









Informing the Use of Visual Assets in Industrial Augmented Reality

Michele Gattullo¹ , Giulia Wally Scurati² ,
Alessandro Evangelista¹ , Francesco Ferrise² ,
Michele Fiorentino¹ , and Antonio Emmanuele Uva¹ 

¹ Department of Mechanics, Mathematics, and Management,
Polytechnic Institute of Bari, 70126 Bari, Italy

{michele.gattullo, alessandro.evangelista}@poliba.it

² Department of Mechanical Engineering, Polytechnic University of Milan,
20156 Milan, Italy

Abstract. Augmented Reality is considered one of the most promising technologies for Industry 4.0. Augmented Reality allows to display the needed information at the right time and to locate it in the desired space, superimposing it to the real world. In this way, it could simplify the work of industrial operators in a variety of tasks, from planning and commissioning to assembly and maintenance. Despite the growing research interest and efforts for the development of Augmented Reality applications and technologies for the industrial domain, Augmented Reality is still rarely used in real industrial procedures, often remaining at a conceptual level. In fact, developers fail to completely answer the actual needs of industries and workers because of a lack of knowledge of the context and the absence of guidelines to drive the design and development of Augmented Reality applications. One of the open issues is the choice of the proper visualization methods to display technical information in Augmented Reality, which is not trivial. Both the paper-based documentation, based on text and illustrations, and digital documentation, including CAD models and image or video-based tutorials, present different criticalities when adapted to the Augmented Reality technology. Other visual material has been developed specifically to be used in Augmented Reality, but it is difficult to use because of a lack of standards which are instead present for the former paper-based and digital documentation. In this work, we categorize and compare different Augmented Reality visual assets, illustrating their advantages and disadvantages and providing directions for future research.

Keywords: Augmented Reality · Industry 4.0 · Graphical User Interface · Technical documentation

1 Introduction

Pushed by the fourth industrial revolution (Industry 4.0), Augmented Reality (AR) is a candidate to be implemented in industrial facilities for bridging virtual information and real environment [1]. AR is a valid technology for technical communication since it

allows the reduction of the cognitive load when performing procedural tasks as for maintenance or assembly procedures [2–5].

The number of AR applications in the industrial domain is continuously increasing in the last decades. Despite this, AR is still rarely used in real industrial procedures, but it often remains at a conceptual level in the laboratories [6]. One of the main reasons is the scarce collaboration between developers, academic researchers and the industrial world [7]. Therefore, the literature is full of various concept AR applications, but poor of guidelines about the use of AR techniques in the industrial domain. Among the open issues, there is the choice of the proper visualization methods to display technical information in AR.

Technical documentation evolved in the last years, starting from paper-based, characterized by text and illustrations, often black and white. Digital documentation, through the use of computer graphics, allows the use of new channels to convey information: interactivity of the contents, color, and animation. Examples of digital contents are CAD models and multimedia as image-based or video-based tutorials. Furthermore, AR allows the registration of graphical visual assets directly on the real world, with technical advantages as objects localization.

However, the authors of AR technical documentation need guidelines on how exploiting these novel technologies, considering both their opportunities and their limitations (e.g., occlusion). In fact, AR offers many options for the GUI designers, but currently, there are no specific standards to follow for industrial technical assets. Existing standards are generalist, like the Augmented Reality Markup Language (ARML) [8] and the Augmented Reality Application Format (ARAF) [9]. However, these standards can be used as a starting point for future implementations, for example for the nomenclature. In this work, we took inspiration from the Separation of Concerns principle, which is at the basis of ARML 2.0 [10], for the analysis of AR interfaces. According to this principle, an augmented scene is composed of three main components: (1) a real-world object (called feature), (2) its projected location on the augmented scene (called Anchor), (3) a virtual model associated to it (Visual Asset). The IEEE is developing a family of standards for virtual and augmented reality [11], that address aspects such as safety, how different technologies should be defined, and how virtual and real objects should work together.

Literature shows that the use of technical visual assets in existing industrial AR is varied and, in some cases, not optimized, because interface design is left to programmer discretion without a dedicated study. In many AR applications, there are no detailed descriptions of the interface, as well as motivations for the choice of the visualization methods. Thus, some visual assets are not used in an optimal way, while others are completely neglected.

According to the authors, there is a lack in the literature of a specific study about technical visual assets that can be used in an AR industrial interface. This study aims to orientate technical writers in the choice of visual assets, supporting them in the designing of future AR manuals.

In this work, we report a list of visual assets used in industrial AR applications (Fig. 1), deriving from a literature review on this topic [12]. We also provide a heuristic evaluation of visual assets, considering the most common issues that we experienced in these years during the development of test cases for industrial companies.

2 Visual Assets

2.1 Text

The traditional way to convey information is using text, since language, both written and oral, is the oldest and most solid way humans use to communicate technically. Text can provide any kind of information: through descriptions, it is possible to procure correct understanding and visualization of the problem, through commands it is possible to give instructions on how to proceed. Text instructions have been used since from the first paper manuals and are still used in AR applications. Authoring of text is very simple, but the need for translation of manuals in every language is a challenge for the use of text in the new AR interfaces.

We included in this category both 2D text and 3D text, as well as both texts, displayed with and without labels. In fact, the amount of information is not affected by the way the text is rendered.

2.2 Signs

We applied the definition of Peirce [13]: “a sign is a thing which serves to convey knowledge of some other thing, which it is said to stand for or represent.” Signs are of three classes: icons, indices, and symbols. An icon has a physical resemblance to the thing or concept being represented, like the camera icon in a smartphone, or gloves in an obligation symbol. A symbol may have no resemblance between the signifier and the signified, so the connection between them must be culturally learned: numbers and alphabets are good examples, or the shutdown symbol on a device, or the information symbol. An Index shows evidence of what is being represented, as a skull with two crossed bones is used to warn of toxic material.

Signs are regulated by standards that could be either International Standards, as the ISO 3864 for safety symbols, or internal practices. The information contained in signs is very focused, and this is what distinguishes signs by photographs, where an amount of information could be present.

2.3 Photographs

In this class, we put just photographs of real-world as acquired by a camera. Use of photographs is very common in manuals, especially digital manuals and instructional websites, as iFixit.com [14]. In fact, authoring is very easy since it requires just taking a picture of a real object.

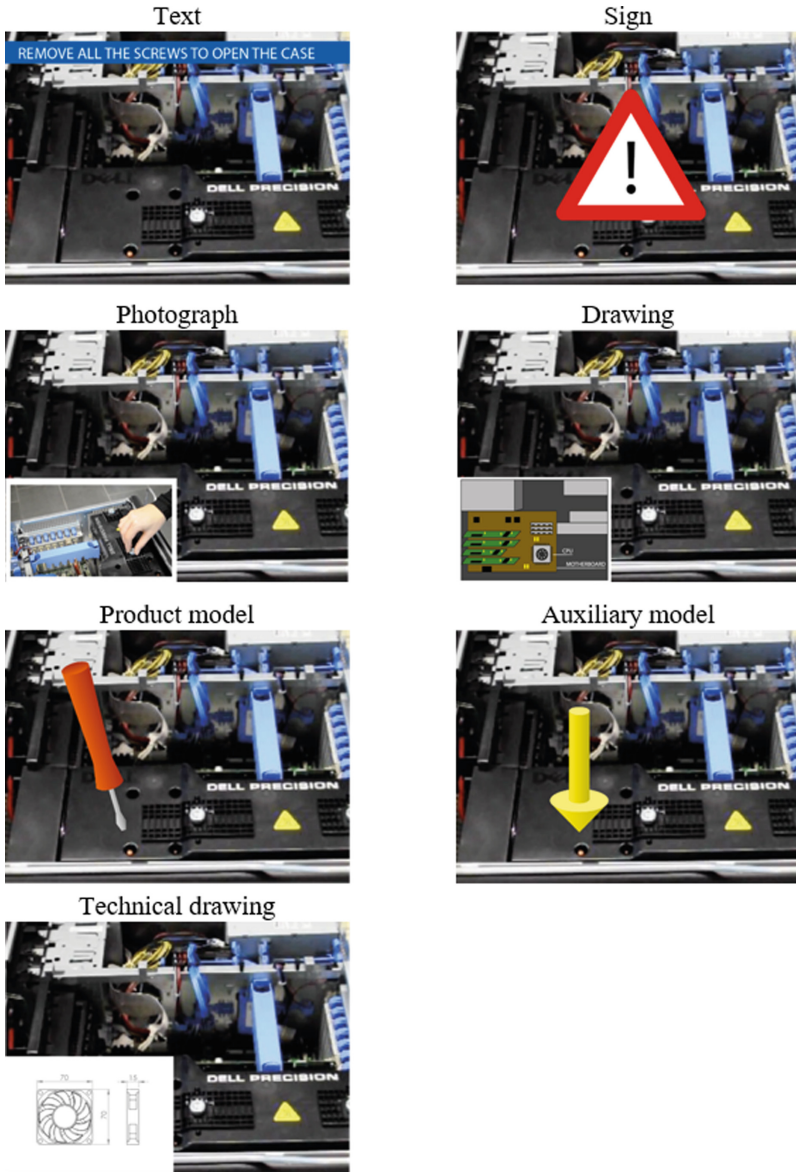


Fig. 1. Examples of using different visual assets in the same context.

2.4 Videos

In this class, we put just video recordings of real-world as acquired by a video camera or webcam. In the latest years, the number of video tutorials to convey technical information is widely increasing on the web since they are among the easiest and fastest way for “Do It Yourself” (DIY) authors to show technical instructions and for the end-users to learn them.

2.5 Drawings

In this class, we put digitized 2D drawings except for technical drawings on any support. Examples are freehand sketches, maps, charts. We also included in this group the annotated photographs, i.e. a combination of a photograph and 2D auxiliary models and/or text.

2.6 Technical Drawings

Technical drawings are an international standard to deliver constructive and functional information about products (ISO 128).

In this category, we included 2D representations in the form of technical drawings displayed as a static image on canvas, but also 3D graphical annotations according to ASME Y14.41 – 2003.

2.7 Product Models

We used the definition provided by Wang et al. [15]: “product models are 3D virtual models of product and parts”. In most of the industrial applications, product models are the digital representation of real objects modeled with CAD tools (e.g., machinery parts, components, tools). The information conveyed by product models has an explicit meaning since it does not change as the feature it is anchored changes. For example, the 3D animation of a screw that fits in a hole has the same meaning for every hole in every real component.

2.8 Auxiliary Models

We also used the definition provided by Wang et al. [15]: “auxiliary models are virtual models for auxiliary instructions”. The auxiliary models are 2D and 3D models, used by technical authors for delivering generic information to the operator. Some examples are arrows, circles, abstract sketches. Unlike the product models, the information conveyed by auxiliary models has not an explicit meaning, since it could change according to the feature it is anchored to: for example, an arrow can be used to locate objects in the real scene, to indicate a direction, to express a pushing action, and so on. We did not make a distinction between 2D and 3D elements since the same information can be conveyed by the 2D and 3D version of an auxiliary model.

3 Heuristic Evaluation of Visual Assets

We made heuristic evaluations to analyze the visual assets described in the previous section considering as heuristics the most common issues that we experienced in these years during the development of test cases for industrial companies [2, 5, 16]. A variable that could influence the choice of visual assets could be the display device used. However, we did not consider it in the heuristics list because our considerations are valid for all the devices commonly used in industrial AR (HWDs, handheld devices, desktop screen), except for Spatial AR that request specific attention.

- *Occlusion*. The use of AR to provide information for the operator implies that visual assets are merged with the real world. Then, visual assets can occlude a great portion of the real world. Occlusive visual assets may limit situation awareness and operator safety.
- *Registration*. As to registration, using the definition provided by Gabbard et al. [17], we can distinguish screen-fixed and world-fixed (or conformal) visual assets. The formers are rendered at a fixed location on the display and are generally not spatially anchored to any specific objects in the scene. These last are rendered such that they are perceived to exist at specific locations in the real world. Registration of world-fixed visual assets is affected by positioning and orientation errors due to tracking robustness. This misalignment influences the information correctness differently according to the visual asset used.
- *Communication time*. We defined it as the time that occurs between the moment since the visual asset appears until the moment the user comprehends the information associated. Different types of visual assets convey information in different ways then the communication time varies consequently. Some visual assets can provide a small piece of information in a more synthetic way, some others can provide more details about an operation. The information provided by a visual asset can be instantaneous or can be provided in a time cycle (e.g., an animation). In this second event, it is important that the user watches the augmentation for all the duration of the cycle. When it does not occur due, for example, to user distractions or tracking failures, the information provided by the visual asset can be compromised.
- *Authoring*. This is one of the most important issues for the development of AR interfaces. The creation of the virtual contents may require competencies that are not typically part of technical writers' background: 3D modeling, computer graphics/animation skills, programming.
- *Associated information*. The information provided by a visual asset can be either all included in the augmentation or in the combination of the augmentation with the real world. In this second case, the same visual asset can provide different information depending on the real object is associated with. If all the information relies on the visual asset, there could be more than one meaning conveyed by that visual asset.
- *Computational cost*. The hardware resources (e.g., CPU, GPU, RAM) employed during the rendering of the AR scene depend on the type of visual asset.

3.1 Text

Occlusion caused by text depends on the amount of text and visual style used. Descriptive instructions make use of many words and it is difficult to summarize them. The use of styles with large outlines and billboard is needed to improve the contrast between text and background and then legibility.

In most AR GUIs there is a dedicated area for text information, thus used screen-fixed. If text is world-fixed, information comprehension is not influenced by registration accuracy because a perfect matching between virtual text and the real scene is not needed.

The time spent by the user to acquire text information depends on text length. If some frames are lost due to tracking inaccuracy, users can miss some words, then communication time increases.

Authoring of text is very easy and does not request particular skills. Furthermore, text instructions can be imported by traditional manuals, if available in digital format. However, they need to be translated into all the languages requested by end-users.

A complete technical sentence must not have more than one meaning. Single words, often used in AR manuals, can lead to different interpretations and depend on the real object or event associated with (e.g., “right” could mean the direction of moving or a confirmation of a correct task execution).

Finally, rendering of text in AR interfaces require very low hardware resources.

3.2 Signs

Usually, a sign is monochromatic, and the background can be transparent. The dimension of the sign is the minimum to be distinguishable. Then, the occlusion caused by signs is very limited.

In an AR GUI, signs can be screen-fixed if coupled with other visual assets that indicate the real object the sign refers to. However, if they are world-fixed, they do not require a perfect registration with the real scene.

The memorization of the sign requires a little communication time, thus there is not the risk to miss information if some frames are lost.

The use of signs by AR designers is very simple since they would choose the desired one from a vocabulary of signs. Thus, authoring is limited to the definition of the vocabulary that should come from organizations for standardization. However, currently there are no standards for signs for all the industrial tasks (just for safety instructions), then the design of a vocabulary of signs is left to AR designers. Once that this vocabulary is defined, it can be used for all the AR industrial manuals designed.

The same sign could express different meanings: for example, arrows are used to indicate movements in a specific direction or forces to exert but are also used as pointers to focus the user’s attention on an object or area. Ambiguity is easily avoided using common standards that are still not available for all the industrial tasks.

Rendering of signs in AR interfaces requires very low hardware resources.

3.3 Photographs

Opaque photographs occlude a large portion of the real world, hiding what is behind them. Use of transparent or reduced-size photographs could reduce occlusion, but information comprehension could be compromised.

Photographs are preferably used screen-fixed because AR designers can control the GUI area occluded by a photograph. However, if they are world-fixed, they do not require a perfect registration with the real scene.

Communication time depends on the amount of informative details present in the photograph acquired.

Authoring of photographs is accomplished by taking pictures from the real scene. It requires time for authorizations needed, to prepare the scene, to acquire pictures, to post-process them. However, in an industrial scenario not always it is possible to acquire this multimedia due to safety problems, patents problems, difficult to access to the components, and so on.

If in the photograph more than one object is present, this can lead to information ambiguity. This situation is very common in industrial facilities.

Computational cost depends on the original photograph resolution that can be reduced for the AR application.

3.4 Videos

As to occlusion and tracking issues, we can make the same considerations made on photographs.

Communication time depends on the video duration. If some frames are lost, there could be some missing information for users.

Authoring of videos is similar to photographs. However, post-processing requires more effort than a photograph because of scenes editing.

Most of the videos used as visual assets show a preview of the operation to accomplish, made by another operator. Then, users must repeat what they watch in the video, thus there is a very low risk of interpretations different from the one intended by the AR designer.

Computational cost depends on the original video resolution, that can be reduced for the AR application, and on the duration.

3.5 Drawings

As to occlusion and tracking issues, we can make the same considerations made on photographs and videos.

Communication time is related to the quantity of information included in the drawing. However, since a drawing is specifically designed by AR designers, the information presentation could be optimized to reduce communication time.

Authoring of drawing depends on the type of drawing. Annotated photographs require an authoring effort similar to photographs with, in addition, the annotation. Sketches or other user-generated drawings require a higher authoring effort.

A drawing is specifically designed to provide objective information, then the meaning associated is unique.

Computational cost depends on the picture resolution that can be optimized for the AR application.

3.6 Technical Drawings

Differently from photographs and drawings, technical drawings cannot be made transparent and reduced in size because some lines can be missed. Thus, occlusion of the real scene is hard to reduce.

Similarly, to photographs and drawings, technical drawings are preferably used screen-fixed. If they are world-fixed, a perfect registration with the real scene is not required.

Communication time depends on the complexity of the represented object and the quantity of information included in the technical drawing: shapes, dimensions, assembly information, bill of materials, and so on. The comprehension of the information requires skills in the engineering field and the knowledge of technical drawing standards.

Authoring of technical drawings is related to the complexity of the represented object. It requires skills in the engineering field since international standards must be followed. However, the use of 3D CAD software, compared to software for 2D drafting, helps in the creation of drawings starting from a 3D model of the object, but post-processing of the drawing is needed: views management, adding/removing lines, quotation alignment, and so on.

By definition, a technical drawing is a representation of an object aimed at transmitting technical and functional information objectively. Thus, technical drawings convey only the information specified by the designer.

The computational cost of technical drawing is usually lower than the one of photographs.

3.7 Product Models

World-fixed product models are used to create the impression of a scene in which the virtual and the real coexist. They can replace (e.g., in assembly tasks) or overlap to a real object (e.g., for locating tasks). However, when operators accomplish the task, large product models can hide the real scene where they are operating. In this case, the use of transparent models can help to reduce the occlusion of the real world.

Product models are highly sensitive to registration accuracy. A misalignment between the augmentation and the real object it refers to causes low user acceptance and can lead to interpretation errors.

For static models, communication time depends on its complexity. If the product model is animated, it depends also on the animation duration. If some frames are lost, there could be some missing information for users, increasing communication time.

Authoring of product models requires competence in 3D modeling. Authoring effort is related to the complexity of geometry and animation.

Product models have a unique meaning due to their resemblance to real objects. The use of animations provides a preview of the operation to accomplish, further improving the uniqueness of meaning.

Rendering of product models requires a high computational cost, especially if animated. It can be reduced with the decimation of the model mesh.

3.8 Auxiliary Models

Auxiliary models fill a small portion of the GUI, thus there is a little occlusion of the real world.

Auxiliary models are little sensitive to registration accuracy. This sensibility depends on the model size. A small misalignment between the augmentation and the real object it refers to does not cause interpretation errors.

The association of the information to this visual asset is very immediate and exploits the affordance of real objects. There is no risk of missing information if some frames are lost.

The authoring of auxiliary models is very simple. Designers can choose the desired one from a list of existing models or also draw a new one.

Auxiliary models have not a specific meaning, but it strongly depends on the real object they are associated with. For this reason, the use of these visual assets in an AR GUI makes sense only if they are world-fixed and coupled with other visual assets.

Rendering of auxiliary models in AR interfaces requires very low hardware resources.

4 Results of Heuristic Evaluation

As we can see from the analysis provided in the previous section, the reported issues are critical with some visual assets, while they can be overcome using other visual assets. We made heuristic evaluations to assess the usability of visual assets in Industrial AR. The six authors were the evaluators. For each visual asset, they reported if it is a good choice or not in terms of the considered heuristic: occlusion, registration, communication time, authoring, associated information, computational cost. We synthesized these evaluations in Table 1.

Table 1. Synthesis of visual asset analysis based on the considered issues: red dots indicate bad choices, green dots indicate good choices, blank cells indicate situations where the choice depends on multiple factors.

	occlusion	registration	communication time	authoring	associated information	computational cost
Text	●	●			●	●
Signs	●	●	●	●	●	●
Photographs	●	●		●	●	
Videos	●	●		●	●	●
Drawings	●	●	●		●	
Technical drawings	●	●	●	●	●	●
Product models	●	●	●	●	●	●
Auxiliary models	●		●	●	●	●

It is not by chance that occlusion, one of the issues most specifically related to AR visualization, is not a critical point for signs and auxiliary models, which are often specifically designed for AR applications, while the other visual assets are mainly adapted from previous documentation. Many of the latter also present other different issues: in particular, authoring is critical for product models, technical drawings and videos, since the development and update of this kind of documentation requires a significant amount time and efforts. For product models, registration accuracy is fundamental, since misalignments can lead to interpretation problems and errors. This adds difficulty to the implementation of AR applications, also depending on the AR technologies and lighting conditions, which in industrial environments may not be optimal.

Technical drawings demand a high communication time, meaning that the operator consulting the information needs time and effort to elaborate it and understanding it. Since one of the main advantages of using AR is the immediacy and simplicity with which information can be found and displayed, this criticality is particularly relevant. The use of photographs without additional explanation can instead lead to interpretation issues and ambiguity, especially if complex or several objects are depicted. Issues regarding the understanding of the related information are also a strong concern for signs and auxiliary models. In fact, as stated above, those are often newly designed for AR applications. For this reason, they are not subject to industry regulations and standards, differently from the other visual aspects, which inclusion in technical documentation has a long tradition.

5 Conclusion

From the results of the heuristic evaluation, it is possible to make the following insights for Industrial AR designers.

Signs and auxiliary models are the only visual assets that present only one critical issue; hence they are particularly suitable for AR communication of information and instructions. The development and adoption of new standards for their use seem to be the most sensible path to open the way to the employment of AR solutions supporting industrial tasks and procedures.

However, this is not trivial: the creation of international industrial standards requires a complex validation process, which should be implemented for all the designed assets until an exhaustive amount is approved to be used for AR communication, as it happens for ISO graphical symbols. For this process to be optimized and effective, an intense collaboration is needed between industries, academic researchers and developers.

Also, the text presents only one critical issue, i.e. occlusion. However, the authoring of text instructions is highly dependent on the translation effort. Thus, research works on text optimization for translations as [18], are of valuable interest for the community.

References

1. Paelke, V.: Augmented reality in the smart factory: supporting workers in an industry 4.0 environment. In: 19th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA (2014)
2. Uva, A.E., Gattullo, M., Manghisi, V.M., Spagnulo, D., Cascella, G.L., Fiorentino, M.: Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. *Int. J. Adv. Manuf. Technol.* **94**, 509–521 (2018)
3. Henderson, S., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Trans. Vis. Comput. Graph.* **17**, 1355–1368 (2010)
4. Henderson, S.J., Feiner, S.: Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In: Proceedings of the IEEE 2009 International Symposium on Mixed and Augmented Reality, ISMAR 2009, pp. 135–144 (2009)
5. Fiorentino, M., Uva, A.E., Gattullo, M., Debernardis, S., Monno, G.: Augmented reality on large screen for interactive maintenance instructions. *Comput. Ind.* **65**, 270–278 (2014)
6. Bottani, E., Vignali, G.: Augmented reality technology in the manufacturing industry: a review of the last decade. *IIESE Trans.* **51**, 284–310 (2019)
7. Fite-Georgel, P.: Is there a reality in industrial augmented reality? In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, pp. 201–210 (2011)
8. Lechner, M., Tripp, M.: ARML—an augmented reality standard. *Coordinates* **13**, 432–440 (2010)
9. Preda, M.A.M., Choi, B.S.: Augmented reality application format (2012)
10. Lechner, M.: ARML-augmented reality markup language. *Ginzkeyplatz* **11**, 5020 (2010)
11. IEEE-SA- Augmented Reality. <https://standards.ieee.org/initiatives/ar/stds.html>
12. Gattullo, M., Evangelista, A., Uva, A.E., Fiorentino, M., Gabbard, J.L.: What, how, and why are visual assets used in industrial augmented reality? A systematic review and classification in maintenance, assembly, and training. *IEEE Trans. Vis. Comput. Graph.* **49**, 215 (2019)
13. Peirce, C.S.: *Collected Writings*, vol. 8, p. 58. Harvard University Press, Cambridge (1931)
14. iFixit: The Free Repair Manual. <https://www.ifixit.com>
15. Wang, J., Feng, Y., Zeng, C., Li, S.: An augmented reality based system for remote collaborative maintenance instruction of complex products. In: 2014 IEEE International Conference on Automation Science and Engineering (CASE), pp. 309–314 (2014)
16. Boccaccio, A., Cascella, G.L., Fiorentino, M., Gattullo, M., Manghisi, V.M., Monno, G., Uva, A.E.: Exploiting augmented reality to display technical information on Industry 4.0 P&ID. In: *Advances on Mechanics, Design Engineering and Manufacturing II*, pp. 282–291. Springer (2019)
17. Gabbard, J.L., Fitch, G.M., Kim, H.: Behind the Glass: Driver challenges and opportunities for AR automotive applications. *Proc. IEEE* **102**, 124–136 (2014)
18. Gattullo, M., Uva, A.E., Fiorentino, M., Scurati, G.W., Ferrise, F.: From paper manual to AR manual: do we still need text? *Procedia Manuf.* **11**, 1303–1310 (2017)