2022 European Conference on Computing in Construction
Ixia, Rhodes, Greece
July 24-26, 2022

AN ARTIFICIAL INTELLIGENCE AND MIXED REALITY APPROACH FOR OPTIMIZING THE BRIDGE INSPECTION WORKFLOW

Stephan Embers[1], Sven Zentgraf[1], Patrick Herbers[1], Firdes Celik[1], Benedikt Faltin[1], Markus König[1], Jan-Derrick Braun[2], Jessica Steinjan[2], David Schammler[2], Sonja Nieborowski[3] and Ralph Holst[3]
[1]Ruhr-University Bochum, Germany
[2]HOCHTIEF ViCon GmbH
[3]Bundesanstalt für Straßenwesen

Abstract

Detecting damage to bridges at an early stage is very important for financial and environmental reasons. This can only be achieved by regular and frequent inspections by experts, who mostly use paper based methods to document their findings. This paper aims to develop a concept for a bridge inspection tool using multiple types of hardware devices to support on-site bridge inspection personnel in assessing and documenting damages, employing combinations of both AI and MR technologies. Interviews were conducted with structural inspectors from various companies and from different sectors to identify important requirements. Based on these requirements, a concept has been developed, which is compatible with existing databases for infrastructure.

Introduction

Due to the ageing of structures, changes in traffic compositions, influences by climate change, and increasing traffic loads, measures for the maintenance of structures have to be carried out to a considerable extent. In Germany, these maintenance processes are defined in DIN 1076 (DIN 1076:1999-11 1999), with damage assessment defined in RI-EBW-PRÜF (Bundesministerium für Verkehr und digitale Infrastruktur 2017). When inspecting bridges and other civil engineering structures in accordance with these standards, it is essential that the damage found is identified completely and accurately, and that it is assessed in a reproducible and uniform manner. This is indispensable in particular for the assessment of damage progression and thus for the evaluation of the urgency of maintenance measures to be carried out. For this purpose, it is important that the bridge inspection personnel make optimal use of their experience and skills in damage detection, and assessment and are supported in tasks that can be performed more easily through the use of digital technologies. Thus, the introduction of IT-supported processes in the operation of the infrastructure has an important significance for reducing the required maintenance effort. Research is being carried out to identify potential for optimized life cycle management of bridges and to bring it into line with best practice. The rapidly developing technological progress is to be observed, accompanied and promoted by further investigations in the research project on behalf of the Federal Highway Research Institute.

According to the German standard for structural inspections DIN 1076 (DIN 1076:1999-11 1999), there are three phases to the bridge inspection process: Data preparation, on-site data gathering, and data processing. In this paper, we will focus on phase 2, the on-site data gathering. First, the on-site inspectors locate known damages using written documentation and search for possible new damages in the structure. The damages are then documented using digital cameras and categorized in written form. This data is then used to make decisions about the condition of the building and possible actions to be taken. This process still has potential to be optimized through the use of digital technologies.

Augmented and mixed reality (AR/MR) offer new possibilities for visualization of virtual content, information input and information provision. Virtual content can be provided to inspection workers, who thus have the possibility to include and process it in their decision making. Systems used for this purpose can record the real environment and situation via sensors, and cameras and overlay information appropriately. Artificial Intelligence (AI) technologies can be used to detect and analyze damage to structures by training image capture data models. In combination with AR, the AI can process the data recorded by the AR/MR system in real time, recognize patterns and provide the user with an evaluation via the AR/MR system. The application of a combination of Augmented or Mixed Reality and AI or Machine Learning can support the bridge inspection process as shown by previous studies (Salamak & Januszka 2018, Karaaslan et al. 2019, Moreu et al. 2019). These studies show how the use of one device, mostly Head-Mounted Displays (HMDs), can enhance the bridge inspection workflow. Most of the time, the used device is only utilized to display information without interaction, granting only limited options for the user to interact with the system, which represents a challenge when documenting more complex data, such as damage reports. Furthermore navigating large data structures or inspecting structural plans on HMDs can be difficult. Using devices that mitigate these drawbacks, such as tablets, could improve the workflows presented in these studies. DIN 1076 states that a test team of at least two persons has proven to be the most effective unit for a main inspection. This is widely accepted and practiced in the German bridge inspection workflow. Because of this, we propose a digitized workflow for bridge inspections using two different devices. Through the use of two devices, the workflow can be split up, reducing the workload for both inspectors and devices, as well as specialize the different devices to either information visualization or documentation.

Related Work

In order to support on-site bridge inspection workers, augmented reality (AR) for information visualization, and artificial intelligence (AI) for damage detection, will be used. AR is a human-machine interaction paradigm that overlays virtual information on top of a real environment (Dix et al. 2003). This allows the user to see the real environment with additional virtual objects. In general, AR should combine the real and virtual world and be operable in real time. In addition, objects in virtual and real space should have a 3-dimensional relationship to each other (Azuma 1997). The strengths of AR include the real-time interpreted information of the user's environment and the simple representation of it. In addition, by superimposing the digital information on the real environment, the user is not distracted from the real environment, while, for example, an instruction manual is shown on the display. Other strengths of AR include the paperless delivery of large amounts of information and the ability to connect additional devices to the system. Many of these strengths vary depending on the AR application area (Azuma 1997). One such application area is the on-site bridge inspection. To best utilize the strengths of AR, Moreu et al. (2019) conducted interviews with stakeholders to show the potential for detecting, analyzing and highlighting damages on a bridge for the on-site inspector. Salamak & Januszka (2018) introduced a concept on how to implement an AR bridge inspection tool using an HMD. The proposed system would enable the inspectors to capture videos and pictures of damages, and be guided via a visual overlay to points of interest. For the system to work, a digital model of the bridge would be required. Riedlinger et al. (2021) present another approach using a digital model to improve the on-site bridge inspection process using AR, while optimizing both the preparation and post-processing phases using virtual reality as well as supporting the collaboration in the bridge inspection process.

Since manual inspection at small time intervals cannot always be guaranteed, research has focused on computer vision-based automation of damage detection, which would allow not only more frequent, but also more time-efficient and cost-effective monitoring and inspection. However, image-based concrete damage detection is very challenging: images of bridges may contain inter-categorical and intra-categorical varieties in damage, noise elements such as graffiti, plants, insects, trash, discoloration, different lighting conditions, etc. There have been AI based damage detection methods in images, such as support vector machines (Liu et al. 2002), neural networks (Oullette et al. 2004, Rughooputh et al. 2000), and k-nearest neighbor (Kaseko et al. 1994). However, it was the publication of Krizhevsky et al. (2012) and the recent increase in GPU power that brought machine learning into the focus of image-based damage detection. In particular, it was the use of Convolutional Neural Networks (CNNs) that achieved breakthrough results in image recognition. Subsequently, CNNs have been used as the base architecture for most machine learning-based methods of damage detection. This development has been particularly favored by the introduction of advanced recognition methods. These are, in particular, bounding-box object detection methods (Girshick et al. 2014), semantic segmentation methods (Long et al. 2015), and instance segmentation methods (Hariharan et al. 2014). Karaaslan et al. (2019) present an AR system that implements an artificial intelligence (AI) and collaborates with a structure inspector to identify damage to the bridge structure. The proposed AR application is run on an HMD during the bridge inspection, on which the AI is also implemented. Among other things, the AI can detect cracks and spalling, and record the size of these damages. Detected damages are displayed to the structural inspector on the HMD by a virtual overlay over the real damage.

In order to locate and recall damages on a bridge model, a form of localization of the device and the recorded damage has to be performed. The localization problem is a term from robotics which describes the task of determining the pose of a device (position and rotation) from the environment. The pose can be related to a model or an already visited environment. This problem is also called "Kidnapped-Robot-Problem" in the literature, in which a robot has to find out its own position after waking up in a known environment. The best known localization method is GPS, where the position is determined with the help of a satellite network. The rotation is usually determined by a gyroscope and compass in such a system. While GPS can be used all over the world, its accuracy is limited to a few meters. Indoor occlusion, tunnels, or urban canyons can further reduce accuracy. Nevertheless, it can be used outdoors for visualizations with an AR device (Schall et al. 2009). Another method of localization is local signal stations, such as RFID or WiFi beacons. Based on the signal strength of the various beacons, the position of a device can be determined to within less than a meter (Chintalapudi et al. 2010). However, this method requires that the area is covered by a beacon network and the network has been calibrated accordingly. One way to perform localization without external sensors is geometric localization. AR devices can use spatial mapping techniques (structure from motion, depth cameras) to determine the geometry of their environment, and match this with existing geometries. This includes approaches such as point cloud matching (Bueno et al. 2018) or floor plan matching (Herbers & König 2019). These methods can determine very accurate poses, but are unreliable at locations with repetitive or self-similar geometry. In recent years, there has also been increased research on AI-based localization methods. Using CNNs, a network can learn a mapping between images and poses (Walch et al. 2017). Such approaches usually have an accuracy of one to three meters and are very performant in evaluation, but a significant amount of data is required for training.

Methods

To determine possible features and hardware for the proposed bridge inspection system, a requirements analysis was conducted. In the course of the requirements analysis, a detailed analysis of the bridge inspection workflow according to DIN 1076 (DIN 1076:1999-11 1999, DIN 1076 2013) was conducted, to identify crucial process steps within the workflow, which could potentially be optimized by using AR or AI. The procedure during the requirements analysis follows the principles of requirement engineering and takes into account requirements that result from the sequence of the structural inspection process.

Furthermore interviews with structural inspectors from various companies and from different areas were conducted. The interviewees were selected based on their experience in the inspection process. Overall, five persons were interviewed, three of which are structural inspectors specialized in bridge maintenance, and two of which are project managers for public contractors for infrastructure. The interviews were based on the aforementioned detailed analysis of the DIN 1076 norm and conducted in a uniform manner with each participant getting an overview of underlying project. Participants were only asked if a given process step would benefit from an AR or AI support system. Each individual process step was discussed together in order to identify exact points of contact for the AR application to be developed and the use of the supporting AI. Each interview was conducted over the course of one hour and were held online. The results of these interviews were incorporated into the formulation of the necessary requirements for the proposed AR/AI concept.

Additionally, a hardware analysis was conducted to identify possible combinations of different devices. In this analysis, devices will be evaluated according to their potential to visualize and process data, their ease of use, and other criteria to support an on-site inspector. The cost, availability, and battery run time of the devices are also taken in consideration. In order to specify the adapted AR-supported inspection process to be developed in this paper, these requirements are incorporated into the subsequent conceptual design proposal.

Results

The following section is separated into the results of the interview process, the results of the hardware analysis, and the proposed system.

Results Interviews

The detailed analysis of the individual process steps of the bridge inspection sequence according to DIN 1076 (DIN 1076:1999-11 1999, DIN 1076 2013) offers various starting points for a mobile AR application with integrated AI to support the inspection process. The interviews highlight an added value for the use of the new digital technologies seen in the reduction of effort, especially the manual effort in providing the documents on site. Automatic highlighting of damages during recording and a direct comparison to previous condition data is mentioned as an increase in efficiency. An evaluation recommendation as a guideline for the inspector's assessment of damages and the final structure condition grade could improve accuracy as well. Common rating levels are suggested based on the extent and size of the damage. The consolidation and evaluation of the recorded data for standardized documentation is also seen as promising. The localization of the damage on the structure plays a major role when documenting damages. The AI is to identify existing damage patterns and information, and visualize changes since the previous inspection in the AR application. While there were more potential points of interest within the inspection process, the above mentioned steps were identified as the most promising.

Results Hardware

Two different technical AR systems will be utilized in this paper: an AR tablet based on Google's ARCore, and a Trimble XR10 system with an attached Microsoft HoloLens 2. The Trimble XR10 system with HoloLens 2 has been specifically developed for use in areas with increased safety requirements such as construction sites, offshore facilities, or mining sites, and complies with the ANSI/ISEA industry standard. Multiple mobile AR devices based on different operating systems (Apple, Google, Microsoft) have been evaluated. Several concepts for configuring a two-device-approach have been examined: tablet/tablet, tablet/HMD, and HMD/HMD.

The tablet/tablet configuration is a low-cost and robust method, which is intended to replace analog work on paper with digital work on a tablet. Acceptance of and experience with the hardware is high among the population, which simplifies retraining. However, advanced AR immersion is not possible with this configuration.

This contrasts with the HMD/HMD configuration, where both inspectors use a head mounted AR device to perform the inspection. This combination would enable many innovative interactions, such as immersive AR and multi-user spaces. However, since AR-HMDs are a very new technology, the usage of such devices require extensive training, and user acceptance is likely lower. In addition, precise input patterns, such as typing or operating a map, are more complex using an HMD, and slower overall.

For these reasons, the Tablet/HMD concept was chosen, as it combines advantages of the above configurations while minimizing the disadvantages, combining a robust application with innovative methods. Data entry and lookup of old data is performed on the tablet, while data collection and visualization is done with the HMD. During development, emphasis is placed on keeping the coupling between devices as flexible as possible, so that the hardware concept can be easily expanded or changed in the future. For example, it should not be necessary to use an HMD for every inspection, but a single tablet should suffice. This would allow flexible use of the new technologies.

Proposed System

Based on the requirements analysis, the detailed specification of the functional units to be implemented for the Bridge Inspection Support System (BISS) for structural inspection in bridge construction is made. In the following, the intended process of the new system, the system architecture, the visualization of the data via AR and human-machine interaction with the BISS will be addressed. For a bridge inspection with the BISS, an adaptation to the conventional procedure is needed. For this purpose, a new bridge inspection procedure is designed that integrates the use of the BISS (Figure 1).

The adapted process starts in the data layer. This layer contains the asset database with information about the current status of the structure (e.g. bridge) being inspected, as well as information about past inspections and their recorded damages. To be compatible with existing databases, a second database for additional information collected with the new workflow is employed. The mesh database contains 3D data on recorded damage from past bridge inspections and their environmental data, later used for the relocalization of damages. The meshes are saved using the open glTF file format, while the relocalization data is saved in a proprietary file format used by the Microsoft HoloLens. From both databases, bridge data is imported from the data layer to an office desktop at the beginning of the bridge inspection process. The data imported from the databases by the desktop computer contains all data on the structure to be inspected, i.e. construction plans, documented damages from previous inspections, spatial meshes on damages if available, descriptions and assessments of the documented damages, and other data required for the inspection. The data inspector prepares the on-site inspection normally as described in the first phase of DIN 1076 (e.g. acquiring permits, setting inspection dates, etc.). The required data is then transferred to the tablet used for the bridge inspection with the BISS application. Once this has been done, the on-site bridge inspection can begin. During the on-site bridge inspection, the data inspector uses the BISS application on the tablet and the visual inspector uses the HMD. Here, the data inspector sends the relevant structure information to the visual inspector's HMD via the tablet. Using the received data from the tablet, the visual inspector can check the structure for damage. Several functions of the BISS on the HMD are available to the inspector to help them evaluate the detected damage. Documented damages of previous inspections can be visualized using the relocalization feature of the HoloLens. Newly detected damages can be recorded and categorized, including location and surrounding geometry. In future research, these damages can also be highlighted through an integrated AI, which will be able to mark and categorize damages automatically.

As seen in Figure 1, the data layer of the BISS consists of independent databases. The current condition data for bridge structures are stored in the asset database. The data required for a bridge inspection is retrieved via the REST interface provided by the asset database. For this purpose, the data inspector accesses the asset database via the Desktop component and can thus download and prepare the plans, the already documented damage images, and descriptions and assessments. The prepared data is then transferred to the ICDD component via a DIN SPEC 91391-2 compliant REST API and, if available, linked to the spatial meshes, anchor points, and damage masks, and stored in a container.

The Information Container for linked Document Delivery (ICDD) (Höltgen et al. 2021), published in the international standard ISO 21597-1, introduces a data structure that enables the linking of documents linked to a container, using semantic web technologies (EN ISO 21597-1 2020). The DIN SPEC 91391-2 specification (known as openCDE) can be used to enable the exchange and access to different types of information containers (DIN SPEC 91391-1:2019-04 2019). It defines requirements for the development of an open web API for the storage, management and distribution of these containers.

The bridge inspection data stored by the data inspector in the ICDD platform (spatial meshes of the damage, construction plans, damage from previous inspections, and descriptions and ratings) can be downloaded in the Tablet component via the openCDE API provided by the office application in the form of an ICDD container. During the bridge inspection, the damage recorded by the visual inspector via the HMD component is fed directly back to the data inspector in the Tablet component. The data inspector evaluates the transmitted damages and performs evaluations. Figure 2 shows the exact structure of the data inside the ICDD container.

After completion of a bridge inspection, the as-built data and the newly recorded damage and reports are again stored in an ICDD container and transferred back to the ICDD platform via the openCDE API. From here, the new data collected during the on-site inspection can then be fed into the asset database via the REST interface with the help of the Desktop component. The exchange of data within BISS is file-based. The linking of the used data is done by the semantic web technologies implemented in the ICDD platform. An overview for the proposed system architecture can be seen in Figure 3.

In order to be able to visualize the damage history in subsequent inspections, the captured information must be stored in a location-based manner. AR devices use so-called anchor points for this: points in AR space that are assigned to certain environmental features in real space. For example, an anchor point stores local image information (in the form of SIFT features), or the local geometry of the environment (in the form of point clouds or meshes). When entering a place that has been already visited in a previous inspection, the AR device compares the environment with the list of stored anchor points. If a match is found, the anchor point can be placed at the detected location, thus matching the AR space of the previous inspection and the real space of the current inspection again. We propose storing one

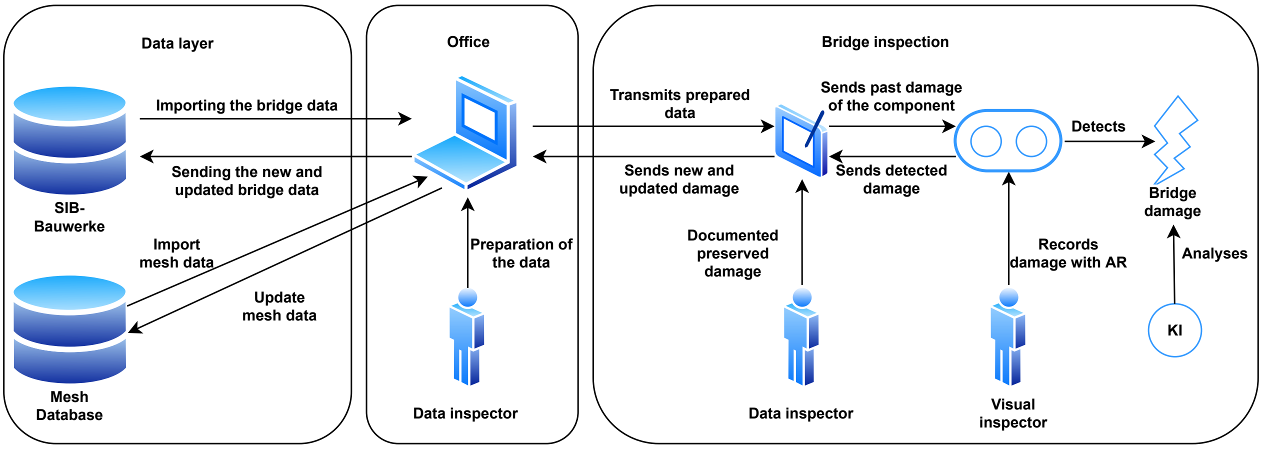

Figure 1: Overview of the adapted bridge inspection procedure for use of a two device AR and AI approach.

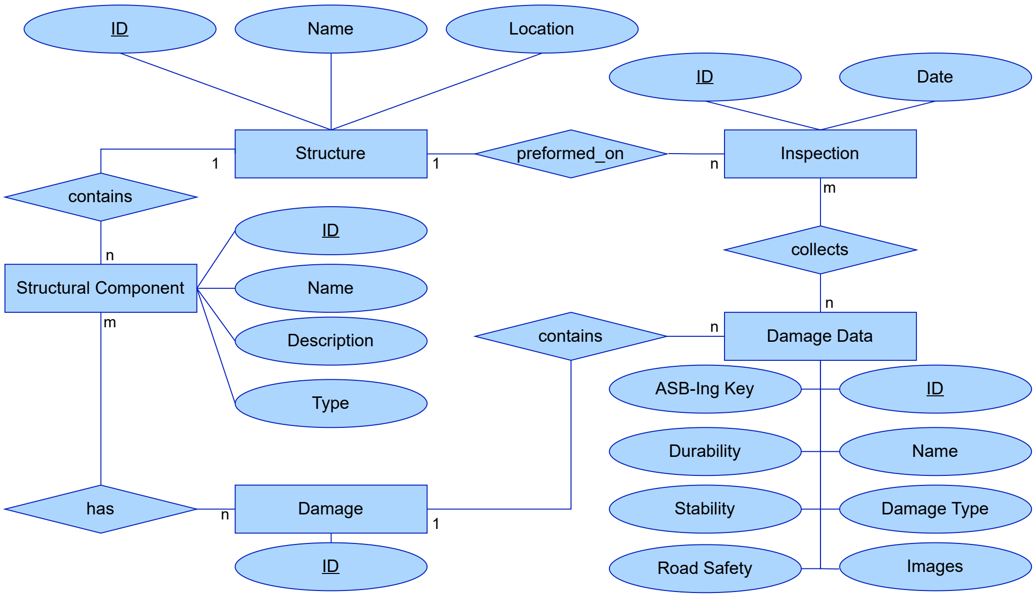

Figure 2: Structure of the ICDD container used in the associated linking of the transferred data.

anchor point should per damage. While it would be possible to use only a single anchor point from which all damage is positioned, the precision of localization decreases the further the device is away from this anchor point. Instead, an elastic network of anchor points is formed, which can alleviate the positioning errors between anchor points. Each recorded damage is assigned an anchor point, forming a node in the network. This network of anchor points is exported from the device at the end of an inspection and stored in the mesh database, where one anchor network is managed per bridge. This also allows the transfer of anchor points to other devices.

An important aspect in the development of new software, especially with newer technologies such as AR and the use of new hardware such as HMDs, is the ease of use and good user-friendliness of the new software. The aspects of ease of use and user friendliness also have a direct impact on user acceptance of the new software. For this reason, we developed a simple and intuitive human-machine interaction concept for the two systems used, the tablet and the HoloLens. Particular attention is given to the different interaction methods of the HoloLens and the user interfaces of both devices with which the user interacts.

To keep the use of the tablet as intuitive and simple as possible, common user interfaces for mobile applications are used. A distinction is made between the two frequently used mobile operating systems, iOS and Android. Both operating systems presuppose their own specifications or guidelines for the creation of mobile applications, which also applies to the user interface. iOS presupposes stricter specifications, so that a very similar user experience is granted across the entire spectrum of applications on iOS, which enables users to have a familiar control scheme over all available applications. In contrast to iOS, the Android operating system gives developers of a mobile application more freedom in the design of user interfaces. However, this has the disadvantage that a uniform implementation of user interface components is not enforced, which the user

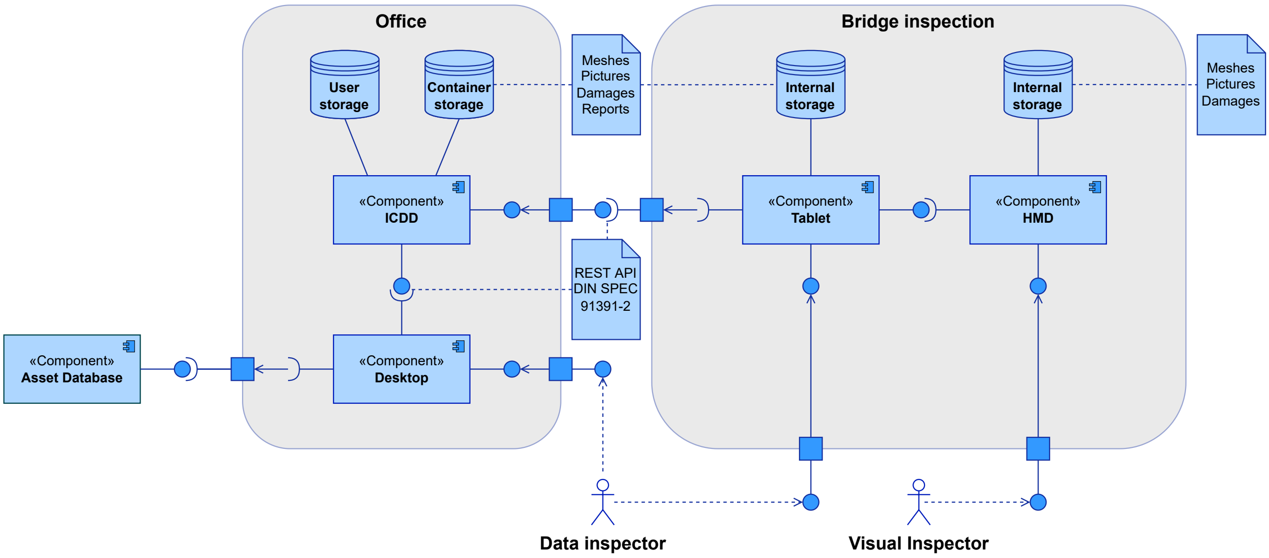

Figure 3: Overview of the system architecture for the bridge inspection support system.

would have to relearn when using a new application. However, there are certain user interface components that often perform the same task in Android applications. These are, for example, icons such as the "hamburger menu" or the "settings gear", which are familiar to users. Due to the additional design freedom for the operation, the wide availability, the easier communication and connection to other devices and the lower costs, the use of an Android device is employed for this concept.

The HoloLens offers several ways for a user to interact with it or its interface. The two commonly used interaction options are voice control and gesture control. With voice control, the user transmits commands to the HoloLens device via voice; usually this is initiated via predetermined keywords. With gesture control, the hands are used as an input method. Here, different hand gestures can be used to pass on commands to the HoloLens, for example via waving or pointing. A subcategory of gesture control is the use of a virtual touch interface, which provides virtual buttons and controls. Furthermore, it should be mentioned that the field of view of the HoloLens is smaller than the field of view of the user, which means that information is mostly displayed in the center of the user's field of view. Because of this, special attention should be paid when placing interface components, so as not to overwhelm the user or obscure his field of view.

Both interaction methods have their advantages and disadvantages. The control via voice commands has the advantage that the input of commands can be carried out completely hands-free, so that the hands can be used freely for other activities. Furthermore, the user interface can be designed more clearly and cleanly, since control elements within the interface can be omitted. This keeps the user's field of vision free of distractions. However, voice control requires a quiet environment and a clear voice. In environments with a lot of ambient noise, such as strong winds or car noises, the user may not be understood correctly or not at all by the speech recognition system, which can partially or completely limit the operation of the software. Even in optimal conditions, the voice input can be misunderstood, which in the worst case can lead to a work step being completely reset. This not only causes loss of working time, but also user dissatisfaction due to the repetition of voice commands. In addition to the problem of speech recognition, there is also the fact that the user needs to wait several seconds between individual commands for the speech recognition and the execution of the commands. Although this can be solved by stringing together several commands at once, the risk of misinterpretation by the speech recognition is higher.

In contrast to voice control, gesture control can be operated independently of ambient noise. The most common gestures in gesture control are waving, wiping, and pointing, but a variety of other gestures can also be programmed to be recognized by software. The gestures are mainly performed with one hand, which leaves the other hand free for other activities. Similar to voice control, the user interface on HoloLens can be kept free of additional controls, that would otherwise obscure the user's field of view. Gesture control also enjoys higher user acceptance because parts of gesture control are already established, for example swiping on a smartphone. Nevertheless, misinterpretation of the detected gestures can also occur with gesture control. This can happen due to poor visibility or poor execution of the gestures. Problems can also occur with complex instructions to the software, which require either very complex gestures, high precision, or a large sequence of gestures. Finally, a user would have to learn the usable gestures and their function, which is an additional barrier to entry.

An alternative to pure gesture control is the introduction of virtual input elements within the AR environment that can be used by pointing gestures or "pressing". This has the advantage that more complex commands or inputs are easier to execute, for example via the use of a virtual keyboard. This type of control also requires only one hand. In addition, there is no misinterpretation of speech or gesture recognition, and errors can be corrected more quickly. The operation of virtual input elements is very identical to the real counterpart, but haptic feedback is missing. This further reduces the barrier to entry for users. When using virtual input elements, special attention must be paid to their placement. The input elements must be visible to the user, but must not restrict the user's field of vision too much. This can happen quickly, especially with more complex applications. Input elements should always be in the user's field of vision, otherwise they have to be searched for by looking around, which takes additional time and can lead to frustration. To keep input elements in the user's field of view, they can follow the user's field of view so that the input elements are always displayed in the same place for the user.

The proposed concept consists of a combination of gesture control and virtual input elements. Through this combination, an error-free and simple operation can be ensured. In addition, the advantages of both methods can be further built upon. By using pure gesture control, the user's field of vision can be kept free during a building inspection, and additional information is only displayed when it is needed. Furthermore, the virtual input elements provide simple control elements for recording structural damage when they are needed. During the recording of structural damage, only the required information and input elements are displayed for a better overview. These are bound to the user's field of view so that the user does not lose sight of the input elements when looking around. As soon as the user has finished entering the required information, the input elements relevant for this input are hidden until they are needed again. This means that the user has the clearest possible field of vision at all times.

Some interactions, such as the selection of structural components or displaying structural drawings, would require an increased amount of display space, possibly obscuring the entire field of view of the user. Additionally navigating long list of structural components or structural drawings can be difficult. To alleviate this problem, these interaction are outsourced to the tablet of the data inspector for easier use.

Conclusion

This paper investigates how the existing process for bridge inspections is structured and identifies potentials for digitalization through computer-based processes and new technologies. Based on this, a digitalized two device process was developed for bridge inspectors to support inspection process in the assessment and documentation of damages, employing combinations of AI and MR technologies. Hardware requirements were investigated and appropriate technologies were defined. A concept was designed in the form of a tablet application on the Android operating system and an AR application on the Microsoft HoloLens. The proposed two-device configuration exploits the advantages of AR technologies (immersion, hands-free use), while mitigating disadvantages (complex operation, user acceptance). The system should allow damage to be digitally annotated and located. A damage history for each component can be viewed to monitor development and changes since previous damage inspections using the integrated AI. Spatial anchor points allow previously recorded damages to be easily retrieved and superimposed for comparison with the current state.

In the next step a demonstrator will be developed, incorporating the envisioned concept. Test will be conducted to verify the optimal method of communication between the two on-site devices. Emphasis will be placed on modularity, resilience, and user-friendliness of the system. The trained AI networks will be prepared for deployment and integrated on the devices. Furthermore, the process and the demonstrator will be evaluated by domain experts by means of an effort-benefit analysis.

Acknowledgments

This paper is based on parts of the research project carried out at the request of the Federal Ministry For Digital and Transport and its Network of Experts "Knowledge - Ability - Action", represented by the Federal Highway Research Institute, under research project FE-No. 69.0006/2020. The authors are solely responsible for the content. We would like to thank the members of our project's advisory board for their useful advice and feedback.

References

Azuma, R. T. (1997), 'A survey of augmented reality', *Presence: teleoperators & virtual environments* **6**(4), 355–385.

Bueno, M., Bosché, F., González-Jorge, H., Martínez-Sánchez, J. & Arias, P. (2018), '4-plane congruent sets for automatic registration of as-is 3d point clouds with 3d BIM models', *Automation in Construction* **89**, 120–134.

Bundesministerium für Verkehr und digitale Infrastruktur (2017), 'RI-EBW-PRÜF'.

Chintalapudi, K., Padmanabha Iyer, A. & Padmanabhan, V. N. (2010), Indoor localization without the pain, *in* 'Proceedings of the sixteenth annual international conference on Mobile computing and networking', MobiCom '10, Association for Computing Machinery, pp. 173–184.

DIN 1076 (2013), 'Bauwerksprüfung nach DIN 1076. bedeutung, organisation, kosten'.

DIN 1076:1999-11 (1999), 'Ingenieurbauwerke im Zuge von Strassen und Wegen - Überwachung und Prüfung'.

DIN SPEC 91391-1:2019-04 (2019), 'Gemeinsame datenumgebungen (CDE) für BIM-projekte - funktionen und offener datenaustausch zwischen plattformen unterschiedlicher hersteller - teil 1: Module und funktionen einer gemeinsamen datenumgebung'.

Dix, A., Finlay, J., Abowd, G. D. & Beale, R. (2003), *Human-computer interaction*, Pearson Education.

EN ISO 21597-1 (2020), 'Information container for linked document delivery - exchange specification - part 1: Container'.

Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2014), Rich feature hierarchies for accurate object detection and semantic segmentation, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)'.

Hariharan, B., Arbeláez, P., Girshick, R. & Malik, J. (2014), Simultaneous detection and segmentation, *in* 'Computer Vision – ECCV 2014', Lecture Notes in Computer Science, Springer International Publishing, pp. 297–312.

Herbers, P. & König, M. (2019), 'Indoor localization for augmented reality devices using BIM, point clouds, and template matching', *Applied Sciences* **9**(20), 4260. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

Höltgen, L., Cleve, F. & Hagedorn, P. (2021), Implementation of an open web interface for the container-based exchange of linked building data, *in* '32. Forum Bauinformatik 2021', p. 9.

Karaaslan, E., Bagci, U. & Catbas, F. N. (2019), 'Artificial intelligence assisted infrastructure assessment using mixed reality systems', *Transportation Research Record: Journal of the Transportation Research Board* **2673**(12), 413–424.

Kaseko, M. S., Lo, Z.-P. & Ritchie, S. G. (1994), 'Comparison of traditional and neural classifiers for pavement-crack detection', *Journal of Transportation Engineering* **120**(4), 552–569. Publisher: American Society of Civil Engineers.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012), 'ImageNet classification with deep convolutional neural networks', *Communications of the ACM* **60**(6), 84–90.

Liu, Z., Suandi, S. A., Ohashi, T. & Ejima, T. (2002), Tunnel crack detection and classification system based on image processing, *in* 'Machine Vision Applications in Industrial Inspection X', Vol. 4664, SPIE, pp. 145–152.

Long, J., Shelhamer, E. & Darrell, T. (2015), Fully convolutional networks for semantic segmentation, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', pp. 3431–3440.

Moreu, F., Lippitt, C., Maharjan, D., Aguero, M. & Yuan, X. (2019), 'Augmented reality enhancing the inspections of transportation infrastructure: Research, education, and industry implementation'.

Oullette, R., Browne, M. & Hirasawa, K. (2004), Genetic algorithm optimization of a convolutional neural network for autonomous crack detection, *in* 'Proceedings of the 2004 Congress on Evolutionary Computation', Vol. 1, pp. 516–521.

Riedlinger, U., Oppermann, L., Neumann, S., Holst, R., Hill, M., Bahlau, S., Klein, F. & Mertens, M. (2021), Supporting bridge inspectors with interactive mixed reality visualizations of bim process and geometry data, *in* 'Proceedings of the International Bridge Conference 2021'.

Rughooputh, H. C. S., Rughooputh, S. D. D. V. & Kinser, J. M. (2000), Automatic inspection of road surfaces, *in* 'Machine Vision Applications in Industrial Inspection VIII', Vol. 3966, SPIE, pp. 349–356.

Salamak, M. & Januszka, M. (2018), BrIM bridge inspections in the context of industry 4.0 trends, *in* 'Maintenance, Safety, Risk, Management and Life-Cycle Performance of Bridges', CRC Press, pp. 2260–2267.

Schall, G., Mendez, E., Kruijff, E., Veas, E., Junghanns, S., Reitinger, B. & Schmalstieg, D. (2009), 'Handheld augmented reality for underground infrastructure visualization', *Personal and Ubiquitous Computing* **13**(4), 281–291.

Walch, F., Hazirbas, C., Leal-Taixe, L., Sattler, T., Hilsenbeck, S. & Cremers, D. (2017), Image-based localization using LSTMs for structured feature correlation, *in* 'Proceedings of the IEEE International Conference on Computer Vision', pp. 627–637.