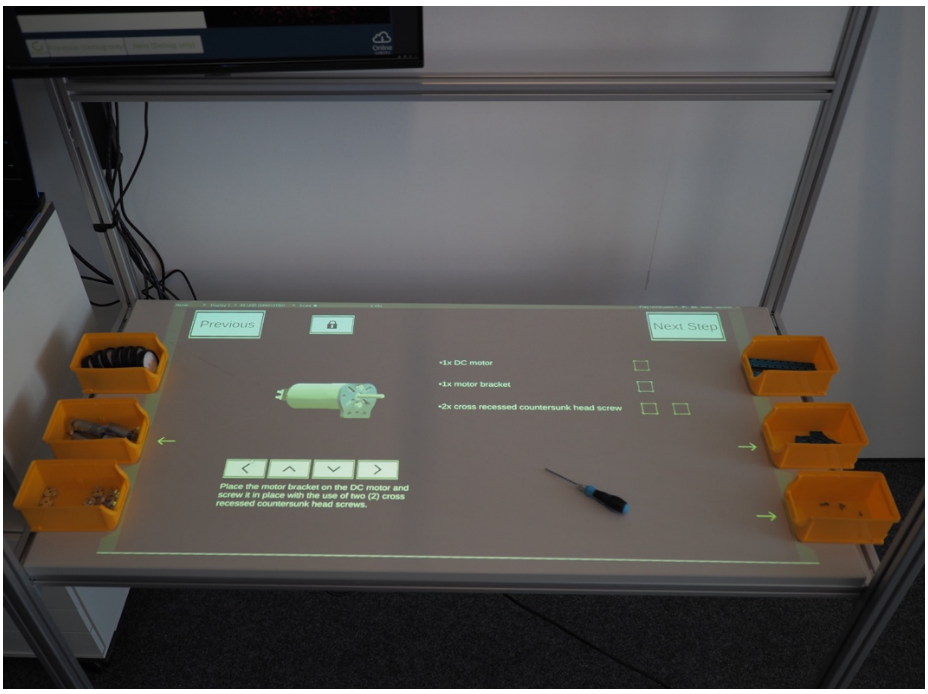

FİGURE 1. The implemented projection stand with highlighted storage containers, instructions, and the 3D model.

D. Evaluate Solution

The individual components have been verified to ensure they fulfil the given requirements. Because the system will be used exclusively for demonstrating purposes, an evaluation and formal validation has not been performed. However, before implementing such a system in a production environment, evaluation should be performed with the target group to ensure the system is supporting them adequately.

V. CONCLUSION

To reduce effort required when introducing and implementing IAR applications in the industry a modular approach offers various advantages. Individual components can be implemented by specialists and reused in other systems. To better fulfil special requirements of individual users, by integrating new components into an existing system. However, such a modular approach increases the complexity when designing a system. In addition to requirements for the overall system, it needs to be decomposed into individual modules and components. A system designer needs to select the functionality for each component so that the overall system can fulfil the requirements, while the individual components are general enough to make them reusable.

This contribution presents a methodology for defining requirements for such a component-based architecture. Based on established standards and best-practices from the literature, a four-step process is described. Guideline questions are provided that support in the requirement definition phase of IAR systems. While the provided case-study focuses on a laboratory setting, the generated insights can be transferred to other use-cases, as well.

Acknowledgment

The headings of the Acknowledgment and the References sections must **not** be numbered. Editorial Board wishes to acknowledge all authors for their kind attentions in complying with the formatting instructions in this template.

References

[1] P. Fite-Georgel, “Is there a reality in industrial augmented reality?” in *International symposium on mixed and augmented reality*, in Science and technology proceedings. IEEE, Oct. 2011.

[2] S. Büttner, M. Prilla, and C. Röcker, “Augmented reality training for industrial assembly work - are projection-based AR assistive systems an appropriate tool for assembly training?” in *2020 CHI conference on human factors in computing systems*, New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–12. doi: 10.1145/3313831.3376720.

[3] M. Funk, A. Bächler, L. Bächler, T. Kosch, T. Heidenreich, and A. Schmidt, “Working with augmented reality?” in *10th international conference on PErvasive technologies related to assistive environments*, New York, NY, USA: Association for Computing Machinery, Jun. 2017, pp. 222–229.

[4] E. Bottani and G. Vignali, “Augmented reality technology in the manufacturing industry: A review of the last decade,” *IISE Transactions*, vol. 51, no. 3, pp. 284–310, Feb. 2019.

[5] V. Bilous, R. Porsch, and K. Spanoudakis, “Augmented reality: Increasing availability and its implication for SMEs,” *The Future of Smart Production for SMEs*, pp. 349–356, Oct. 2022, doi: 10.1007/978-3-031-15428-7_31.

[6] W. Fang, L. Chen, T. Zhang, C. Chen, Z. Teng, and L. Wang, “Head-mounted display augmented reality in manufacturing: A systematic review,” *Robotics and Computer-Integrated Manufacturing*, vol. 83, p. 102567, Oct. 2023, doi: 10.1016/j.rcim.2023.102567.

[7] T. Masood and J. Egger, “Adopting augmented reality in the age of industrial digitalisation,” *Computers in Industry*, vol. 115, p. 103112, 2020.

[8] “Ergonomie der mensch-system-interaktion – teil 210: Menschzentrierte gestaltung interaktiver systeme,” Deutsches Instut für Normung e.V., DIN EN ISO 9241-210, 2020.

[9] M. Runji, Y.-J. Lee, and C.-H. Chu, “Systematic literature review on augmented reality-based maintenance applications in manufacturing centered on operator needs,” *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 10, no. 2, pp. 567–585, Mar. 2023, doi: 10.1007/s40684-022-00444-w.

[10] T. Kosch, Y. Abdelrahman, M. Funk, and A. Schmidt, “One size does not fit all,” *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, Sep. 2017, doi: 10.1145/3123024.3124395.

[11] A. Syberfeldt, M. Holm, O. Danielsson, L. Wang, and R. L. Brewster, “Support systems on the industrial shop-floors of the future – operators’ perspective on augmented reality,” *Procedia CIRP*, vol. 44, pp. 108–113, 2016.

[12] M. Eswaran, A. K. Gulivindala, A. K. Inkulu, and M. V. A. Raju Bahubalendruni, “Augmented reality-based guidance in product assembly and maintenance/repair perspective: A state of the art review on challenges and opportunities,” *Expert Systems with Applications*, vol. 213, p. 118983, Mar. 2023, doi: 10.1016/j.eswa.2022.118983.

[13] G. Stübl, G. Ebenhofer, H. Bauer, and A. Pichler, “Lessons learned from industrial augmented reality applications,” *Procedia Computer Science*, vol. 200, pp. 1218–1226, 2022, doi: 10.1016/j.procs.2022.01.322.

[14] D. Röltgen and R. Dumitrescu, “Classification of industrial augmented reality use cases,” *Procedia CIRP*, vol. 91, pp. 93–100, 2020, doi: 10.1016/j.procir.2020.01.137.

[15] J. L. Siewert, M. Neges, and D. Gerhard, “Ein klassifizierungssystem für industrielle augmented reality anwendungen,” pp. 401–416.

[16] M. Bauer *et al.*, “Design of a component-based augmented reality framework,” in *Proceedings IEEE and ACM international symposium on augmented reality*, in ISAR-01. IEEE Comput. Soc, 2001. doi: 10.1109/isar.2001.970514.

[17] M. Schumann, P. Klimant, C. Kollatsch, and V. Wittstock, “Modulares framework zur entwicklung von augmented-reality-anwendungen,” Feb. 2013.

[18] P. Fraga-Lamas, T. M. Fernandez-Carames, O. Blanco-Novoa, and M. A. Vilar-Montesinos, “A review on industrial augmented reality systems for the industry 4.0 shipyard,” *IEEE Access*, vol. 6, pp. 13358–13375, 2018.

[19] T. Kuster *et al.*, “A distributed architecture for modular and dynamic augmented reality processes,” in *2019 IEEE 17th international conference on industrial informatics (INDIN)*, IEEE, Jul. 2019. doi: 10.1109/indin41052.2019.8972101.

[20] E. Laviola, M. Gattullo, A. Evangelista, M. Fiorentino, and A. E. Uva, “In-situ or side-by-side? A user study on augmented reality maintenance instructions in blind areas,” *Computers in Industry*, vol. 144, p. 103795, Jan. 2023, doi: 10.1016/j.compind.2022.103795.

[21] M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. L. Gabbard, “What, how, and why are visual assets used in industrial augmented reality? A systematic review and classification in maintenance, assembly, and training (from 1997 to 2019),” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, pp. 1443–1456, Feb. 2022, doi: 10.1109/tvcg.2020.3014614.

[22] D. Gerhard, M. Neges, J. L. Siewert, and M. Wolf, “Towards universal industrial augmented reality: Implementing a modular IAR system to support assembly processes,” *Multimodal Technologies and Interaction*, vol. 7, no. 7, 2023, doi: 10.3390/mti7070065.

[23] M. Quandt, A. A. Alla, L. Meyer, and M. Freitag, "Success factors for the development of augmented reality-based assistance systems for maintenance services," in *7. WGP-jahreskongress*, R. H. Schmitt and G. Schuh, Eds., Apprimus Verlag, 2017.

[24] M. Quandt, B. Knoke, C. Gorldt, M. Freitag, and K.-D. Thoben, "General requirements for industrial augmented reality applications," *Procedia CIRP*, vol. 72, pp. 1130–1135, 2018, doi: 10.1016/j.procir.2018.03.061.

[25] M. Lorenz, S. Knopp, and P. Klimant, "Industrial augmented reality: Requirements for an augmented reality maintenance worker support system," in *Adjunct proceedings of the 2018 ieee international symposium on mixed and augmented reality (ismar)*, 2018, pp. 151–153. doi: 10.1109/ISMAR-Adjunct.2018.00055.

[26] S. Aquino, M. Rapaccini, F. Adrodegari, and G. Pezzotta, "Augmented reality for industrial services provision: The factors influencing a successful adoption in manufacturing companies," *Journal of Manufacturing Technology Management*, vol. 34, no. 4, pp. 601–620, Feb. 2023, doi: 10.1108/jmtm-02-2022-0077.

[27] M. Gattullo, G. W. Scurati, M. Fiorentino, A. E. Uva, F. Ferrise, and M. Bordegoni, "Towards augmented reality manuals for industry 4.0: A methodology," *Robotics and Computer-Integrated Manufacturing*, vol. 56, pp. 276–286, Apr. 2019, doi: 10.1016/j.rcim.2018.10.001.

[28] M. Quandt, T. Beinke, and M. Freitag, "User-centered evaluation of an augmented reality-based assistance system for maintenance," *Procedia CIRP*, vol. 93, pp. 921–926, 2020, doi: 10.1016/j.procir.2020.03.053.

[29] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. Swan, "A systematic review of 10 years of augmented reality usability studies: 2005 to 2014," *Frontiers in Robotics and AI*, vol. 5, Apr. 2018, doi: 10.3389/frobt.2018.00037.

[30] A. Syberfeldt, O. Danielsson, and P. Gustavsson, "Augmented reality smart glasses in the smart factory: Product evaluation guidelines and review of available products," *IEEE Access*, vol. 5, pp. 9118–9130, 2017, doi: 10.1109/access.2017.2703952.

[31] K. Tainaka, Y. Fujimoto, T. Sawabe, M. Kanbara, and H. Kato, "Selection framework of visualization methods in designing AR industrial task-support systems," *Computers in Industry*, vol. 145, p. 103828, Feb. 2023, doi: 10.1016/j.compind.2022.103828.

[32] T. Bosch, G. van Rhijn, F. Krause, R. Könemann, E. S. Wilschut, and M. de Looze, "Spatial augmented reality," in *Proceedings of the 13th ACM international conference on PErvasive technologies related to assistive environments*, New York, NY, USA: Association for Computing Machinery, Jun. 2020. doi: 10.1145/3389189.3397975.

[33] M. Jost, A. Luxenburger, S. Knoch, and J. Alexandersson, "PARTAS: A personalizable augmented reality based task adaption system for workers with cognitive disabilities," *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, Jun. 2022, doi: 10.1145/3529190.3529208.

[34] M. Gattullo, G. W. Scurati, A. Evangelista, F. Ferrise, M. Fiorentino, and A. E. Uva, "Informing the use of visual assets in industrial augmented reality," 2019, pp. 106–117. doi: 10.1007/978-3-030-31154-4_10.

[35] M. Funk, T. Kosch, and A. Schmidt, "Interactive worker assistance," *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Sep. 2016.