Augmented Reality for Construction Site Monitoring and Documentation

Stefanie Zollmann, *Member, IEEE,* Christof Hoppe, Stefan Kluckner, *Member, IEEE,* Christian Poglitsch
Horst Bischof, *Member, IEEE,* and Gerhard Reitmayr, *Member, IEEE*

***Abstract*—Augmented Reality allows for an on-site presentation of information that is registered to the physical environment. Applications from civil engineering, which require users to process complex information, are among those which can benefit particularly highly from such a presentation. In this paper, we will describe how to use Augmented Reality (AR) to support monitoring and documentation of construction site progress. For these tasks, the staff responsible usually requires fast and comprehensible access to progress information to enable comparison to the as-built status as well as to as-planned data. Instead of tediously searching and mapping related information to the actual construction site environment, our AR system allows for the access of information right where it is needed. This is achieved by superimposing progress as well as as-planned information onto the user's view of the physical environment. For this purpose, we present an approach that uses aerial 3D reconstruction to automatically capture progress information and a mobile AR client for on-site visualization. Within this paper, we will describe in greater detail how to capture 3D, how to register the AR system within the physical outdoor environment, how to visualize progress information in a comprehensible way in an AR overlay and how to interact with this kind of information. By implementing such an AR system, we are able to provide an overview about the possibilities and future applications of AR in the construction industry.**

***Index Terms*—Augmented Reality, Visualization, Aerial 3D reconstruction, Civil Engineering**

I. Introduction

Augmented Reality is an interface that overlays digital information onto the user's view, spatially aligned to the current physical environment [1]. The user's view is often a camera image of their physical surroundings. The video image is augmented with digital information and rendered on the display device, which may be a head-worn display or a mobile device. Such an overlay allows for the presentation of information that is relevant to a specific task right on-site and aligned to the objects of interest.

This type of information presentation is relevant for various professional tasks. In fact, AR has been identified to be successful for several industrial applications. In particular, civil engineering applications can benefit from such a registered overlay. For instance, in the Smartvidente project, subsurface infrastructure has been overlaid on top of street scenes in order to allow the user to access information about theses structures while at the site [2]. Professional tasks within the construction industries also often require access to information right on-site. For instance, Woodward et al. showed how to overlay construction plans over a user's view of a construction site [3]. This allows the users to inspect these plans within the context of the physical world environment. Similar techniques have been used in indoor environments to inspect the as-built state of a factory in comparison to the as-planned state [4], [5]. However, AR does not only allow for the visualization of as-planned structures, it also provides tools for visualizing the progress of construction sites if this data is available.

A. Motivation

From discussions with partner companies, we learned that automated documentation and monitoring is an important topic for the construction industry. It can support the supervision of contractors' achievements, as well as the detection of schedule derivations or the search for sources of defects and the workers responsible. The last point is particularly interesting for compensation requests and responsibilities. Adequate progress monitoring methods help supervisors and workers to document the current status of the construction work as well as to understand origins of defects.

Nowadays, construction site staff often uses digital photography to document the progress of a construction site. Typically, staff members take individual photos of the construction site on a regular basis and store them together with the construction site plans in a database (Figure 1). This enables the supervisors to relate possible errors or bottlenecks to certain dates and identify the workers responsible. The disadvantage of this approach is that a staff member has to take the photographs manually. This is very time-consuming and leads to areas not being covered very well. Another aspect is that neither the relation between acquired photographs, nor the relationship to the physical construction site is available. This creates a high mental workload for the supervisor when mapping the photographs to the actual construction site.

Another technique that is sometimes applied for construction site monitoring is the use of laser scanners. In this case, 3D information can be created and stored for later inspections. But there are two main problems with this technique 1) these devices are bulky and have to be carried around, which is a problem on rough terrain such as is found on a construction site. 2) Inspecting the resulting 3D data is still challenging, since the spatial relationship to the actual environment is missing.

S. Zollmann, C. Hoppe, C. Poglitsch, H. Bischof and G.Reitmayr are with the Institute of Computer Graphic and Vision, Graz University of Technology, Graz, 8010 Austria, e-mail: zollmann@icg.tugraz.at

S. Kluckner is with Siemens AG Austria, Corporate Technology, Imaging and Computer Vision, Video Analytics.

Manuscript received July 15, 2013;

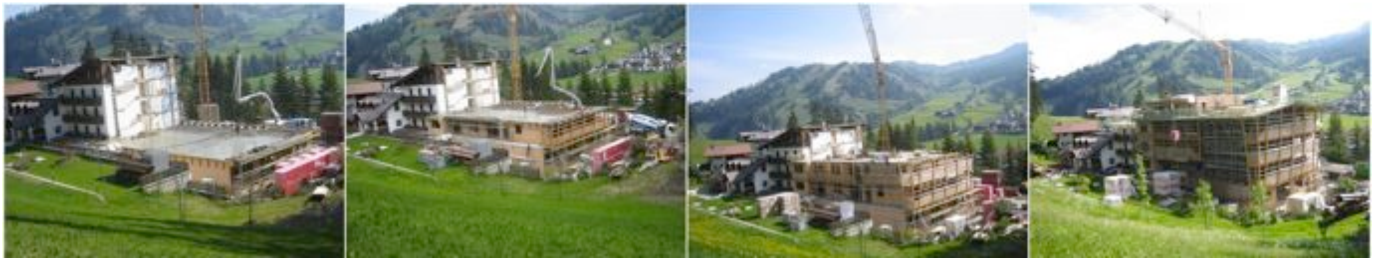

Fig. 1. Construction site monitoring using camera images. The image sequence represents different steps of the construction progress. This information can be difficult to analyze and search for sources of mistakes. Furthermore, it is tedious for the user to map the 2D camera image information into the actual physical 3D environment.

AR in combination with aerial 3D reconstruction, provides a solution for these problems and allows for on-site construction site monitoring by capturing data of interest and providing the required information in direct relationship to the physical construction site. The idea behind our approach is to create 3D information literally on-the-fly by using aerial vehicles and providing a mobile device that allows one to inspect this progress information directly on-site. For this purpose, we superimpose progress information in an AR overlay on a mobile setup that is registered to the actual physical environment.

AR visualization techniques have already been applied by other research groups in construction environments, for instance for displaying construction plans on-site [6], [3]. Golparvar-Fard et al. discussed the use of AR visualization for supervising the progress on construction sites within the scope of the D4AR project [7]. Their system computes progress information that represents the status of the construction site, such as the current level of completion. The corresponding value is visualized by color coding the 3D physical world object. While this approach allows one to study differences between planned data and the current physical world situation, our approach aims at providing an overview as well as detailed information about the progress.

B. Contribution

In this article, we propose a system that allows for the visualization of progress information of a construction site directly on-site using AR. For this purpose, 1) we developed a system consisting of three main components: an *Aerial Client* that captures aerial images for 3D reconstruction on regular basis, a *Reconstruction Client* that performs aerial reconstructions and remote localization and finally an *AR client* that visualizes progress information spatially registered to the physical environment on-site. Each component is able to communicate with the others and to exchange data over a network.

In order to be able to overlay progress information correctly aligned to the physical world, we need accurate registration techniques that tell the AR client its pose in relation to the surrounding environment. For this purpose, we 2) introduce a set of registration techniques that differ in the amount of sensors they are using. We will discuss sensor-based registration techniques that incorporate inertial sensors, camera and GPS, and remote localization techniques that only make use of a camera, but depend on a 3D point cloud.

Besides the accurate registration of data, it is important to visualize the data in a comprehensible way. Since the simple overlay of information can be problematic, we propose 3) a set of visualization techniques that generate comprehensible overlays of progress information. Within the visualization section of this article, we will describe methods for visualizing single as well as multiple points in time overlaid onto the physical world environment. For visualizing selected points in time, we show methods for overlaying 3D mesh information as well as describing techniques for creating abstracted representation that are easier to comprehend. Furthermore, we will describe filtering methods that allow for the visualization of multiple points in time in one view.

Several tasks in construction site monitoring and documentation also require interactive input, for instance in order to leave comments about specific objects. AR allows to store such annotations directly in relationship to the actual environment. For this purpose, we will show tools for annotating the physical environment based on the 3D information available from the 3D reconstruction. Furthermore, the availability of 3D information about the physical environment allows interactive surveying tasks to be performed on the construction site.

II. Background

The recent developments in mobile technology allow advanced interfaces such as AR to be moved from the laboratory into the field. Their ability to present information on-site, where it is needed, holds many advantages for professional applications. In particular, civil engineering tasks that require field workers to work with complex information, such as information from GIS databases and maps, can greatly benefit from these systems [8]. The potential of AR for the architecture, construction and engineering (ACE) industries has already been identified by Shin et al. in 2008 [9].

Application scenarios have, for instance, been shown by King et al. who demonstrated ways in which AR can be used to visualize GIS information for the viticultural industries [10]. They overlaid GIS data, such as harvest yield monitoring values onto a video image on a mobile AR system. In order to register the geo-referenced data to the video image, the authors equipped a laptop with a set of sensors, such as GPS and orientation sensors. A similar approach was used within the Vidente project [8], where Schall et al. introduced a system for visualizing underground infrastructure information, such

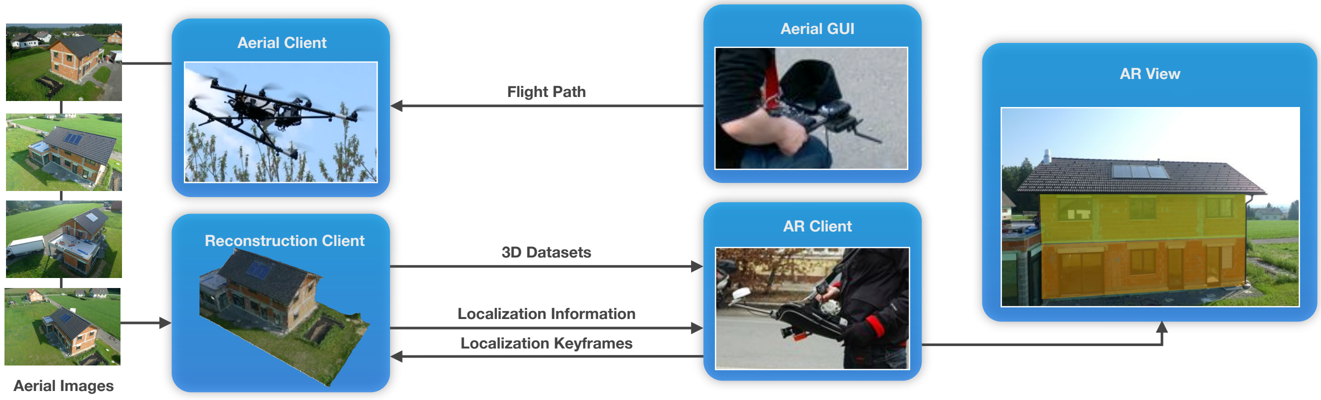

Fig. 2. Mobile AR system for construction site monitoring and documentation. Left) *Aerial Client* captures a set of *Aerial Images* of an area of interest. It publishes the images over the network, and the *Reconstruction Client* uses them to create a 3D representation of the construction site. Right) After computing the 3D information, the *Reconstruction Client* publishes the 3D data over the network as well. An *AR Client* that is spatially registered using external sensor or remote localization accesses the 3D data and renders it in the AR view.

as subsurface electricity lines and gas pipes. Schall et al. increased the mobility of their system by replacing the laptop with an ultra-mobile PC. They equipped the ultra-mobile PC with a set of registration sensors and used it to display GIS data from civil engineering companies. The work of Vidente was continued within the Smartvidente project [2] focusing mainly on improving the registration accuracy and visualization techniques for subsurface visualization. The latest developments in Smartvidente already provided a registration accuracy in the range of 10 cm for positioning and within a sub-angle range for the orientation. This project already showed the high potential of a highly accurately registered AR system for professional industrial tasks.

Other research groups showed that there is also a high potential for the use of Augmented and Mixed Reality in supporting applications within the construction industries. For instance, Woodward et al. proposed a system for visualizing and verifying as-planned information [3]. They integrated a set of different datasources into their system, such as map data, Building Information Models (BIM) and 4D BIM. The registration of this system is based on a standard GPS (accuracy 10-50m) and orientation measurements given by model-based initialization, feature-based tracking and compass information. In order to compensate for inaccurate registration methods, they provide interactive tools for manual positioning and alignment corrections. The idea of visualizing planned data for construction management has already been adopted by companies. Researchers at Bentley for instance showed how to integrate 2D map data into the physical environment of a construction site [11].

In contrast to visualizing as-planned information, Golparvar-Fard et al. presented techniques for visualizing progress information [7], [12] using Mixed Reality. For this purpose, they take time-lapse photographs and perform 3D reconstruction based on this data. By overlaying the 3D information onto registered camera images, the progress can be visualized in relation to the physical environment in order to highlight problematic areas that are behind schedule.

While this body of existing work showed how to either visualize as-planned data on-site or how to visualize progress information using Mixed Reality overlays in desktop environments, in this article we will present a system for capturing and presenting progress information on-site with a highly accurate AR overlay.

III. System

In order to visualize progress information of a construction site on-site with a mobile AR system, we need access to data that represents the construction progress, for instance in the form of multiple 3D datasets from different points in time. For this purpose, we use an *Aerial Client* that flies over the area of interest on a regular basis. During such flight sessions the aerial vehicle captures a set of aerial images from meaningful positions. This set of camera images is then sent to the *Reconstruction Client* which creates a 3D representation based on the images. If such a reconstruction is done on a regular basis, we can create time-oriented data or so-called 4D datasets. Finally, in order to visualize this 4D data on-site, we use a mobile AR client that is capable of exchanging and displaying the progress information on-site and that is spatially registered to the physical world and the geo-referenced data.

For a high level of flexibility and reliable communication between the clients, we use the Robot Operating System (ROS[1]) for exchanging data (Figure 2) over the network. Each client has to registered itself within the system to be able to receive the relevant data. In the following, we will describe each of these components in greater detail.

A. Aerial Client

We need to capture a large amount of meaningful progress data with a low manual input. For this purpose, we use an aerial vehicle in combination with automatic flight path planning. While manned aerial vision was traditionally only used for capturing large scale areas such as the digital surface models (DSM) for complete urban environments [13], smaller

[1] http://www.ros.org

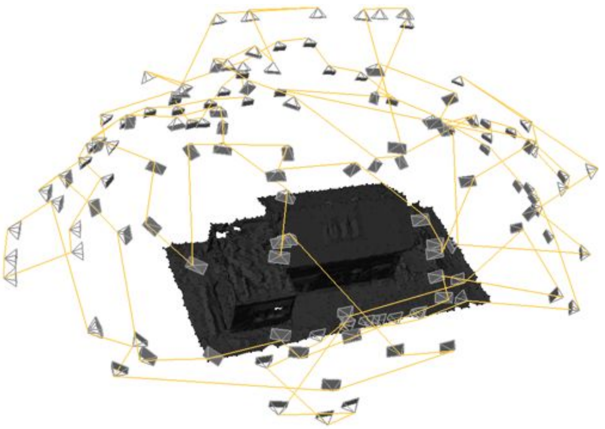

Fig. 3. Automatically calculated camera network (gray) with corresponding flight path (yellow).

and cheaper vehicles, such as micro aerial vehicles (MAV) can be used to reconstruct smaller areas and capture also detailed views. This makes them of interest for capturing the progress of construction sites. For our system, we equipped an AscTec Faclon 8 Octocopter with a standard consumer digital camera. Such a setup is able to automatically capture images of the area of interest from predefined positions [14].

The limited battery power supply of a MAV restricts the flight time to 10 to 20 minutes. After this time the batteries has to be replaced or has to be recharged with average recharging times of 45 minutes. Due to this restriction it is highly important to apply a sophisticated flight management methods that ensure that enough meaningful views are captured before the vehicle runs out of battery power. In order to address this problem, we developed automatic methods for flight path planning. These methods are usually known as Photogrammetric Network Design (PND) and support the capturing process by calculating the best camera poses and the optimal flight path in advance.

In order to create a good set of meaningful images, we set the following requirements for the automatic flight management that guarantee a good reconstruction of the area of interest:

- Minimal number of views.
- Cover the complete area of interest.
- Redundancy between the images.
- Overlap between images
- Adequate viewing angles between views.
- Fully connected image overlap graph.

An additional requirement that is not listed is a minimized path length. For our setup, we decided to not use this requirement, since the flight time of the MAV moving with 10m/s is rather insignificant compared to time the device needs to swing into one position. To meet all above listed requirements within one flight session, we developed a new method for flight management. The approach requires a rough description of the area of interest, for instance in the form of a GIS model, BIM data or a previous rough reconstruction. Based on this rough description, we use a viewpoint sampling around the area of interest and render views from these viewpoints. Afterwards, viewpoints are randomly selected from this set until it is possible to perform a complete reconstruction (Figure 3). We described this method in greater detail in [15].

B. Reconstruction Client

During one flight session the *Aerial Client* is able to create image sets of approximately 200-300 high-resolution and highly overlapping images. These images are published over the network and used by the *Reconstruction Client* as input for 3D reconstruction. The reconstruction is based on Structure-from-Motion (SfM) similar to the methods described by Irschara et al. [16]. SfM uses camera images from different viewpoints of the same area of interest to compute 3D information. For this purpose, SIFT features are extracted from the camera images and matched between the different images. The matches are used to compute the relative camera poses. The known poses and the correspondences in combination are then used to calculate a 3D point cloud.

Additionally, we include available GPS information from the MAV into the reconstruction workflow to reduce computation time and to obtain a geo-referenced 3D model at a metric scale as proposed by Irschara et al. [17].

The output of the reconstruction is 3D scene geometry represented as a sparse 3D point cloud (Figure 4). This 3D reconstruction data can now be distributed over the network and used for visualization. Nevertheless, the sparse geometry often does not include enough information to comprehensible reveal the progress to the user. For providing a meaningful representation, the reconstruction client uses the 3D point cloud data as input for point cloud densification and mesh creation (Section IV). By flying over a construction site on a regular basis and creating and storing 3D representations of different points in time, we are able to compute a 4D dataset of an area of interest. This dataset can then be visualized using a mobile AR client on-site.

C. Mobile AR Client

For overlaying the captured progress information onto the user's view of the physical outdoor environment, we need a mobile AR client that is appropriate for working in outdoor

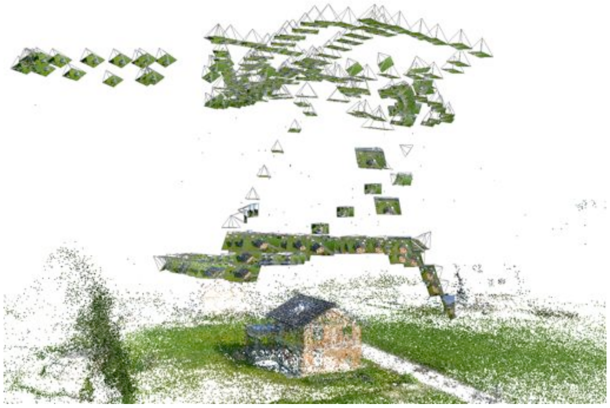

Fig. 4. Sparse point cloud and corresponding camera images.

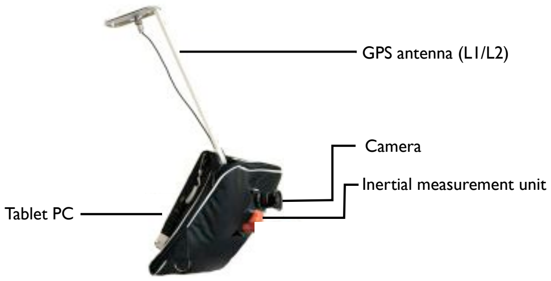

Fig. 5. Augmented Reality Setup equipped with external sensors.

environments and integrates all sensors that are required for video capturing and for achieving an adequate registration.

For addressing the particular needs of workers in construction sites environments, we use a ruggedized tablet PC (Motion J3400, 1.6GHz). This tablet PC provides a screen that is specially built to be viewable outdoors, even under sunlight conditions. We equipped the tablet PC with a camera (VRMagic VRM FC-6 COLOR) and a wide-angle lens for capturing video images for the AR overlay and feature-based tracking. Furthermore, we integrated a set of external sensors that support outdoor registration if no 3D information about the current scene is available (Figure 5). The set of additional sensors consists of an Inertial Measurement Unit (IMU, XSense MTx 3DOF), and a GPS receiver. The IMU is fixed to the camera and is calibrated to the camera coordinate system. Both, camera and IMU are mounted on the back of the tablet, pointing away from the user. The IMU consists of gyroscopes, accelerometers and 3D magnetometers and provides 3DOF orientation measurements. As a GPS sensor, we use a L1/L2 RTK receiver that measures the device's position within centimeter accuracy (Novatel OEMV-2 L1/L2 Real-Time Kinematic). All sensors are connected via USB and synchronized in software. A leather bag covers all cables and the GPS sensor, to protect them from the weather (Figure 5). The system itself can be carried using a shoulder strap that is connected to the bag. Furthermore, it can be used in a fixed position by mounting it to a tripod for hand-free operations.

During runtime, the sensors of the setup work in separate threads and feed the registration methods (Section V). The output of the registration is used to update the transformation matrix of the virtual content. The rendering itself is performed each time a new camera image arrives.

IV. Data Postprocessing

We can use the 3D data that is captured by the *Aerial Client* directly for visualization purpose, since it is registered in 3D to the physical world. Nevertheless for some applications more detailed data, such as a 3D mesh or as-planned data is required. In this section, we describe methods for post-processing the 3D data created by the *Reconstruction Client*. Furthermore, we will show how we can apply external datasources for comparing as-built with as-planned data.

A. 3D Densification and Mesh Generation

Sparse geometry often does not contain enough information to visualize the progress. For computing more dense information, we use the approach of Furukawa et al. that calculates a semi-dense point cloud [18]. In addition to the dense point cloud for several applications, a 3D mesh can be required. To provide 3D mesh data, we also compute a 3D mesh by using Poisson surface reconstruction [19]. Unfortunately, this method is computationally expensive. In order to provide online feedback, we developed a new method for computing 3D mesh representations online as soon as new images from the *Aerial Client* arrive. The basic idea is similar to the method presented by Labatut et al. combining a Delaunay triangulation with global optimization methods [20]. In [21], we discuss this approach in greater detail and show how to use it to provide online feedback for SfM image acquisition.

This reconstruction workflow creates not only single as-built dataset of an area of interest, it also allows us to create 4D as-built data. A 4D dataset represents the as-built status over a selected period of time with a predefined sample rate. When creating the 4D datasets it is important that the datasets are accurately aligned with each other. While the geo-referencing already allows for the course alignment of these datasets, a matching procedure on the sparse 3D points (resulting from the SfM) allows for fine alignment between the different datasets.

Since coarse registration is already available, the matching can be applied to smaller regions. Robust inlier estimation of these smaller regions via RANSAC allows us to determine a global rigid transformation computed from a final set of correspondences. The final step includes a fine ICP based refinement by incorporating the semi-dense point cloud. The obtained mesh is then transformed according to the aggregated transformations and registered to a canonical reference frame. We call such a dataset a time-oriented or 4D representation of a construction site.

B. External Data Sources

When inspecting the progress of a construction site, it is often helpful to not only visualize the as-built status as is provided by the 3D reconstruction, but also to render information about existing surrounding structures as given by GIS data and to compare the current or previous situations with information provided by BIM. BIM describe systems that combine different information about the life cycle of a building such as construction plans, but also plans for building management. Information from BIM provides geometric as well as semantic information about the physical world context similar to information from GIS. Thus, it can be a helpful source of information for on-site progress visualization. While the concept of BIM aims to provide 3D as well as 4D information for construction sites that could be directly used for visualization in AR, in reality a lot of companies still work with 2D CAD plans. In this case, we have to apply a data conversion step that transcodes the 2D information into 3D models that can be used for visualization purposes. Often BIM or CAD data is not geo-referenced, thus for outdoor usage we have to register them according to the physical world. For instance by using at least three point correspondences to a geo-referenced model and the Absolute Orientation Algorithm.

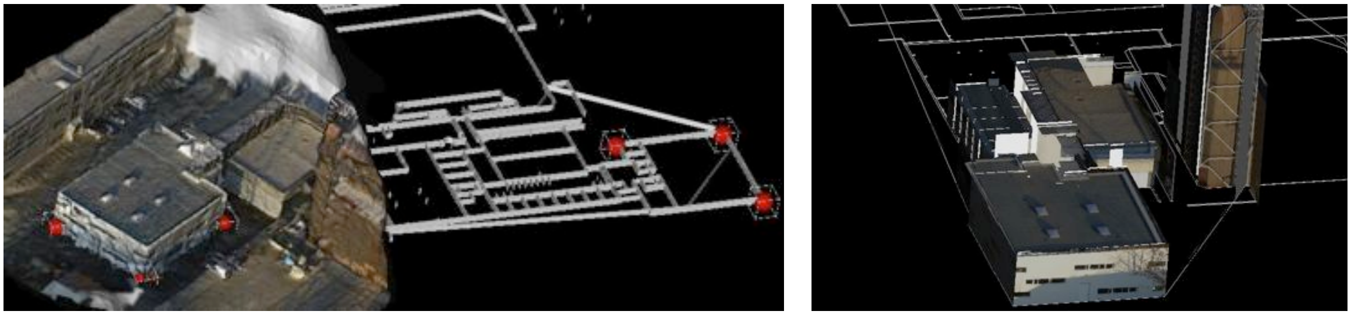

Fig. 6. Blockworld. Left) Computing the absolute orientation. Selecting the 3D points in the 3D mesh data and selecting the corresponding 3D points in the CAD data (red spheres). Right) Textured Blockworld model computed from a 3D mesh of a construction site.

C. Combining 3D Meshes and External Data Sources

Another aspect is that 3D meshes or point clouds can be to complex for understanding a visualization in AR. In this subsection, we will discuss a method for converting 3D meshes into a more simplified representation based on the extracted as-planned data. The simplified representation uses abstract block-shaped objects instead of mesh data. Due to their characteristics, we call this representation *Blockworld* representation. For instance, a 3D mesh representing a wall can be simplified by a box-shaped wall object.

As described in the previous discussion, a lot of as-planned data is already stored in external data sources such as BIM. However, such data is not always available, does not represent the time component or is simply not accurate enough. In these cases, we need an additional method to create this kind of data. The main goal of this subsection is to describe a semi-automatic tool for creating an abstract representation from a 3D mesh and a 2.5D plan. While complete manual modeling of this data would require a large amount of effort, especially when it comes to data which varies over time, these semi-automatic tools allow for a fast and intuitive creation of an abstracted representation with a minimal user interaction.

To create these abstract models, we use different data sources such as:

- 2.5D as-planned data.
- Geo-referenced point cloud data of the as-built situation.
- Geo-referenced camera images that are registered to the point cloud data.

The 2.5D input objects are transcoded into a 3D representation by using an automatic extrusion transcoder [22]. After transcoding, every polygon is represented as an extruded line with a certain height and depth. To be able to adapt the abstract as-planned data to the actual 3D mesh, both datasets have to be registered to each other. Since the 3D mesh data is geo-referenced, the user is asked to perform a rough manual registration by selecting at least three corresponding points from both datasets (Figure 6, Left). Using the selected points and the *Absolute Orientation* method [23] a transformation matrix between the datasets is computed. The transformation matrix is used to align the 3D mesh data to the as-planned data. After achieving this kind of rough registration, automatic methods can be applied to adapt the 3D extrusions to the exact characteristics of the 3D mesh and to texture the 3D extrusions.

In order to automatically adapt the abstract as-planned to the dimensions of the as-built point cloud, we analyze the data of the point cloud and derive height, depth and width measurements. For this purpose, we compute histograms based on the bounding box given by the as-planned data. To compute the height, we project all 3D points within this bounding box on the y-axis (up-vector of the bounding box). This projection provides a histogram, that allows us to compute the maximum using the histogram entries. Similar methods are used to compute the depth and width of the objects.

This adaption step results in a set of rectangular abstract objects. In order to texture these objects, we compute the most appropriate camera for texturing and apply projective texture mapping.

After performing these manual and automatic adaptions, the result of is an abstract representation that consist of several 3D blocks, the blockworld (Figure 6, Right).

V. Registration

The first prerequisite for visualizing data in AR is the registration between digital data and video image. An accurate registration assures that virtual objects are correctly aligned to the physical world. There are different ways of achieving an accurate registration, ranging from simple maker-based registration techniques [24], to sensor fusion [25] and to features-based localization approaches [26], [27]. All these technologies use different methods to achieve the same goal; the alignment of the virtual data to the physical world in real-time to achieve a coherent AR visualization. Other techniques focus on achieving high-precision registration with higher computational costs for Mixed Reality overlays [28].

Marker-based and natural-feature-target-based techniques are often used for indoor AR applications. For the outdoor applications, they are usually not an option due to larger working environments and environmental influences. In this section, we will describe two methods of achieving reliable registrations in unknown outdoor environments. The first method is a sensor-fusion-based approach combining panoramic mapping and tracking with external sensors such as IMU and a GPS receiver. The second method is applicable if a 3D model of the physical environment is available. In this case, we apply remote localization methods that rely on a server-client structure in combination with panorama-based or model-based tracking.

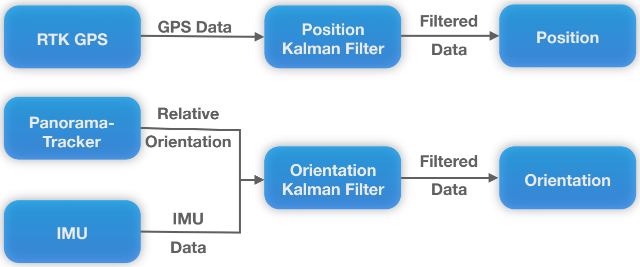

Fig. 7. Multi-sensor fusion system architecture. Data provided by the GPS sensor is filtered by a *Position Kalman Filter*. The *Orientation Kalman Filter* fuses data from *IMU* and *Visual Panorama Tracker*.

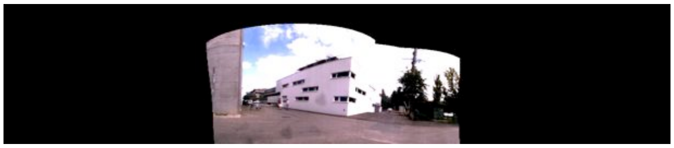

Fig. 8. Panorama generated by the panorama tracker.

A. Multi-Sensor Outdoor Registration

For AR applications that focus on entertainment and tourism in outdoor environments, researchers and companies typically use a combination of built-in sensors in mobile phones or tablet computers. For professional applications, the accuracy that these sensors provide is not sufficient. They often show positioning inaccuracies in the range of several meters and in addition orientation sensors are often a subject to drift [29]. In order to provide accurate overlays of the virtual data, we apply a registration method that is able to achieve registration accuracy in the centimeter and subangle range.

In order to achieve such a highly accurate position and orientation estimate of the AR system in outdoor environments, we combine the measurements from different sensors:

- L1/L2 Real Time Kinematic (RTK) GPS
- Inertial Measurement Unit (IMU)
- Vision-based Panoramic Tracker

The GPS receiver performs dual frequency measurements and applies RTK for accurate positioning resulting in position measurements with centimeter accuracy . We use correction data from a professional provider whose reference station delivers a correction signal to the device. Requirements for receiving the correction signal are a network connection as well as a correct configuration of the correction signal. For this purpose, we use an Open Source NTRIP application[2]. We apply a Kalman filter for the positioning (Figure 7) to compensate for positioning noise of the GPS receiver. This filter produces smooth movements and can be adjusted to the amount of smoothness of the expected movement [25].

For estimating their orientation, AR systems often use inertial sensors with gyroscopes, magnetometers and accelerometers. Unfortunately, these sensors are subject to drift and also sensitive to environmental influences, such as electromagnetic interference that often occur in urban environments. We combine the orientation measurements from an IMU with relative measurements provided by a vision-based panorama-tracker [25] to avoid these problems.

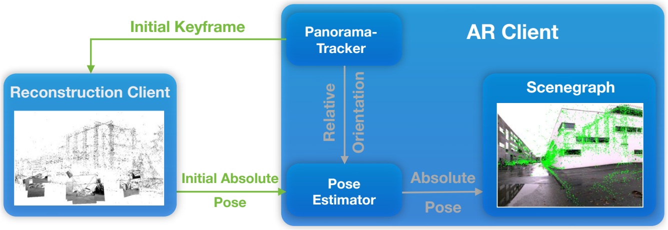

Fig. 9. Model-based localization and panoramic tracking. The panorama tracker updates the pose estimation with relative orientation measurements. In order to determine the absolute pose, an initial keyframe is send to the localization client.

The panorama tracker is based on feature detection and matching. During a mapping step, it creates a panoramic representation of the environment that is stored for further tracking. As the panoramic map is based on a cylindrical representation of the environment, the panorama tracker assumes only rotational movements. Similar to SLAM approaches [30], the idea is to 1) determine the pose of a new frame in relation to already mapped data, and 2) add features from this newly localized frame to the existing map that is then used again for further pose estimation [31], [32]. In order to determine the pose of a new image relatively to the panoramic map, we extract features from the camera image and match them against the map.

By combining the measurements of the IMU and the panorama tracker in an *Orientation Kalman* filter as shown in Figure 7, we are able to achieve robust absolute orientation measurements that are mainly drift-free [29]. Other inaccuracies can result from the magnetic deviation, a value that describes how much the absolute orientation measurement differs from the geographic north. This measurement depends on the current location and has to be configured manually or can be computed from the current GPS location.

B. Remote Localization and Online Tracking

However, even professional external sensors are only able to deliver sufficient accuracy under perfect conditions, such as in unoccluded areas with a high number of visible satellites. In order to provide accurate measurements also under more difficult conditions, we propose additional registration techniques that are based on remote localization and online tracking. These approaches require a 3D model of the environment as input for localization. Typically, in our scenarios such 3D models are available since they are created by the aerial 3D reconstruction. Thereby, it is important that the reconstructed data covers the area of interest to allow for a robust localization. Additionally, we can extend the area by integrating new images into the 3D reconstruction.

In order to integrate a model-based localization into our system, we use the talker-listener structure based on ROS as described in Section III. On the *AR Client*-site, either a visual panorama tracker or a model-based 6DOF tracker receives absolute pose measurements and uses the video stream as tracking input.

[2] GNSS Surfer http://igs.bkg.bund.de/ntrip/download

1) Panoramic tracker: The panoramic tracker uses the incoming camera images to calculate its orientation relative to an initial keyframe. We compute the geo-referenced absolute pose of this panoramic representation by sending the initial keyframe to the *Reconstruction Client* (Figure 9). The *Reconstruction Client* calculates the pose of the keyframe in relationship to the geo-referenced model and sends the localization information back to the *AR Client*. The *AR Client* combines the absolute pose with its relative orientation measurements to compute the current pose of the device.

The accurate registration in relationship to the geo-referenced model allows for the rendering of accurate AR visualizations as long as the user performs purely rotational movements. If the user moves to a new position, the motion model of the panoramic tracker that assumes rotational movements, loses the tracking. Accordingly, the tracker starts to create a new panoramic representation of the environment with a new initial keyframe. Therefore, we have to perform a relocalization and send the new initial keyframe to the localization client again (Figure 9). The process of sending the image data as well as receiving the localization information is implemented in a separate thread. The panoramic tracker runs in real-time and has been shown to perform in realtime even on mobile phones [33].

2) Model-based tracker: While the panoramic tracker is only able to handle rotational movements, we provide another registration method that uses the 3D point cloud data for initializing a tracking model. This allows for translational as well as rotational movements. The model-based tracker uses the talker-listener concept of ROS as well. Similar to the panoramic tracker, the *AR Client* publishes an initial camera frame and waits for answers that contain a localization matrix of this frame in relationship to the geo-referenced point cloud. Based on this image data, the remote *Registration Client* computes the localization matrix and publishes the result within the network. After receiving the localization matrix, the *AR client* initializes the tracking model for model-based tracking with the initial pose, the initial camera image, and the 3D points of the environment. For this purpose, we create a *Point Feature* for each 3D point that is visible in the initial camera image. A Point Feature contains 3D information of a point cloud element and 2D image patch data given by the 2D image data and the 2D location of this 3D point in the initial camera image. Based on this initialized model, movements relative to the initial localization matrix can be calculated as soon as a new camera image arrives. We compute correspondences between the 2D image patch data of all stored visible Point Features of the reference frame and the incoming camera image. The 3D information of the corresponding Point Features are used to compute the relative transformation between reference camera image and the current camera image. Finally, the *AR Client* combines this information into an absolute transformation matrix in reference to the geo-referenced model. The complete tracking step takes approximately 13 ms on the tablet PC (Motion J3400). As long as the model-based tracker finds enough correspondences between the model and new incoming camera frames, the initialized model can be used.

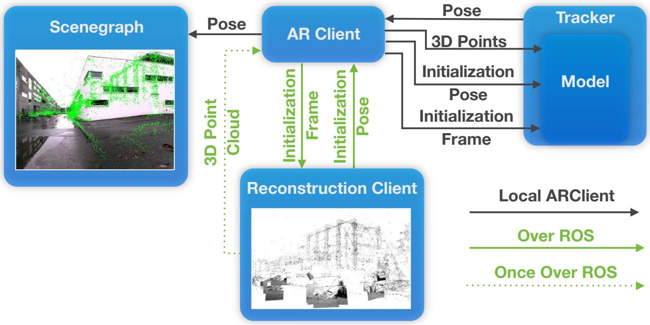

Fig. 10. Model-based tracking. An initial camera frame is remotely localized and combined with 3D point cloud data in order to initialize a local model for tracking.

However, if the AR client moves too far away from the initial pose, there are not enough correspondences available to compute the transformation. Thus, we have to compute a new model. The process starts again by publishing the camera frame and waiting for localization answers (Figure 10). This re-initialization step takes between 500 ms and 1000 ms on the tablet PC. As long as the model-based tracker computes accurate registration values in relationship to the geo-referenced model, the AR client can create accurate geo-referenced AR overlays (Figure 10, Left).

VI. Visualization

Having a mobile AR setup available that is registered to the physical world allows for the visualization of progress information on-site. In the following, we will discuss several ways of presenting the progress information. We will first discuss techniques that simply overlay the 3D progress information, continue with techniques that use filtering to avoid information clutter, and finally we will present an approach that allows for the presentation of progress information by displaying multiple points in time in the AR view. Furthermore, we will discuss their advantages and disadvantages and propose solutions in order to address these problems.

A. Overlay Techniques

The most straight forward way of presenting progress information in an AR overlay on-site is the overlay of data on top of a video image using different compositing techniques. The compositing step thereby controls the way in which digital progress information is combined with the video image of the physical world.

1) Naïve Overlay: The naïve overlay renders the 3D progress information on top of the video images (Figure 11 and Figure 12, Left). This means that it fully replaces the video information at fragments that contain 3D information. The biggest problem of overlaying 3D progress information onto the camera image is that most of the video information is occluded by the 3D information. Thus, the user often loses the spatial relationship between digital information and physical world environment.

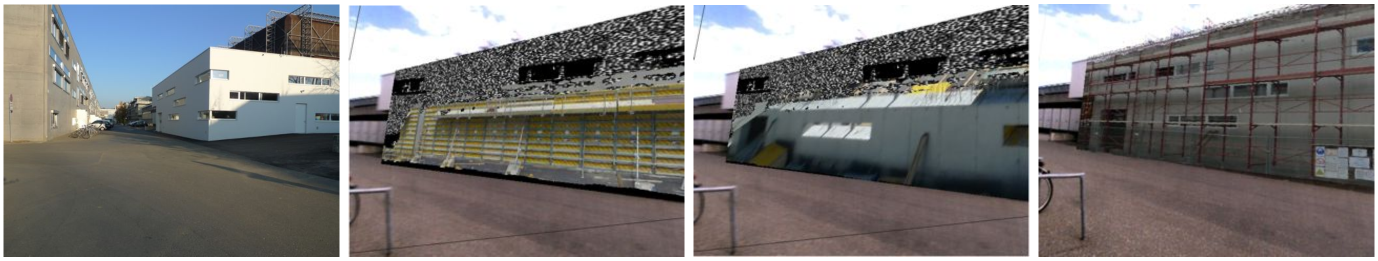

Fig. 11. One wall of a construction site at different points in time represented by textured blockworld models. Left) Scaffold at the beginning of the construction site. Middle) Wall of first floor. Right) All floors finished with a scaffold in front.

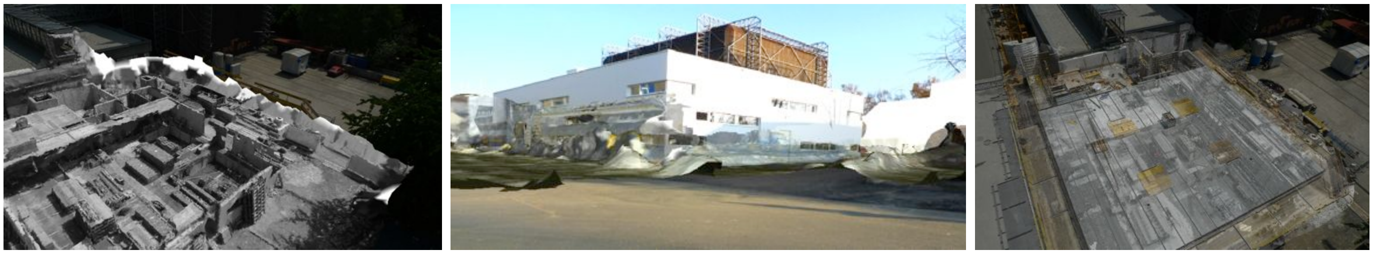

Fig. 12. Overlay techniques for presenting a previous as-built status embedded into a video image of a construction site. Left) Video image of the construction site. Middle Left) Naïve overlay, the 3D mesh rendered in grayscale is simply rendered on top of the video image. In this case, nearly the complete video view is occluded by 3D information. Thus the spatial relationship may get lost. Middle Right) Using alpha blending to combine 3D mesh data and video image. This allows the presentation of both information in one view, but often leads to information clutter and misses important image structures. Right) Ghosted view preserving important image structures in the blended compositing.

2) Blending: In order to present video information as well as 3D progress information in the same view, fragment-based blending can be applied. In this case, we display video content and 3D information with half transparency (Figure 12, Middle). Considering this on a fragment base, at each fragment where 3D information is available, we compute the average value between both fragments and achieve an impression of a transparent occluding layer. If the 3D information is not only available for previous states of the area of interest, but also for the current status, we can use this information to create correct occlusions. Based on this information, only those fragments whose depth are larger than the depth of the current reconstruction are rendered transparently. The disadvantage of this technique is that since both information layers are displayed with the same importance weighting, important structures may get lost. For instance in Figure 12 (Middle), the outlines of the house and the windows are important for understanding the shape of the building. Therefore, the rendering should take care that this information is preserved.

3) Ghostings: Ghosting techniques are well known for X-Ray visualization in Illustrative Renderings [34] as well as for X-Ray renderings in Augmented Reality [35]. Image-based ghostings preserve important image structures, by extracting edges or salient regions from the camera image. For creating a ghosted view, virtual content and camera image are composed through the use of transparency blending with a defined alpha value similar to the previously described blending technique. Additionally, the ghosting techniques extract important image structures and render them on top of this composed view. As shown in Figure 12 (Right), we extract edges and overlay them on top of the blended compositing. Thus, we preserve important structures, such as the edges of the rooftop. This allows the observer to better understand the spatial structures and the order. Nevertheless, for complex scenes these techniques are often subject to information clutter, since the already complex environment is overlaid with complex 3D information.

B. Filtering Techniques

Spatial information filtering combined with a set of interactive Focus&Context tools can be used to explorer a limited number of points in time. For this purpose, the user interactively selects a focus area to define where information about a previous point in time is to be displayed. The information filtering method then only displays virtual information in the dedicated area. At the same time, the current status of the construction site, which is represented by the camera image, is exclusively shown in the context region.

Focus&Context tools address the problems of information clutter by allowing the user to define a focus region either in image or 3D space. Within this focus region, the virtual information representing a previous point in time is displayed. The remaining part of the view outside the focus area is called the context region. In the context region, the current status of the construction site is displayed using the current camera image of the physical environment. The interactive tools that we will discuss in the following comprise 2D sliders, 2D magic lenses, 3D magic lenses and 3D sliders. For a fast prototyping, we use shaders that allow for a fast adaption of the spatial information filtering based on different focus areas.

1) 2D Slider: 2D sliders are tools that are often used for interactive side-by-side visualizations, in particular for before-after comparison of urban or historical scenes. For instance, 2D sliders are used for side-by-side comparison for the documentation of urban development. Here, current photographs of

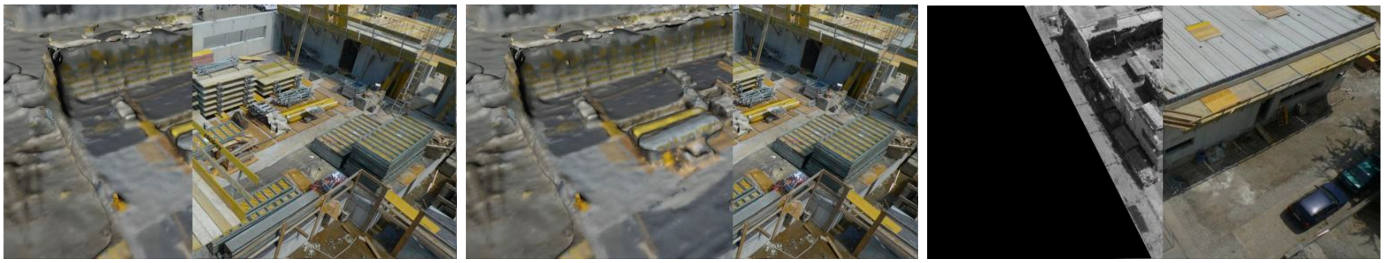

Fig. 13. Side-by-side visualization using 2D sliders. The mouse input defines the border between video image and virtual content. Left and Middle) The virtual content is overlaid on the right side of the border. Right) On the right side of the border only the virtual data representing a previous point in time is displayed.

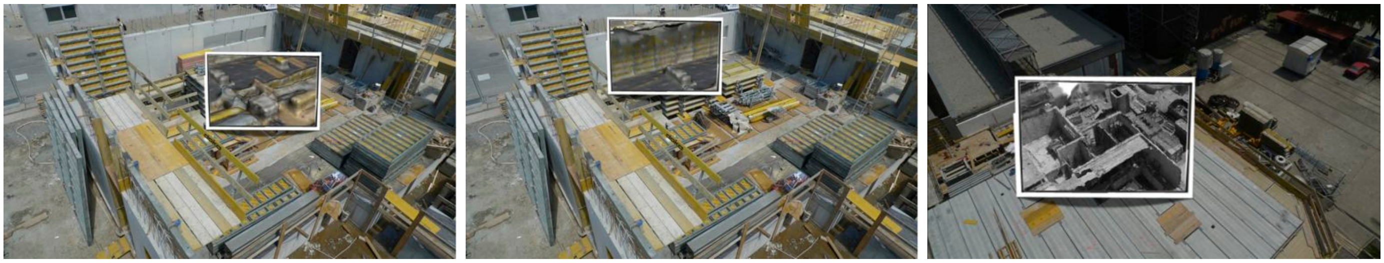

Fig. 14. 2D Magic Lens. Inside the focus region (highlighted in white) the virtual content is displayed. Left) Overlay of virtual data inside the context region. Right) Grey shading inside the focus area supports the differentiation between virtual information and physical world.

buildings are displayed next to older photographs. In order to provide such a visualization, it is important that both images (past and present) are captured from the same view point. The process of taking a picture from the same position is called re-photography. The photographer of the newer image has to take the photograph from the same position as the older photograph was taken from. In order to find a similar position, photographers search for specific features in the scene that are easily to recognize such as walls or roofs. Recently, researchers even developed automatic methods that guide the photographer to the right pose by using SIFT features [36] and pose estimation [37].

The drawback of re-photography is that it limits the observer to views that were captured in the past. If the previous status is available as a 3D reconstruction, it is also possible to choose different viewpoints. By overlaying the 3D information onto the camera image of the current environment, the technique moves from re-photography to AR (Figure 13). Similar to the visualization methods used for re-photography, the user can move a slider interactively in image space to control the border between video image and the virtual overlay containing the 3D information. In our application, the 3D information contains a previous status of the construction site, but the technique could also be used for different urban scenarios as long as the 3D data is available. When the user clicks on an arbitrary location in the 2D view, the x-coordinate of the mouse pointer is used to define the border. The fragment shader then displays the video image for all fragments with x-coordinates larger than the border's x-coordinate. For other fragments the virtual data is either overlaid onto the video image (Figure 13, Left) or rendered exclusively (Figure 13, Right).

2) 2D Magic Lens: Another technique that is often used to create before-and-after effects is the embedding of a cutout of old photographs into new ones. This technique is similar to magic lens interfaces from Information Visualization [38]. In contrast to the 2D slider, a 2D magic lens allows the user to move a dedicated focus area to the regions of interest. Mouse coordinates define the center of the focus region. The focus region can have different shapes, such as a rectangular or circular shape. Based on the shape definition, the fragment shader tests if a fragment is inside the focus area and renders the virtual content if this is the case. Thereby the virtual content is again either displayed using an overlay or exclusive rendering. In addition, the virtual content can also be rendered using different shadings, such as Toon shading or a gray scale shading to highlight the difference to the actual physical environment (Figure 14).

The advantage of the 2D Focus&Context tools is that the visualization requires no additional information or post processing of the reconstructed data. However, they have the disadvantage of not addressing self-occlusion, since they simply define a 2D region in the user's view. This means we can filter information in image space but not in depth. Furthermore, it is difficult to present more than one point in time.

3) 3D Slider: In order to address the problem of self-occlusion, we investigated tools that allow focus and context regions to be defined in 3D. This spatially filters the information in all three dimensions. Such a 3D filtering is particularly interesting for the visualization of 3D data that was reconstructed with aerial vision, since the users may want to inspect one specific element that is occluded by other previous structures in their view. By defining a focus area in 3D, it is possible to exclusively visualize information for this selected region.

Another interesting aspect of 3D tools is their ability to con-

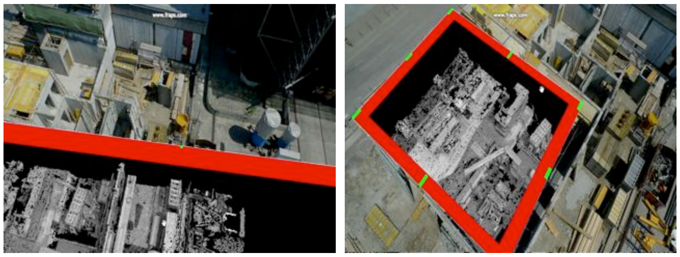

Fig. 15. 3D tools for information filtering. Left) A 3D slider allows the user to divide the view in an area where a previous status of the construction site is shown and an area where the current status is shown. Right) The 3D magic lens magic lens defines a three dimensional region where the virtual content is displayed.

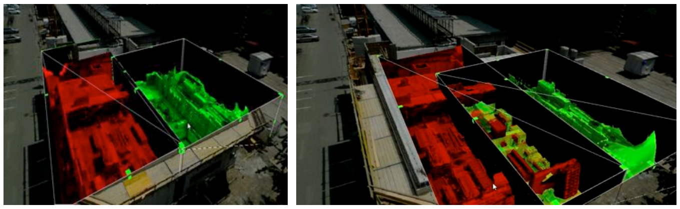

Fig. 16. 3D Focus&Context tools using different color codings. Thereby green represents an earlier stage of the construction site as red. Two 3D magic lenses display different points in time.

vey information about depth. This supports the understanding of spatial relationships between objects at the present time and objects that were at this location in the past.

Similar to the 2D slider, the 3D slider allows for the separation of the visualization into a focus area showing the 3D information and a context area showing the camera image. The difference between the techniques is that the 3D slider does the separation in 3D. The 3D slider is defined as a large wall that the user can move in the scene. To provide a convincing visualization, the wall has to be aligned in relationship to a plane in the 3D world. We provide a 3D manipulation tool that allows the user to interactively manipulate the alignment and the dimension of the 3D slider. The intersection plane between virtual geometry and sliding wall provides information about the depth and height of the virtual content.

4) 3D Magic Lens: In order to allow the selection of a smaller focus region, we provide a 3D magic lens. The 3D magic lens allows the user to define a box-shaped focus area in the AR view. The box can be interactively moved and scaled to fit the requirements of the user. The virtual content is displayed inside the box. For all elements that are outside the focus area, video image information is shown. Similar to the 3D slider, the alignment has to be done in relation to 3D planes in the scene to achieve a convincing visualization. An extension of this magic lens would align itself to the planes in the scene, for instance by detecting planes in the 3D point cloud.

So far, we have only discussed visualization tools that support the overlay of one 3D dataset. This only allows for the visualization of one point in time at once. However, visualizing the progress of a construction site often requires the comparison of multiple points in time to inspect the progress. With the existing tools, this is only possible by presenting different points in time one after another. Unfortunately, such an approach is often subject to change blindness [39]. To address this problem, we further investigated methods that allow for the presentation of multiple complex datasets in one view. For this purpose, we visualized different points in time using different visualization techniques. This allows for a differentiation between multiple datasets. For instance, we use a color-coded rendering to display different 3D datasets. Thereby, the different color-codings are implemented by enabling the corresponding color buffers (Figure 16). This technique again only makes sense in combination with Focus&Context techniques, because otherwise the view is too cluttered to understand the changes. Combining the color-coding with multiple magic lenses allows for the comparison of multiple points in time.

C. 4D Abstraction Approach

In order to visualize changes in a comprehensible way, perceptional issues like change blindness have to be addressed [39]. Change blindness describes the problem of not noticing changes. This problem often arises when many changes occur in the view seen by a person. To avoid change blindness, it is important that the user can keep track of selected changes for instance by providing additional visual hints [40]. This is particularly important for outdoor AR visualizations, since the environment is changing all the time and important changes in the data may be concealed. The aforementioned techniques do not address the issue of change blindness since they are not able to visualize multiple points in time in one view.

To address both, information clutter and change blindness, we developed an approach that follows Shneiderman's *Visual Information Seeking Mantra* [41]. The mantra defines visualization guidelines as follows: There should first be an overview, then zooming and filtering to concentrate on important information elements and finally it should be possible to access details on demand. In the following, we will show how the mantra can be applied when visualizing multiple complex 3D datasets in AR. According to the mantra, we start with defining three visualization levels varying in their level of detail.

1) Visualization Levels: According to the information seeking mantra, each level displays the virtual content on a different scale:

- L0: Scene overview level
- L1: Object time overview level
- L2: Object detail level

Each visualization level is a detailed view of the higher visualization level. Transition techniques such as Overview&Detail and Focus&Context allow for transitions between the different visualization levels and relating them to each other. We will first discuss each level of abstraction in detail, before we describing how we move from one to the other.

The first visualization level L0 represents abstract information about individual objects that summarize time-oriented attributes per object. These attributes comprise abstract information such as completion or progress and will be represented by a per-object color-coding similar to the approach in [7].

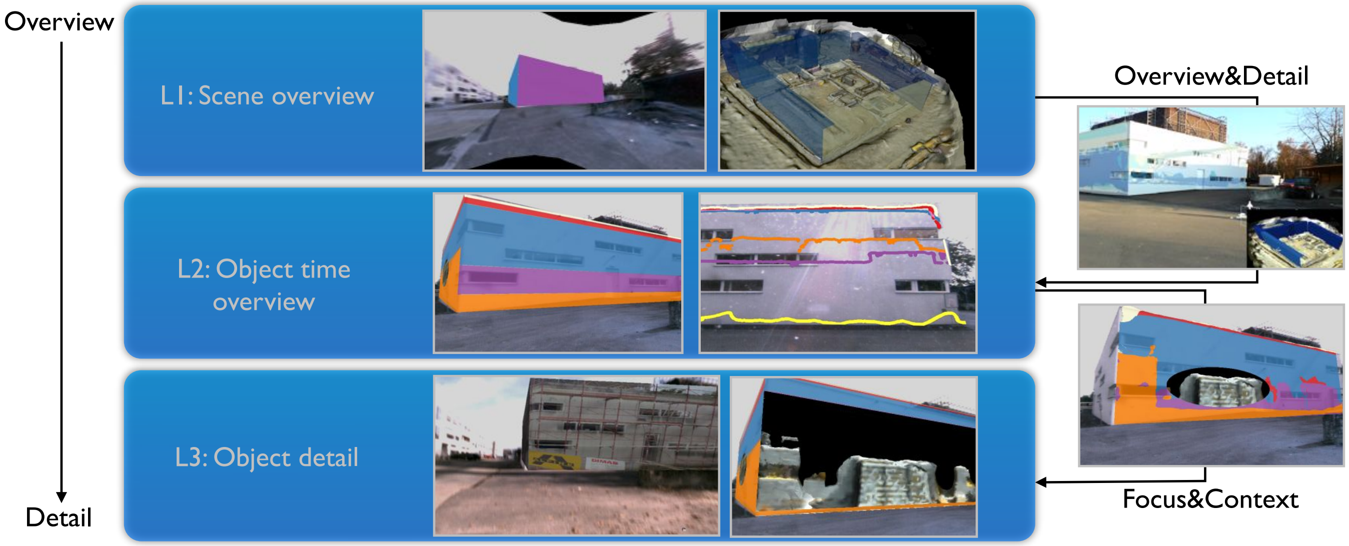

Fig. 17. 4D visualization concept. Multiple levels of detail allow time-oriented data to be explored in an AR visualization. The highest level of abstraction provides just enough information to provide an overview of the data in relation to the objects in the AR environment. The second level of abstraction presents traditional time-oriented visualizations registered in AR to enable an effective analysis of time-oriented data. Its registration in AR additionally provides information about its relation to real world structures. The third level provides structural detail of the object of interest at a selected point in time and 3D space. In order to first study the data in a higher level of abstraction before analyzing it in more detail we interactively combine all levels using Overview&Detail and Focus&Context techniques.

While this first level allows for a global overview of the scene according to *overview first*, the second level L1 presents more detailed information about one object according to the second item of the mantra. This level of information displays the variation over time of one attribute of a selected object, such as

- Height Lines: representing different heights of an object at different times (compare Figure 18, Left).
- Geometric Completion: representing the completion at multiple points in time in reference to the object geometry using a color coding (Figure 18, Middle Left and Middle Right)
- Block diagram for average completion: block diagram showing the average completion in relation to the geometry at multiple points in time (Figure 18, Left).

The third visualization level L2 provides the *details on demand* by presenting a complete 3D rendering of the one object at a selected point in time. We can present the data at this level as pure 3D dataset, as an orthogonal snapshot of the 3D data in order to reduce rendering times or by using the abstract blockworld representation described in Section IV.

2) Transitions between Visualization Levels: We provide interactive Overview&Detail tools to combine these visualization levels. This allows the user to move between the visualization levels.

In order to use these techniques, we have to define overview and detailed information regarding our visualization levels. The definition depends on the particular transition between visualization levels, since the lower level corresponds to a lower level of detail. For instance, for the transition between L0 and L1, L0 contains overview information and L1 contains detail information.

Our idea is to adapt traditional methods for managing overview and detail information in *Information Visualization* to our need to move between the 4D visualization levels. As described by Cockburn et al. [42], Overview&Detail techniques use either a temporal or spatial separation of context and detail information. Focus&Context techniques allow the presentation of both in one view by visually combing this information.

As described above, the first visualization level L0 gives an overview of the complete area of interest (e.g. complete construction site). In contrast, the second visualization level L1 provides more detail that will lead to clutter in an overview visualization. The falloff in detail of the visualization levels makes the transition between the first and the second level a perfect candidate for Overview&Detail techniques.We provide two techniques for transitions between overview and detail visualizations:

- Zooming interface
- World in Miniature (WIM)

The *zooming interface* allows the user to zoom out of their current view. For this purpose, the user either leaves the AR view and zooms out to a virtual overview (Figure 19) or zooms out of the current position using an extended view visualization [43]. The extended view visualization still provides the AR overlay, but based on a simulated wider angle of view, similar to the egocentric panoramic view of Mulloni et al. [44]. The wide angle of view allows the user to inspect objects that are not in their current field of view. Based on the amount of zooming this provides an overview of the scene where the user can inspect multiple objects at once. If the distance exceeds a defined threshold, we assume that the interest in detailed information vanishes. Accordingly, we switch to the overview level L0 and present the more abstract representations of the scene objects. Thus, the user receives

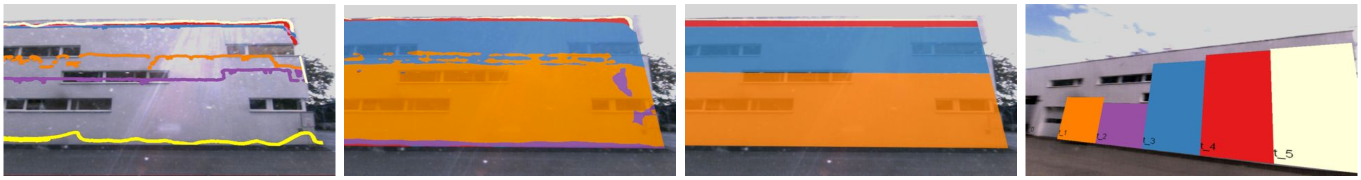

Fig. 18. Object time overview visualization: providing information about multiple points in time for a single object. The different points in time are thereby color coded. Left) Height Lines showing different heights of the wall at different points in time. Middle Left) Geometric Completion shows the completion in geometric reference to the wall. This provides a summary which parts of the wall were added at which time. Middle Right) Completion diagram showing the average completion for multiple points in time. Right) Block diagram showing the average completion of the wall for multiple points in time.

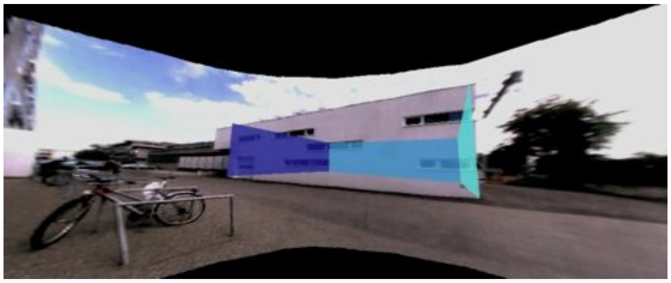

Fig. 19. Overview&Detail techniques. Using an extended field of view for providing a scene overview.

not only a spatial overview but also an overview in terms of provided information.

In the *WIM* visualization, a bird's eye view provides the user with an overview of the scene, while at the same time showing the information from the second visualization level L2 in the main view. With this tool the user can select a point of interest that determines the look-at vector of the camera. Finally, the camera is translated away from the scene using the up-vector of the scene.

In contrast to the techniques described previously, for the transition between L1 and L2 visualizations, the full scale of the visualization is critical for their interpretation. Thus, spatial Overview&Detail techniques are not suitable. Neither are temporal Overview&Detail techniques suitable, since they are subject to change blindness [39]. In order to address these issues, we apply Focus&Context techniques to combine the visualizations in a single view:

- Overlay: presents both overview information and detail information in one view as an overlay using blending.
- 2D sliders: allow for separation of the focus and the context region. Visualize the time overview on one side and the detail object information on the other side (Figure 20, Left).
- Magic lenses: The area inside the magic lens displays virtual representation from a former point in time, while its context area shows the abstract information about multiple points in time (Figure 20).
- Distorted View: View distortion techniques enlarge the focus area and reduce the context area of a visualization [45]. We adapted this technique to embed detailed object information of one point in time while showing a compressed overview of the relative changes in the same view (Figure 21).

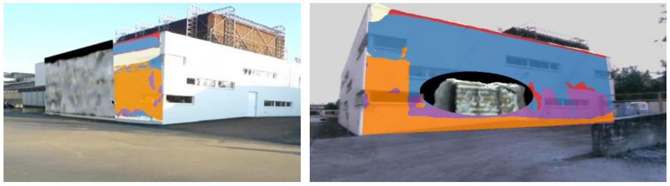

Fig. 20. Transitions between visualization levels with a 2D Slider and a Magic Lens. Left) Side-by-side visualization of time overview information and object detail. Right) The magic lens area provides a detailed view on a scaffold used to built the wall.

For more details and information about the implementation of 4D visualization techniques please refer to [46].

VII. Interacting with the Physical World

This is of particular relevance for construction site related tasks. For instance, for construction site monitoring, it is often important to understand spatial dimensions in greater detail. This can be supported by interactive surveying tools. Furthermore, for documentation purposes it is often advantageous to store annotations which correspond to physical objects. If these annotations can be stored not only as notes, but also within the spatial context of the physical environment, this often makes it easier to access the information when inspecting the area of interest later on. In the following, we will describe interactive tools that we integrated into the mobile AR client for surveying and for annotating the physical environment on-site. Both interactive techniques require the availability of a 3D representation of the physical environment. This can be achieved using the 3D *Reconstruction Client* as described previously.

A. *Surveying*

The main goal of a surveying task is to measure the dimensions of an object in the user's proximity. Traditionally, for surveying, the user has to make use of a set of additional devices such as a laser measurement device. Since our AR system has access to a 3D representation of the physical environment, we can provide an interactive tool that takes these measurements within the AR view. In order to survey an object the user has than to select a point in space by clicking this point in the 2D view space. Based on this 2D coordinate and the known transformation of the AR client, we compute a ray in 3D in relationship to the physical environment. The 3D ray is then used to compute the 3D intersection point with the 3D representation of the environment. Finally, we use this 3D

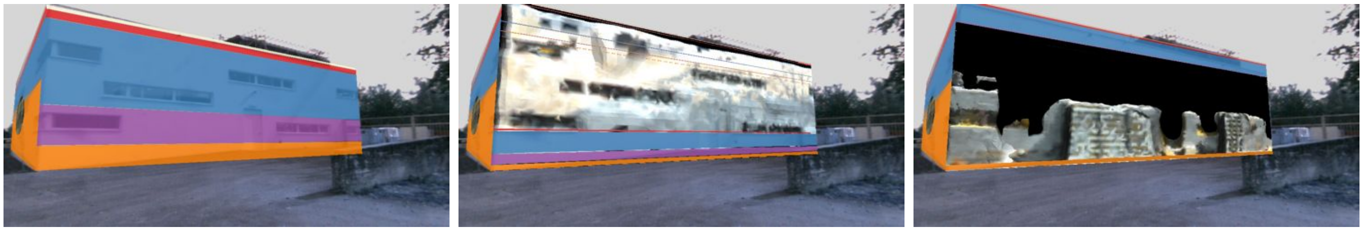

Fig. 21. Transitions between visualization levels with distorted view. Left) Average Completion Diagram. Middle and Right) By selecting a point in time within the completion diagram the user gets a detailed visualization of the wall at the point point in time he selected. To provide still the abstract information, the completion diagram is distorted and presented in the same view.

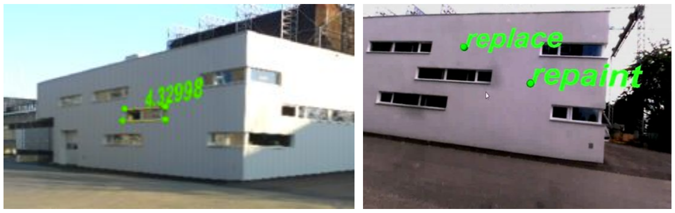

Fig. 22. Interactive techniques. Left) AR surveying of a window. Right) AR annotations used to store annotations spatially aligned to the building.

coordinate as a first surveying point and highlight it in the AR view. We can either display the geo-referenced position of this surveyed point or use it to compute dimensions of objects of interest. For instance, in Figure 22 (Left), the interactive input is used to determine the width of a window. For this purpose, the user selects multiple points and for each pair of points the dimension is then displayed in the AR view.

B. Annotations

While surveying tasks are usually used to get immediate feedback, annotations are mostly used to leave information for later usage or to exchange information with other users. Regarding construction site documentation, interactive annotation tools can be used to store information right next to the annotated object. For instance, in Figure 22 (Right) we show an example where a building is annotated with a series of suggestions for modifications. The main idea is that the construction site staff responsible can store these suggestions directly on the object suggested for modifications. Workers can then access this information later on directly on site. In order to be able to place these annotation in relationship to the object of interest, we again need a 3D point in the environment similar to the surveying techniques. Thus, we compute an intersection point with the 3D representation and put a text object at this 3D input position.

VIII. Conclusion

In this paper, we introduced an approach for using Augmented Reality for on-site construction site monitoring and documentation. Our approach combines aerial vision with a mobile AR interface and thus allows relevant information to be accessed directly on-site. In order to achieve this, we developed methods for aerial data capturing and 3D reconstruction. Furthermore, we introduced methods to post-process this data. Using the reconstructed data and external sensors, we are able to register the mobile AR setup in relationship to the physical world. For this purpose, we implemented different techniques that either use additional sensors, such as GPS and IMU or are purely vision-based. Having the data and the registration available, we can visualize the progress information for documentation and monitoring purposes directly on site. In this context, we developed a set of visualization techniques that address the special needs for the visualization of progress information. Finally, we discussed method of annotating and surveying objects within the AR environment. With this approach combining different components, we were able to show that AR can support the documentation and monitoring of construction sites.

We tried the approach in several field tests to learn more about the suitability for outdoor usage and robustness of the registration techniques. These field tests were performed during different seasons and day times. We experienced that the registration techniques are working robustly under different lighting and environmental conditions. Furthermore, we received first feedback from our partner company confirming the high potential for various future business.

However, there are still some unresolved issues. So far we have only included multiple as-built datasets into our visualization. Another interesting aspect here is how to make use of change detection methods. For instance, Golparvar-Fard used a voxel-based labeling to automatically detect progress [12]. This kind of information is important in order to be able to highlight important changes in the construction environment. For future work, we plan to integrate change detection data to support our visualization. Another issue is the availability of data. While 4D BIM data is already an important topic regarding the storage of the life cycle of buildings, a lot of construction companies still work with 2D maps. In order to address this issue, we proposed a method for converting data into a 3D representation, However, having 4D as-planned data available allows for more possibilities such as the direct comparison to the as-built status at selected points in time.

Acknowledgment

This work was supported by the Austrian Research Promotion Agency (FFG) FIT-IT project Construct (830035).

References

[1] R. T. Azuma, “A survey of augmented reality,” *Presence: Teleoperators and Virtual Environments*, vol. 6,

no. 4, pp. 355–385, 1997. [Online]. Available: http://nzdis.otago.ac.nz/projects/projects/berlin/repository/revisions/22/raw/trunk/Master' sDocs/Papers/ASurveyofAugmentedReality.pdf

[2] G. Schall, S. Zollmann, and G. Reitmayr, “Smart Vidente: advances in mobile augmented reality for interactive visualization of underground infrastructure,” *Personal and Ubiquitous Computing*, pp. 1–17, 2012. [Online]. Available: http://link.springer.com/article/10.1007/s00779-012-0599-x/fulltext.html

[3] C. Woodward, M. Hakkarainen, O. Korkalo, T. Kantonen, K. Rainio, and K. Kähkönen, “Mixed reality for mobile construction site visualization and communication,” in *10th International Conference on Construction Applications of Virtual Reality (CONVR2010)*, 2010, pp. 1–10.

[4] R. Schoenfelder and D. Schmalstieg, “Augmented Reality for Industrial Building Acceptance,” in *IEEE Virtual Reality Conference (VR 2008)*, 2008, pp. 83–90. [Online]. Available: http://ieeexplore.ieee.org/Xplore/login.jsp?url=http://ieeexplore.ieee.org/iel5/4472735/4480728/04480755.pdf?arnumber=4480755\&authDecision=-203

[5] K. Pentenrieder, C. Bade, F. Doil, and P. Meier, “Augmented Reality-based factory planning - an application tailored to industrial needs,” in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2007)*. Ieee, Nov. 2007, pp. 1–9. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4538822

[6] G. Klinker, D. Stricker, and D. Reiners, “Augmented reality for exterior construction applications,” *Augmented Reality and Wearable Computers*, vol. 35, p. 1, 2001.

[7] M. Golparvar-Fard, F. Pena-Mora, and S. Savarese, “D4AR - a 4 dimensional augmented reality model for automation construction progress monitoring data collection, processing and communication,” *Journal of Information Technology in Construction*, vol. 14, pp. 129–153, 2009.

[8] G. Schall, D. Schmalstieg, and S. Junghanns, “VIDENTE-3D Visualization of Underground Infrastructure using Handheld Augmented Reality,” *Geohydroinformatics-Integrating GIS and Water Engineering” CRC Press/Taylor and Francis Publisher: CRC*, vol. 1, pp. 1–17, 2010. [Online]. Available: http://www.icg.tu-graz.ac.at/Members/schall/geohydro.pdf/download/

[9] D. Shin and P. Dunston, “Identification of application areas for Augmented Reality in industrial construction based on technology suitability,” *Automation in Construction*, vol. 17, no. 7, pp. 882–894, 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0926580508000289

[10] G. R. King, W. Piekarski, and B. H. Thomas, “ARVino - Outdoor Augmented Reality Visualisation of Viticulture GIS Data,” in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2005)*, 2005, pp. 52–55.

[11] S. Côté, P. Trudel, R. Snyder, and R. Gervais, “Augmented Reality for Underground Infrastructure and for Construction,” 2012. [Online]. Available: http://ismar.vgtc.org/ismar/2012/info/overview/demos

[12] M. Golparvar-Fard and F. Pena-Mora, “Monitoring changes of 3D building elements from unordered photo collections,” in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on Computer Vision for Remote Sensing of the Environment*, 2011, pp. 249–256. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6130250

[13] O. Faugeras, S. Laveau, L. Robert, G. Csurka, and C. Zeller, “3-D Reconstruction of Urban Scenes from Sequences of Images,” *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, 1995. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.1119

[14] I. Colomina, M. Blázquez, P. Molina, M. Parés, and M. Wis, “Towards a new paradigm for high- resolution low-cost photogrammetry and remote sensing,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2008. [Online]. Available: http://scholar.google.com/scholar?hl=en\&btnG=Search\&q=intitle:Towards+a+new+paradigm+for+high-+resolution+low-cost+photogrammetry+and+remote+sensing\#0

[15] C. Hoppe, A. Wendel, S. Zollmann, K. Pirker, A. Irschara, H. Bischof, and S. Kluckner, “Photogrammetric Camera Network Design for Micro Aerial Vehicles,” in *Computer Vision Winter Workshop*, Mala Nedelja, Slovenia, 2012.

[16] A. Irschara, V. Kaufmann, M. Klopschitz, H. Bischof, and F. Leberl, “Towards fully automatic photogrammetric reconstruction using digital images taken from UAVs,” in *Proc. International Society for Photogrammetry and Remote Sensing Symposium*, 2010.

[17] A. Irschara, C. Hoppe, H. Bischof, and S. Kluckner, “Efficient structure from motion with weak position and orientation priors,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Workshop on Aerial Video Processing*. Ieee, Jun. 2011, pp. 21–28. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5981775

[18] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, “Towards Internet-scale multi-view stereo,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2010)*. IEEE, 2010, pp. 1434–1441. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5539802

[19] M. Kazhdan, M. Bolitho, and H. Hoppe, “Poisson surface reconstruction,” in *Proc. Symposium on Geometry Processing*, 2006, pp. 61–70.

[20] P. Labatut, J. Pons, and R. Keriven, “Robust and efficient surface reconstruction from range data,” *Computer Graphics Forum*, vol. 28, no. 8, 2009. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2009.01530.x/full

[21] C. Hoppe, M. Klopschitz, M. Rumpler, A. Wendel, S. Kluckner, H. Bischof, and G. Reitmayr, “Online Feedback for Structure-from-Motion Image Acquisition,” in *Procedings of the British Machine Vision Conference 2012*. British Machine Vision Association, 2012, pp. 70.1–70.12. [Online]. Available: http://www.bmva.org/bmvc/2012/BMVC/paper070/index.html

[22] S. Zollmann, G. Schall, S. Junghanns, and G. Reitmayr, “Comprehensible and Interactive Visualizations of GIS Data in Augmented Reality,” *Advances in Visual Computing*, pp. 675–685, 2012. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-33179-4_64

[23] B. Horn, “Closed-form solution of absolute orientation using unit quaternions,” *JOSA A*, vol. 4, no. April, pp. 629–642, 1987. [Online]. Available: http://www.opticsinfobase.org/abstract.cfm?\&amp;id=2711http://www.opticsinfobase.org/abstract.cfm?\&id=2711

[24] H. Kato and M. Billinghurst, “Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System,” p. 85, Oct. 1999. [Online]. Available: http://dl.acm.org/citation.cfm?id=857202.858134

[25] G. Schall, D. Wagner, G. Reitmayr, E. Taichmann, M. Wieser, D. Schmalstieg, and B. Hofmann-Wellenhof, “Global pose estimation using multi-sensor fusion for outdoor Augmented Reality,” in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2009)*. IEEE, Oct. 2009, pp. 153–162. [Online]. Available: http://dl.acm.org/citation.cfm?id=1681509.1682321

[26] J. Ventura and T. Hollerer, “Wide-area scene mapping for mobile visual tracking,” in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2012)*, no. November, 2012, pp. 3–12. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6402531

[27] C. Arth and M. Klopschitz, “Real-time self-localization from panoramic images on mobile devices,” *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2011)*, pp. 37–46, Oct. 2011. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6162870http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6162870

[28] H. Bae, M. Golparvar-Fard, and J. White, “High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (AEC/FM) applications,” *Visualization in Engineering*, vol. 1, no. 1, p. 3, 2013. [Online]. Available: http://www.viejournal.com/content/1/1/3

[29] G. Schall, A. Mulloni, and G. Reitmayr, “North-centred orientation tracking on mobile phones,” in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2010)*. IEEE, Oct. 2010, pp. 267–268. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5643600

[30] A. J. Davison, “Real-Time Simultaneous Localisation and Mapping with a Single Camera,” in *IEEE Ninth IEEE International Conference on Computer Vision*. IEEE Computer Society, Oct. 2003, pp. 1403–1410. [Online]. Available: http://dl.acm.org/citation.cfm?id=946247.946734

[31] G. Reitmayr, T. Langlotz, D. Wagner, A. Mulloni, G. Schall, D. Schmalstieg, and Q. Pan, “Simultaneous Localization and Mapping for Augmented Reality,” *2010 International Symposium on Ubiquitous Virtual Reality*, pp. 5–8, Jul. 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5557942

[32] H. Kim, G. Reitmayr, and W. Woo, “IMAF: in situ indoor modeling and annotation framework on mobile phones,” *Personal and Ubiquitous Computing*, vol. 17, no. 3, pp. 571–582, Apr. 2012. [Online]. Available: http://link.springer.com/10.1007/s00779-012-0516-3

[33] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg, “Real-time panoramic mapping and tracking on mobile phones,” in *IEEE Virtual Reality Conference (VR 2010)*. Ieee, Mar. 2010, pp. 211–218. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5444786

[34] S. Bruckner, S. Grimm, A. Kanitsar, and M. E. Gröller, "Illustrative Context-Preserving Exploration of Volume Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1559–1569, 2006.

[35] S. Zollmann, D. Kalkofen, E. Mendez, and G. Reitmayr, "Image-based ghostings for single layer occlusions in augmented reality," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2010)*. IEEE, Oct. 2010, pp. 19–26. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5643546

[36] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004. [Online]. Available: http://dl.acm.org/citation.cfm?id=993451.996342

[37] S. Bae, A. Agarwala, and F. Durand, "Computational rephotography," *ACM Transactions on Graphics*, vol. 29, no. 3, pp. 1–15, Jun. 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1805964.1805968

[38] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose, "Toolglass and magic lenses," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques - SIGGRAPH '93*. New York, New York, USA: ACM Press, Sep. 1993, pp. 73–80. [Online]. Available: http://dl.acm.org/citation.cfm?id=166117.166126

[39] D. J. Simons, "Current Approaches to Change Blindness," *Visual Cognition*, vol. 7, pp. 1–15, 2000.

[40] L. Nowell, E. Hetzler, and T. Tanasse, "Change blindness in information visualization: A case study," in *Proc. of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, 2001, pp. 15–22. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=963274http://dl.acm.org/citation.cfm?id=857721

[41] B. Shneiderman, "The Eyes Have It: {A} Task by Data Type Taxonomy for Information Visualizations," in *IEEE Visual Languages*, 1996, pp. 336–343. [Online]. Available: http://dx.doi.org/10.1109/VL.1996.545307

[42] A. Cockburn, A. Karlson, and B. B. Bederson, "A review of overview+detail, zooming, and focus+context interfaces," *ACM Comput. Surv.*, vol. 41, no. 1, pp. 2:1—-2:31, 2009. [Online]. Available: http://doi.acm.org/10.1145/1456650.1456652

[43] S. Hwang, H. Jo, and J.-h. Ryu, "EXMAR : EXpanded view of Mobile Augmented Reality," *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2010)*, pp. 3–4, 2010.

[44] A. Mulloni, A. Dünser, and D. Schmalstieg, "Zooming interfaces for augmented reality browsers," in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services - MobileHCI '10*. New York, New York, USA: ACM Press, Sep. 2010, p. 161. [Online]. Available: http://dl.acm.org/citation.cfm?id=1851600.1851629

[45] R. Spence and M. Apperley, *Bifocal Display*. Aarhus, Denmark: The Interaction Design Foundation, 2011. [Online]. Available: http://www.interaction-design.org/encyclopedia/bifocal_display.html

[46] S. Zollmann, D. Kalkofen, C. Hoppe, S. Kluckner, H. Bischof, and G. Reitmayr, "Interactive 4D overview and detail visualization in augmented reality," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2012)*, 2012. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=\&arnumber=6402554\&contentType=Conference+Publications\&queryText=Interactive+4D+Overview+and+Detail+Visualization+in+Augmented+Reality

Stefanie Zollmann is a postdoctoral researcher at the Institute for Computer Graphics and Vision at Graz University of Technology. She graduated in Media Systems at Bauhaus University Weimar, Germany, in 2007. Recently, she received a PhD degree from the Institute for Computer Graphics and Vision in Graz. Her main research interests are visualization techniques for augmented reality, but also include mobile augmented reality and spatial augmented reality in general.

Christof Hoppe Christof Hoppe received his B.S. degree in computer science from the University of Kassel, Germany, in 2007. In 2010, he graduated at TU Graz and received his M.S. degree in computer vision and artificial intelligence. Currently he is a PhD student at the Institute of Computer Vision and Graphics at Graz University of Technology, Austria. His research interests are autonomous flying and driving vehicles, realtime 3D reconstruction, Structure-from-Motion and Simultaneous Localization and Mapping (SLAM).

Stefan Kluckner joined Siemens Corporate Technology in the beginning of 2011. Previous to that, he was a scientic research assistant at the Institute for Computer Graphics and Vision and received a MSc and a PhD degree from Graz University of Technology in 2006 and 2011, respectively. His research interests include geometry recovery from photos, semantic interpretation and holistic scene understanding as well as the integration of visual appearance and 3D information.

Christian Poglitsch is a master student for Software Development and Business Management at Graz University of Technology. Recently, he received his BSc degree in the same field. Currently, he is working as teaching assistant at the Institute for Computer Graphics and Vision in Graz. His main research interest is interactive 3D modeling.

Horst Bischof received his M.S. and Ph.D. degree in computer science from the Vienna University of Technology in 1990 and 1993, respectively. In 1998 he got his Habilitation (venia docendi) for applied computer science. Currently he is Vice Rector for Research at Graz University of Technology and Professor at the Institute for Computer Graphics and Vision at the Graz University of Technology, Austria. His research interests include object recognition, visual learning, motion and tracking, visual surveillance and biometrics, medical computer vision, and adaptive methods for computer vision where he has published more than 650 peer reviewed scientific papers and received several (19) awards.

Gerhard Reitmayr is professor for Augmented Reality at the Graz University of Technology. He received his Dipl.-Ing. (2000) and Dr. techn. (2004) degrees from Vienna University of Technology. He worked as a research associate at the Department of Engineering at the University of Cambridge, UK until May 2009 where he was researcher and principal investigator. Research interests include the development of augmented reality user interfaces, wearable computing, ubiquitous computing environments and the integration of these. Research directions include computer vision techniques for localization and tracking and interaction methods.