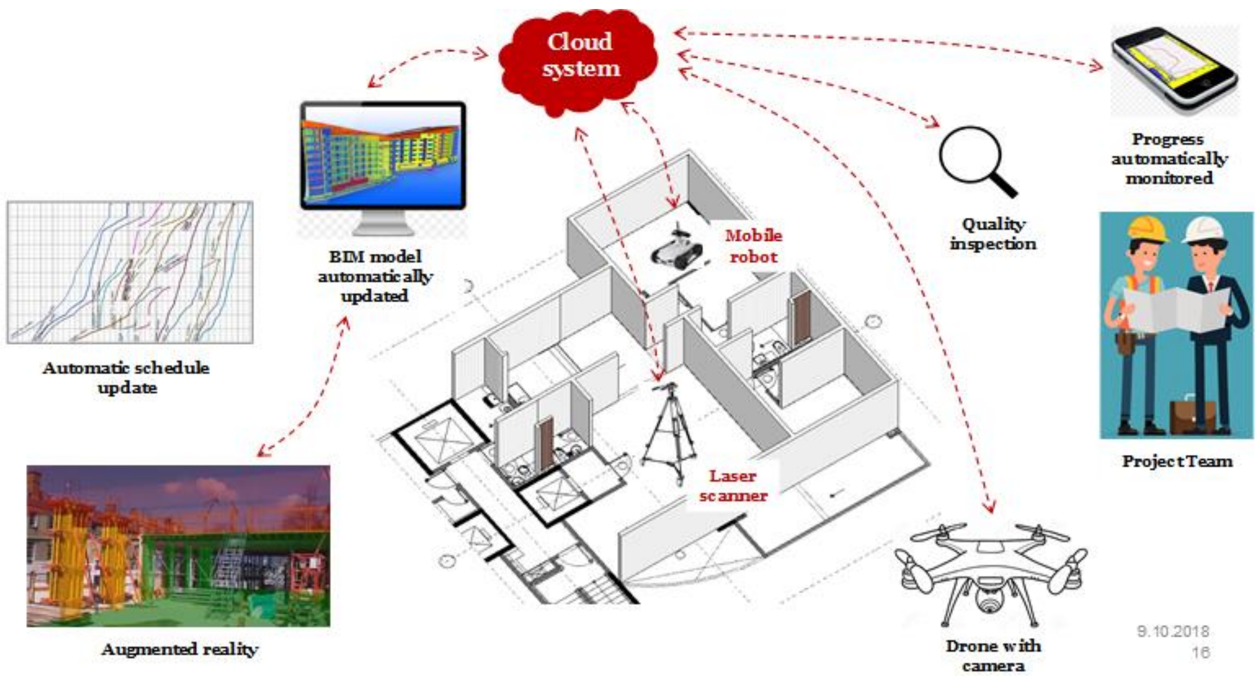
Seppänen, Olli; Xiao, Yu; Khalid Masood, Mustafa; Byvshev, Petr; Pham, Truong; Aikala, Antti; Lundström, Pontus

Reality Capture (RECAP) project final report

Published: 01/01/2020

Document Version
Publisher's PDF, also known as Version of record



Reality Capture (RECAP) project final report

30.1.2020

Olli Seppänen

Yu Xiao

Mustafa Khalid Masood

Petr Byshev

Truong-An Pham

Antti Aikala

Pontus Lundström

Contents

Contents

1. Introduction

The project started from the observation that laser scanning and photogrammetry related technology has matured but their applications in construction have been limited to very few use cases, such as manual comparison of as-built point clouds to the design 3D model to evaluate quality and progress. Our previous research project, Intelligent Construction Site (iCONS), increased the appetite of construction companies for real time data by demonstrating that technology can detect wasted effort which cannot easily be identified by humans. ICONS was not able to solve the problem of automating production control processes because it was based on sensors carried by people or attached to materials or equipment. Although it was possible to evaluate where people, materials and equipment were and analyze their movements, images are required to know what those resources achieved and how well the work was done. The goal of the research project, Reality Capture, was to investigate opportunities for automatic monitoring of progress, automatic quality checking of construction and visualization of progress utilizing Augmented Reality technologies.

The research project was funded by Business Finland, Aalto University and a consortium of companies: Fira, Rudus, YIT, Vionice (acquired by Vaisala during the project) and Umbra. The project's steering group included Otto Alhava (Fira), Sakari Aaltonen (YIT), Mikko Kuusakoski (YIT), Mika Tulimaa (Rudus), Markus Melander (Vaisala), Sampo Lappalainen (Umbra), Matti Vaaja (Aalto), Antti Peltokorpi (Aalto), Yu Xiao (Aalto) and Olli Seppänen (Aalto). Results were reviewed quarterly in steering group meetings.

We set out to investigate the feasibility of using currently commercially available hardware to capture data with minimum human intervention to get an accurate picture of the status of construction project. Because the problem is very wide and possible applications are unlimited, we selected five different applications for research. The applications were selected by the three end-user members of the consortium – 1) Fira, a general contractor also developing technology, 2) YIT, the largest construction company in Finland who have dedicated resources looking at new technologies and 3) Rudus, an innovative precast and ready-mix concrete manufacturer. For each case, we evaluated different methods for data collection and developed prototype algorithms to automatically understand reality. One of the applications resulted in a provisional patent and several applications were novel enough to warrant publication of the utilized method in scientific conference or journal papers.

In addition to end user companies, our consortium included a graphics software technology company Umbra and a computer vision company Vionice (acquired by Vaisala during the project). They were planning to use the findings of the project to develop commercial products. However, it turned out in the project that the data collection problem is much larger than the technical problem of designing the algorithms. Regardless of case, practical obstacles hindered the standardized collection of image data. The issues encountered were not technical but rather process related. The current construction process is not well suited for standardized collection of data and in order to enable machine vision approaches, a systemic change of construction process is required. Because of data collection issues, it is presently difficult for technology companies to create a business case for computer vision in buildings. Either the technology companies would need to be heavily involved in data collection, for example with services, or the construction process should change first.

This final report first describes the requirements for Reality Capture from the Finnish consortium and our international partners. Secondly, we describe how the use cases were selected. The results related to each use case are described next, including details related to algorithm implementation, data collection issues and practical guidelines for commercial companies. Finally, the results are compared to goals and future development directions are introduced.

2. Requirements for Reality Capture

We interviewed the consortium companies and companies from United States / California (3 companies), Brazil (4 companies) and China (3 companies). The companies were asked about the main challenges in their production control and quality inspection processes, do they use any real time tools for monitoring progress or quality control, what would be the expected benefits of real-time information based on images and/or laser scans, and which use cases would be most interesting to them and potential interest to validate the results of research.

The main challenges were found to be quite consistent between countries. All companies agreed that it is currently impossible to get real-time and precise data. Lack of real time data means that it is hard to keep people working on the right things and prevent them from starting work that they are not supposed to work on. This leads to a very fast initial round of construction but a lot of working time goes to “pick-up work” where something was wrong or something was missing and the work did not get completed. In Brazil, there were additional issues with quality inspections because they cost a lot of money and it is hard to get trained inspectors, leading to low quality. Everyone agreed that both production control and quality processes are very manual, tedious work where it is very easy to miss something critical. People see things that are wrong but do not report them.

Some digitalization of processes has been attempted by companies. In Finland, Congrid is a software used for by Fira and YIT for quality inspections. Fira has developed SiteDrive to allow for self-reporting of progress by the workers. In Brazil, California and China, similar tools digitalizing quality issues were found (e.g. Snagger, Autodoc, Procore, BIM 360 Field). In California and China, Robotic Total Stations and laser scanning was done a lot to verify geometry. In addition, virtual mock-ups based on VR technology were used by a company in California to make quality requirements clear for workers.

All respondents emphasized the importance of checking productivity, for example by having a time series of images from the same location. Augmented Reality applications where images taken in the same location at different times are added on top of reality would allow the user to know what is inside a structure. Especially quality checking Mechanical, Electrical and Plumbing systems would be important for construction companies because that scope is not understood well and it is hard to evaluate progress or quality manually. Comparison of plans to actual came up from each country. A commercial system should be able to say at the minimum if an element has been installed and if it is in the right location. However, it would be very beneficial also to look for visual quality issues like colors, gaps etc. which are typically not in a BIM model and which need to be inspected manually by an experienced inspector. Overall, the requirements were found to be consistent between countries and there were no automated commercial solutions, except for comparison of point clouds and BIMs where several software packages were being piloted in California but they were not yet in wide commercial use.

3. Selection of case examples

Because of limited budget the consortium agreed in the beginning that we will focus on carefully picked specific example problems where both images and point clouds can be utilized and both productivity and quality aspects are considered. Each end-user company selected two case examples for consideration and we tried to ensure that all aspects are included.

For progress controlling, we had three case examples. Fira selected two examples where the progress in bathrooms and kitchens of a plumbing renovation project should be automatically evaluated based on images only, without any access to design information. These types of projects do not typically have BIM models, so using design information would be impractical. In contrast, YIT selected a case where there was a BIM model and point clouds were created based on two top-down cameras mounted on a crane. The goal was to evaluate the progress of structural works by using the images and / or point cloud captured by the technology. For progress analysis, the case study selection worked out well because we had two completely different challenges to solve and the problems were slightly different from problems that have been solved before. Earlier algorithms for point cloud based progress use point clouds from laser scanners which provide a very dense point cloud. The crane camera case required new innovations because crane cameras create a very sparse and incomplete point cloud but give very detailed information in images taken from top down.

Regarding automatic quality control, we considered two cases. YIT selected installing molding as a case because there are many quality errors, such as gaps or bumps in molding or door frames which get noticed by tenants or apartment buyers. A lot of molding is installed each day, so it is easy to become blind and not notice every single quality error. The idea was to automate the quality control process so that workers could take a video of each apartment after they finished the work and the application could automatically highlight potential quality issues. Rudus selected the reinforcement of precast concrete stairs as their main point of interest. Each staircase has a lot of rebars and it is possible that some are missed before pouring concrete. The task was to count different types of rebars so that the numbers could be compared to reinforcement schedule (case 1) and to compare rebars to BIM models of complex staircases to highlight if something was missing (case 2). These cases related to quality were sufficiently different to cover many interesting aspects: 1) quality without BIM model requiring classification of images 2) quality without BIM model requiring object recognition from images 3) quality with BIM model requiring comparison of images with BIM models.

For each case, data collection was first planned by trying to find an easy way to collect data that could be scaled in a commercial solution. Then data was collected and, if required by the algorithm, classified. Different algorithms were then tested to see how good results could be achieved with the data we were able to collect. Finally, a way to visualize the results, preferably by using some form of AR approach was developed as a prototype. The following section describes the results of each case example.

4. Case study 1 – Detecting progress in bathroom and kitchen renovation

4.1 Overview of requirements

Construction companies collect photos of different stages of a renovation process but such photos often contain only the timestamps and the name of the project as meta-data. We are looking for a system that extracts the temporal logic from the photos of ordered classes to provide relevant statistics of a construction process, in our case - bathroom/kitchen renovation. The goal was to have construction workers collect photos with a data collection mobile application. The progress would then be automatically assessed in each room and the site manager could use a monitoring tool to remotely evaluate progress and make necessary planning and management decisions (Fig.1).

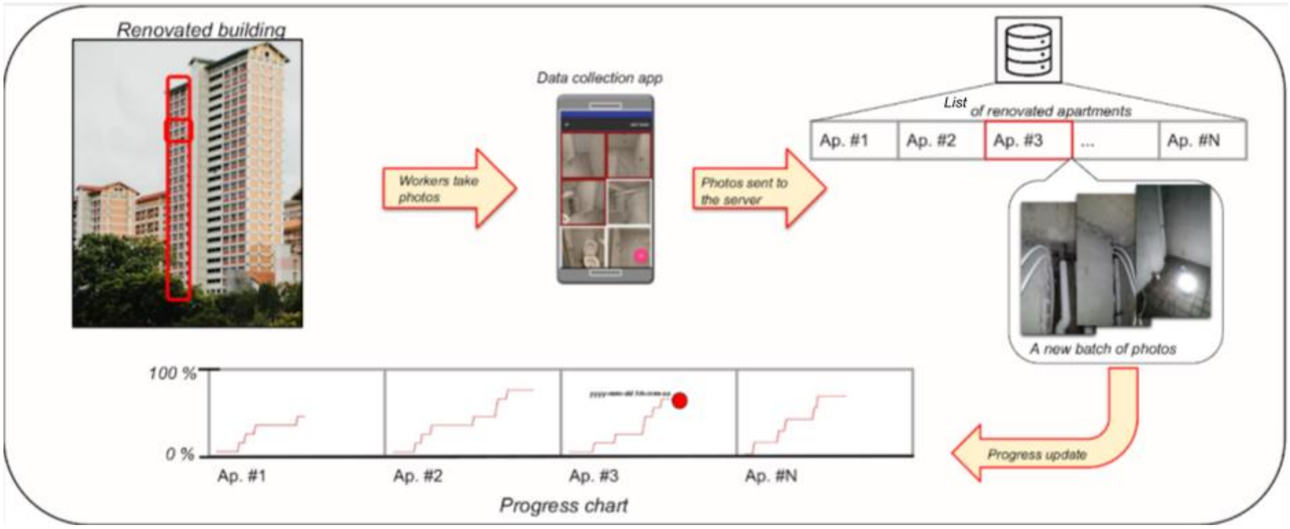


Fig.1 Progress inspection system.

4.2. Developed solution: Deep learning algorithms for progress detection.

We developed a computer vision-based progress inspection system that provides sequential stage identification. The system processes every new batch of photos associated with specific attribute (for example: location, apartment number, date and time,...etc.) and outputs the progress update in real time. In the core of the system is a deep learning model that requires a substantial amount of training examples that cover the whole process. For this purpose we developed an Android data collection app (sec. 4.3). During the training the model is given pairs of images with their temporal order information. The learned temporal dynamics of the process helps to recognize the visual changes in the renovated environment and estimate the progress of the renovation. The proposed solution improved recognition accuracy for both bathrooms and kitchens: from 41.2% to 48.4% and from 37.6% to 42.3%. A more pronounced improvement is observed in Kappa Index which reflects the temporal precision: from 0.64 to 0.80 and from 0.70 to 0.77. Kappa Index varies from zero to one, zero means full temporal disagreement and one means full agreement.

Table 1. Interpretation of the Kappa Index.

Kappa Index	Strength of agreement
<0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very Good

Table 2. Bathroom results

Method	Accuracy	Kappa Index
Baseline CNN	41.2%	0.64
SIFT	27.3%	0.35
Proposed Solution	44.4%	0.80

Table 3. Kitchen results

Method	Accuracy	Kappa Index
Baseline CNN	32.6%	0.70
SIFT	22.3%	0.34
Proposed Solution	42.4%	0.77

4.3. Data collection and validation

Android app for data collection

We developed an Android data collection application, which is designed for taking and labeling photos on-site. Fig. 2 illustrates the user interface: the data collector can take photos and associate them with the corresponding location and renovation step.

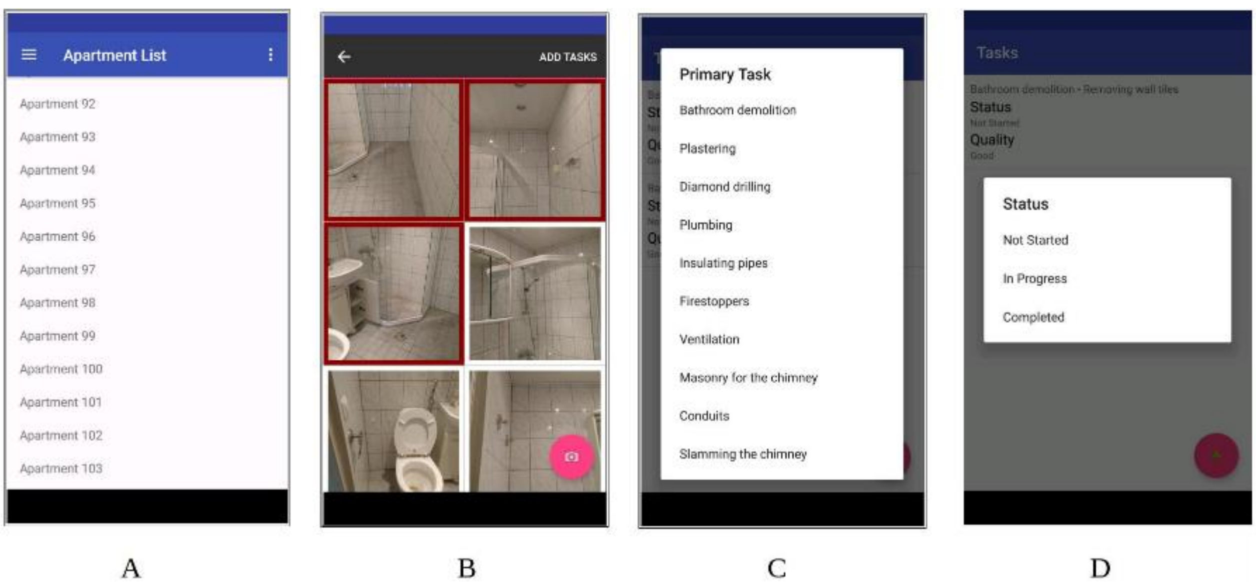


Fig. 2. User interface of Android application for collecting data. (A) Apartment names list. (B) Grid of bathroom photos. (C) Task adding action. (D) Progress stage selection list.

Description of datasets

The Reno-2018 dataset consists of two parts: bathroom photos and kitchen photos, each covering a different renovation process. The bathroom part is composed of photos, taken from 7 bathrooms at 10 different renovation stages. The kitchen part covers 8 renovation stages observed in 6 kitchens. All photos were taken during February-March of 2018. Samsung SM-G920F mobile camera (resolution 5312×2988) and Sony E6653 (resolution 3840×2160) were used to take photos. The photos were labeled by construction specialists according to the internal renovation documentation norms. After the preprocessing, the dataset contains 3708 bathroom and 1486 kitchen images.

Fig.3. Example photos of 10 stages of bathroom innovation.

Fig.4. Example photos of 8 stages of kitchen innovation.

4.4. Challenges and learnings

We see at least three challenging aspects of the datasets. First, the division of a continuous process into stages can be subjective, but such discretization is necessary, as the data collection was performed in batches. Second, a stage can contain visual features of the previous stage, for example, parts of the conduits system may be visible during the laying the concrete floor stage (See Fig. 3 for examples). Third, due to the practical aspects of the renovation, the stages are not strictly ordered: some classes may include images with visual clues from a different temporally adjacent class.

The Reno-2018 formulates an ordinal image classification problem – a discrete work progress estimation task. One alternative option was to define the task as a regression task: for every photo (or batch of photos) provide a % of completed work. The choice of defining the task as a classification, rather than a regression problem was made based on the following reasons. First, the image classification deep learning methods are more developed and understood. Second, photo acquisition was performed in irregular time slots, resulting in a more chronologically fragmented dataset. And third, the true completion degree holds ambiguity that is difficult to control for, namely should we take the actual time passed since the initial phase, or should we use an educated but still subjective progress estimation given by a construction manager.

In total we collected $\sim$7000 photos from 7 bathrooms and 6 kitchens that were originally distributed across 30 classes. Majority of the classes were underrepresented, having less than 50 photos. This led to the decision to dismiss some classes and merge similar consecutive classes. We also kept the initial and the final stages, despite them being underrepresented.

The modern computer vision techniques rely strongly on high quality labeled data. The difference in performance for the bathroom and kitchen cases indicates the necessity of more dense representation of the stages. To this day, the proposed deep learning models require powerful GPUs which makes usability of the methods limited to the specific set-ups.

4.5. Guidelines for developing a commercial solution

An effective commercial solution requires a streamlined data collection routine. The proposed data-driven methods can benefit from new incoming labeled data. The launch of the application should include a "beta-testing" period where potential users would correct for false stage identification. Such a feedback loop would greatly enhance the model performance.

5. Case study 2 – Progress detection based on crane camera images

5.1 Overview of requirements

The objective of this case was to study the potential of crane cameras for automated progress monitoring of construction sites. Crane cameras offer a fully automated, convenient and cost-effective alternative to other technologies such as laser scanners and drones to capture visual data of construction sites. However, their utility for progress monitoring remains largely unexplored. YIT installed a crane camera system developed by Pix4D which generates 2D images and 3D point clouds. We were required to automatically extract useful information from the 2D and 3D data that could be used to infer construction progress.

5.2 Data collection

Two cameras were mounted on the jib of a crane on the Tripla construction site in Helsinki. The cameras snapped pictures when the jib of the crane moved. The resulting 2D images offered an overhead perspective of the building under construction. Once enough images are collected for a day, the Pix4D solution automatically creates a 3D point cloud from the images. The images and point clouds were stored on a cloud server. For the VBUILT algorithm (Section 5.3.1), 40 point clouds spanning August 17th, 2018 to November 23rd, 2018 were used. For the MBUILD algorithm (Section 5.3.2), eight point clouds from the same duration were selected. For the solution described in Section 5.3.3, 84 images were selected.

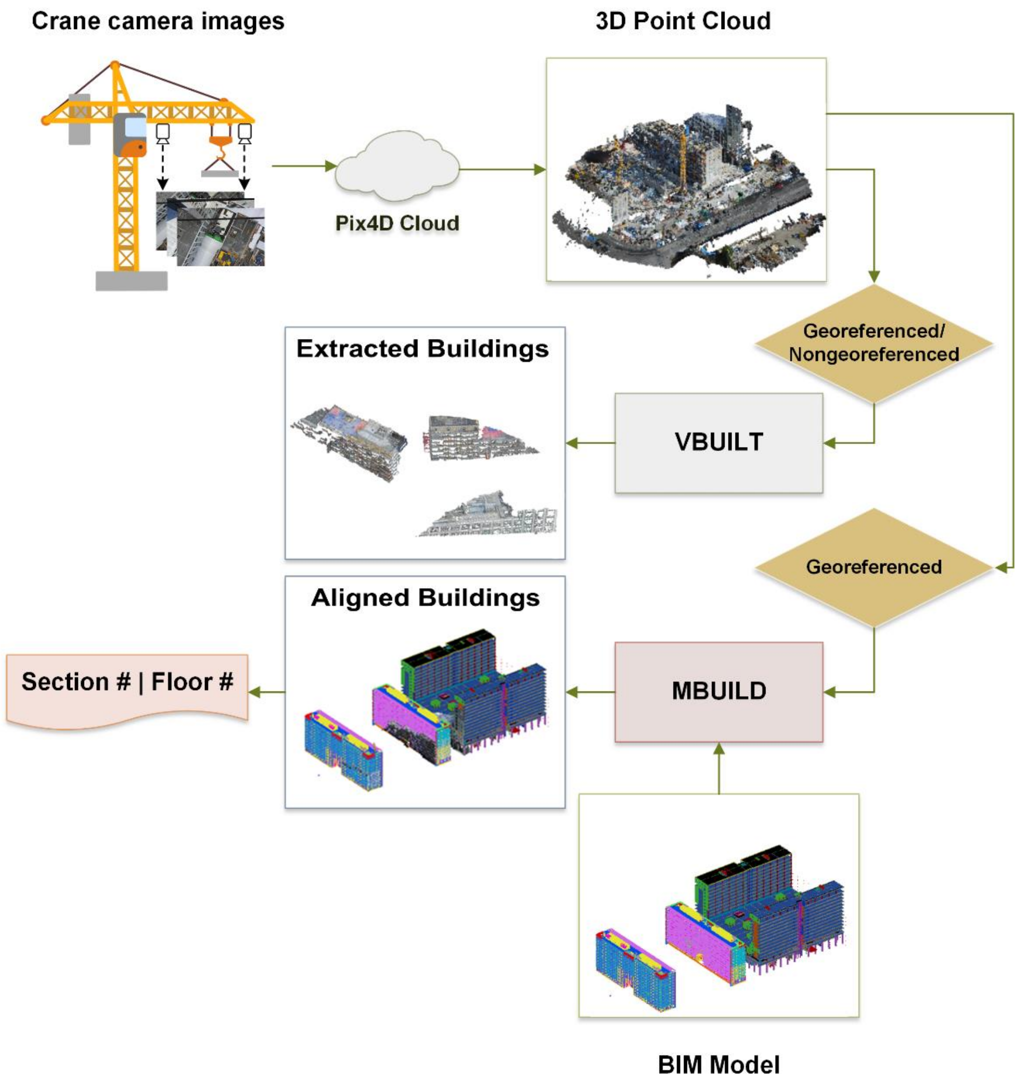


Figure 5: Algorithms (VBUILT and MBUILD) for processing crane camera point clouds

5.3 Developed solutions

The solutions for processing the 3D point clouds are illustrated in Figure 5 and explained below.

5.3.1 VBUILT: Volume-based Building Extraction for As-Built Point Clouds

This algorithm was designed to automatically extract buildings from the construction site point cloud for both georeferenced and non-georeferenced point clouds. The basis of the solution was the idea that buildings generally have larger volumes than non-building elements. First, we extracted the ground plane of the construction site using a model fitting algorithm called MSAC (M-Estimator Sample Consensus), followed by clustering the above-ground point cloud using Euclidean distances, after which we determined the volumes of the 3D convex hulls (essentially, an outward wrapping of the 3D clusters). The largest convex hull volumes were labeled as buildings and the non-building elements were then removed. The algorithm thus outputs a construction site point cloud containing only the buildings. Testing the solution on 40 point clouds spanning August 2017 to November 2017, the accuracy was 100%, meaning that buildings were always correctly distinguished from non-building elements.

5.3.2 MBUILD: BIM-based building extraction and alignment for multi-building point clouds

In order to effectively use the BIM model, this solution requires that the point clouds are georeferenced. The core idea of the algorithm is to use the BIM point cloud (obtained by converting the BIM model) as a *building-pass filter* so as to remove the non-building portion of the site. First, the as-built point cloud is converted into the coordinate system of the BIM model. Then, building-pass filtering is used to remove most of the non-building portion of the site. Ground removal is performed using the z-histogram of the base portion of the building, which is much faster than the MSAC-based ground removal used in VBUILT. Finally, alignment is performed by aligning the xy projections of the BIM and as-built point clouds using the standard Iterative Closest Point (ICP) algorithm, while the yz projections are aligned by matching the edges of their bounding boxes. Note that building-pass filtering is performed strategically at multiple stages in the algorithm to obtain a good delineation of the building. This algorithm was tested for eight point clouds. All the point clouds were successfully filtered and aligned. From the aligned cloud, we calculated the floor number by observing the height of each section of the building. The results are shown in Table 4. Note that floor estimation was performed for two sections of one building only, since the point cloud corresponding to the other sections of the building and to the other buildings was too fragmented due to incomplete coverage by the crane cameras.

Table 4: Floor estimation results

		Actual floor		Estimated floor	
PointCloud#	Day (dd/mm/yyyy)	Section 1	Section 2	Section 1	Section 2
1	17/08/2018	8	6	8	6
2	20/08/2018	8	6	8	6
3	01/10/2018	10	9	10	9
4	01/11/2018	11	10	12	10
5	05/11/2018	12	10	12	10
6	07/11/2018	12	11	12	11
7	13/11/2018	12	11	12	11
8	19/11/2018	12	11	12	11

The output of this algorithm for a single building was fed to the UMBRA viewer. Figure 6 shows the results.

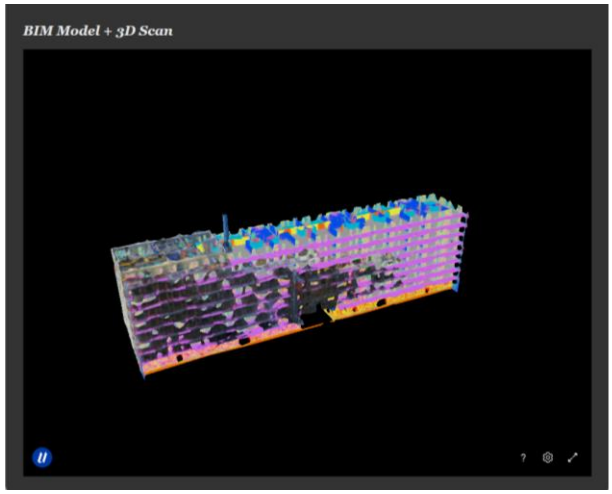


Figure 6: Aligned point cloud viewed on UMBRA viewer

5.3.3 Slab state recognition from crane camera images

Once an aligned point cloud is achieved, we could automatically map from specific points on the point cloud to the 2D images used to create it. The images provide a clear view of the formwork and rebar placement on the precast concrete slabs of the buildings. We implemented a strategy to automatically infer the state of the slab, whether it contained no formwork, formwork only, both formwork and rebar or completed rebar. We framed this as a scene recognition problem and applied the Bag-of-Visual Words technique to train a Support Vector Machine (SVM) classifier to classify the slab state based on the slab images. With only a small number of images (see Table 5) of each slab state needed for training, we achieved a testing accuracy of 100%. The slab images that were input to this algorithm were prepared manually. Figure 7 illustrates this solution.

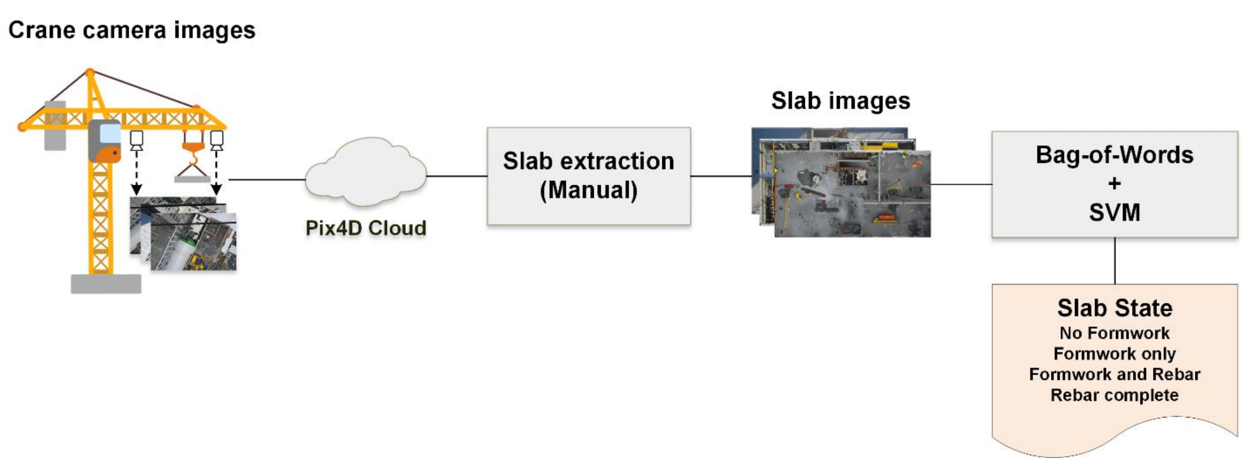


Figure 7: Slab state recognition solution

Table 5: Size of complete image set and training set for slab state recognition

	Count	
Label	**All images**	**Training images**
No Formwork (NF)	39	3
Formwork (F)	30	4
Rebar (R)	7	1
Rebar Complete (Rc)	8	1

5.4 Challenges and learnings

Due to incomplete coverage of the construction site by the cameras, the point clouds were **fragmented**. The most complete portion of the point cloud showed half of the façade of a building and a good view of the roof slabs. The rear was completely missing since it is not within the camera's field of view. Only a small segment of other buildings was available. This made analyzing the point clouds challenging and limited the information that could be extracted. Future projects should consider how more crane cameras can be placed to extend coverage or examine a possible integration with drone images or images from fixed cameras.

Since the **data for the early stages** of the building was not available, we could not test the algorithms for these stages. The VBUILT algorithm especially needs to be tested for when the volume of the building may be less than the surrounding non-building elements. Identifying cranes using their typical geometrical or color features may be required at the early stages to augment the volume-based analysis.

In some cases, point clouds were not georeferenced due to a **failure of image geotagging**. Aligning non-georeferenced point clouds with the BIM model is difficult and unreliable. It is more practical to ensure georeferencing rather than invest resources in aligning non-georeferenced point clouds.

There were large **variations in point cloud density** in some cases. Typically the point cloud density was around 1000 points/m3, but in some cases dropped to 3 points/m3 and in another case increased to 37,233 points/m3. For the VBUILT algorithm, this required adjusting some parameters to work effectively for different densities. Ensuring that the point cloud density stays within a predictable range would help avoid unpredictable results.

The as-built point cloud was slightly **out of scale** with respect to the BIM model, which affected the accuracy of floor estimation. Automatically equalizing the scale of the as-built point cloud and BIM model should be addressed in future work.

Processing the point clouds is usually **computationally intensive**. Computers with powerful General Processing Units (GPU) should be utilized to avoid long processing times.

The UMBRA viewer needed a suitable '**connecting radius**' setting to create a viewable model. The radius needs to be large enough (or the point cloud density needs to be high enough) for the system to connect points to form surfaces.

5.5 Guidelines for commercial solution

A few processes that were manually performed would require automation for a commercial solution. Point clouds and images should be automatically pulled from the cloud to be fed to the algorithms. IFC files should be automatically converted to point cloud format (.las or .ply). An automatic mapping from points on the point cloud to the corresponding 2D images would need to be developed. For the UMBRA system, a strong GPU is needed to avoid Z-fighting in the viewer.

6. Case Study 3 – molding defect detection

The aim of this case study is to recognize bad quality molding from the photos. Often workers become 'blind' to such defects therefore an automatic computer vision solution is necessary to assist in quality control. In the pilot study we had access to 1556 'good quality' examples and 374 'bad quality' examples (table 6).

Table 6. Distribution of photos

	Number of photos
Good quality	1556
Unlevel	10
Gap	202
Ending	19
Damaged	89
Bump	35
Other	19

The accuracy for the binary classification between good/bad quality examples was 84.5%. The limitation was mainly caused by the inconsistency in the data: low number of bad quality examples and their high variety. The defect localization showed partial success (Fig. 8) correctly identifying the area of the defect, but for a more consistent result a substantial amount of labeled data is required. The applied localization method only used the label information and only used correlation statistics to detect defects. That means that the method finds the common visual aspects of the images that contain defects and bases the localization on those statistics.

In order to solve the problem robustly, the data collection strategy requires two improvements. Firstly, more bad quality examples that would represent different error types more completely are required. Secondly, it is a requirement to get more detailed information on the location of defects. With more photos of different type, the scalability and the ability to recognize different types of defects could be improved. With more detailed information on the location of defects, it would be possible improve the precision of defect localization approach.

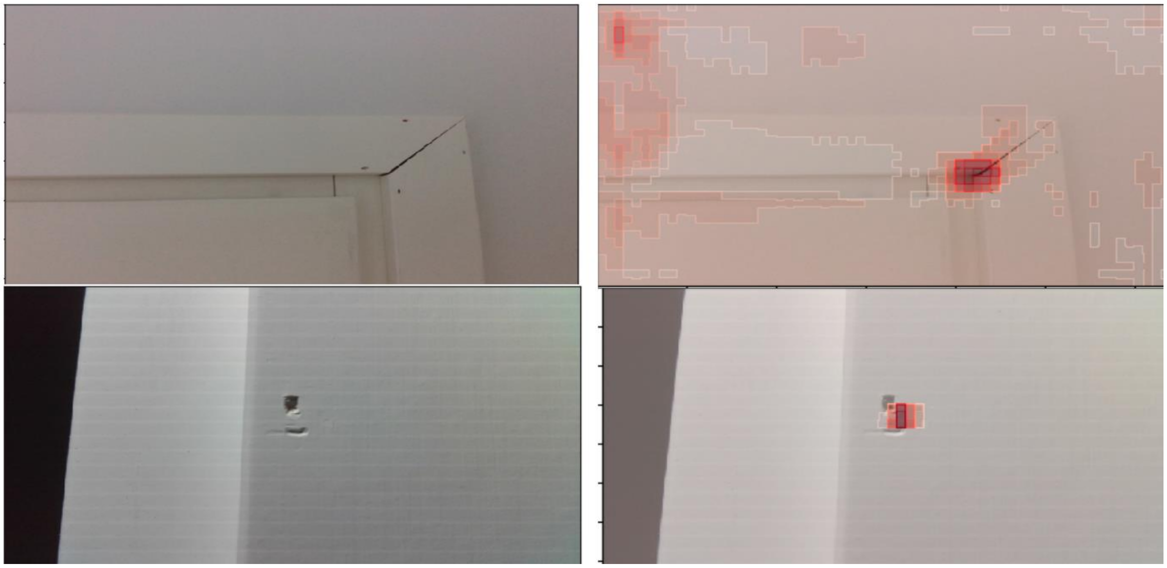


Fig. 8. Defect localization results.

7. Case study 4 – Rebar detection in precast elements

7.1 Overview of requirements

In this case case, the precast staircase factory is interested in an automatic inspection system in which the installed rebars are recognized. The recognized rebars can be compared with the design and the dimensions of the staircase can be measured automatically for estimating the volume of the concrete poured into the staircase rebar frame. We are looking for a system that can localize the staircase of interest in the factory, and then recognize its currently installed rebars.

7.2. Developed solution

Algorithms

As illustrated in Figure 9, the proposed solution contains three step modules. First, the staircase localization module is responsible for detecting the staircase in the photo. The core of the module is implemented by Mask RCNN. After localizing successfully the location mask of the staircase, Rebar detection and classification module will recognize the current rebars in the area of the staircase. In this second module, Mask RCNN is also implemented. We separate staircase localization and rebar detection to increase the accuracy of rebar detection because we can narrow down the search regions by just looking at the localized staircase. The final step is to compare known setup of rebars of the staircase design with current detected rebar installation to estimate the percentage of completion of rebar installation.

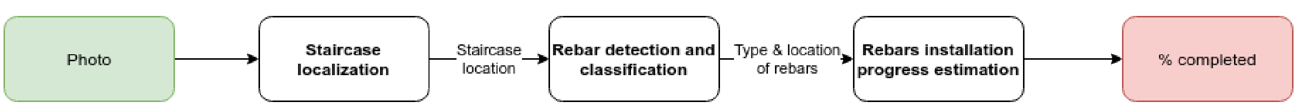


Figure 9. Algorithm for inspecting the progress of installing rebars on staircases.

Web services

Illustrated in figure 10, the solution consists of four main parts: image capture and transfer from the Rudus factory, processing and information storage on an Aalto server, 3D model optimization and viewing provided by the Umbra Composit platform, and an end-user web application.

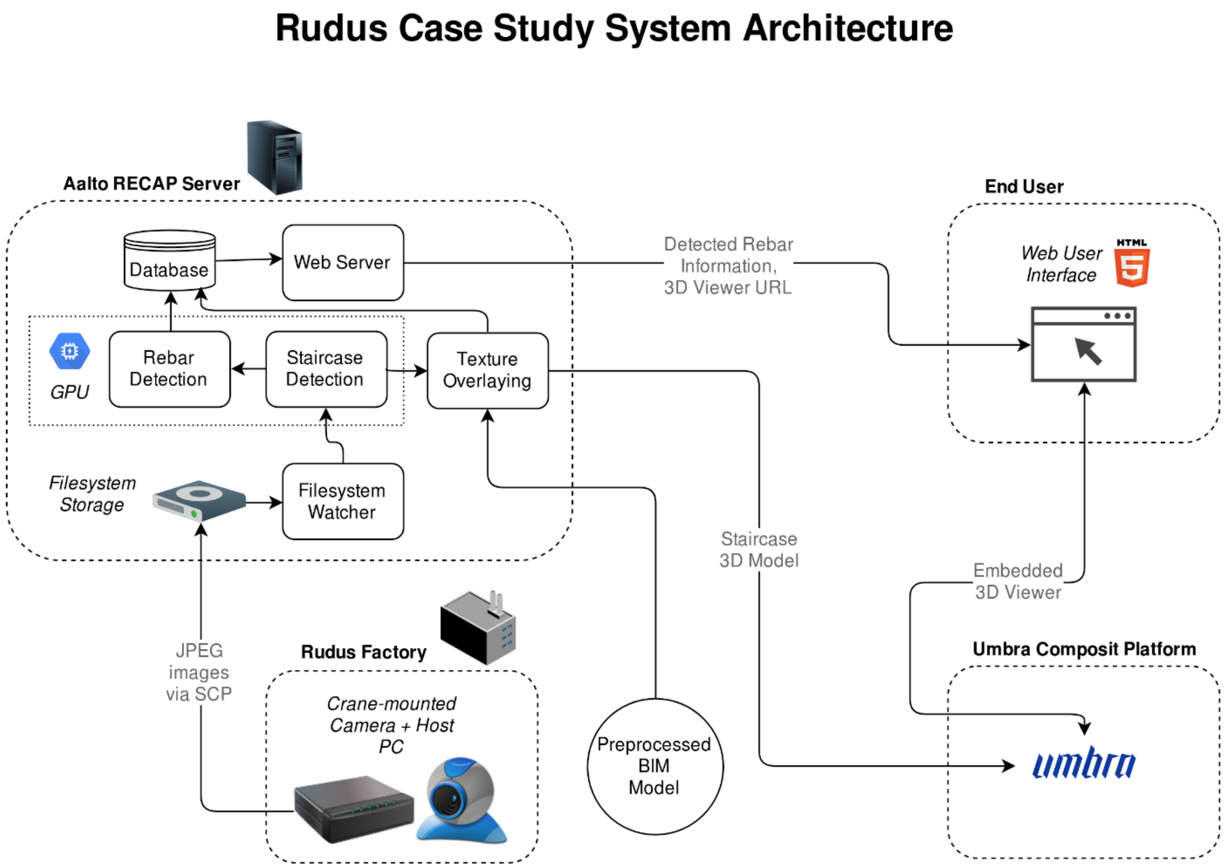


Fig. 10. Architectural overview of the developed solution in the Rudus case study.

At the factory, images are captured periodically (for details about the camera setup, see section 7.3) and transferred to a server managed by Aalto via Secure Copy Protocol (SCP). The received images are fed to a machine learning algorithm that extracts an image of the staircase and detects the individual reinforcing bars.

The cropped and aligned staircase image from the staircase detection stage is attached as a texture onto a 3D model of the staircase (obtained by manually preprocessing a provided BIM model in IFC format) and the model is uploaded into Umbra Composit, a cloud platform for sharing 3D datasets using web technologies . Results from the rebar detection stage (the counts of three types of rebar) are stored in a database alongside with the respective Umbra 3D viewer URL. Finally, a web application, a screenshot of which is shown in figure XZ, fetches the information about detected rebar and embeds the Umbra 3D viewer to show progress to the end user. (Figure 11)

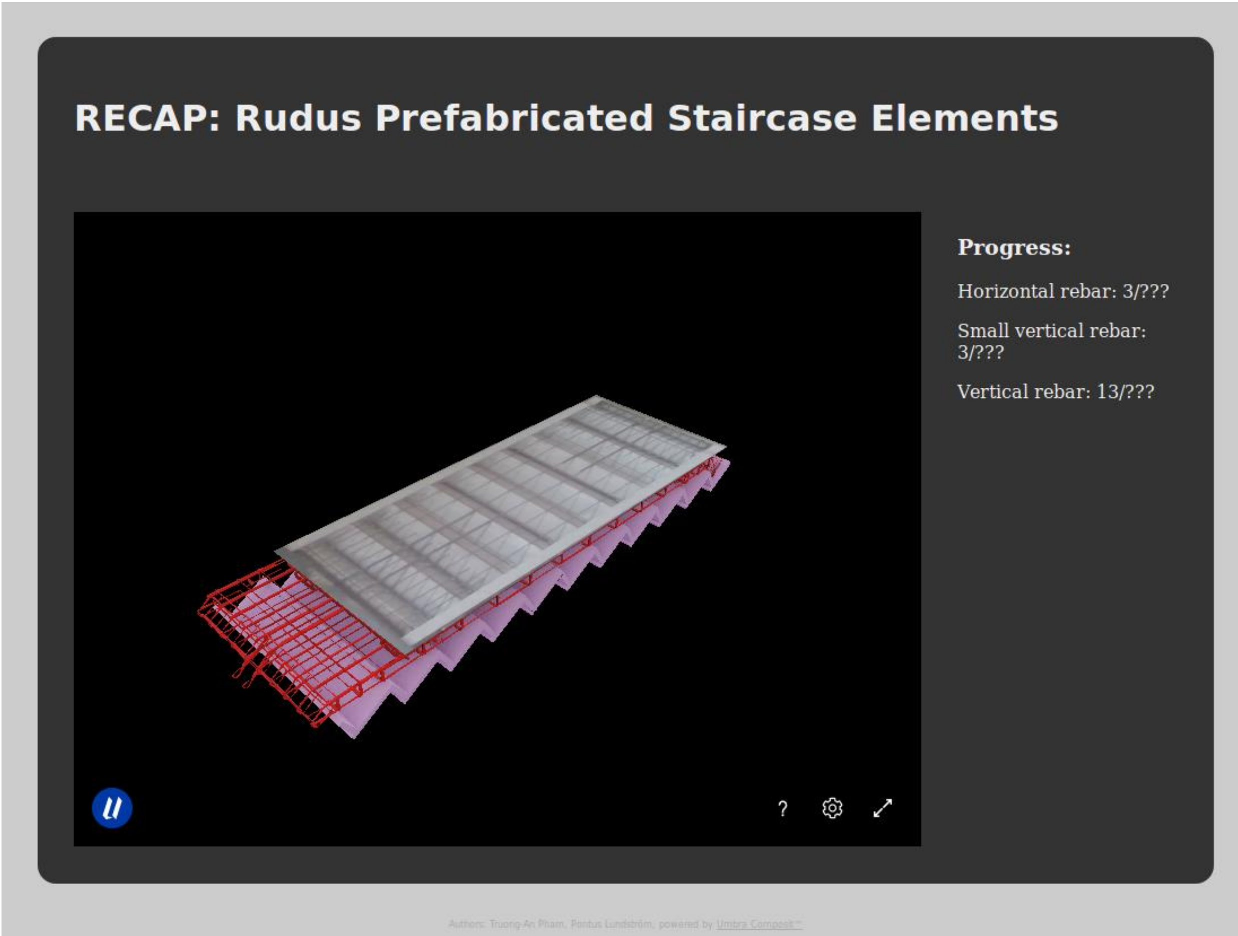


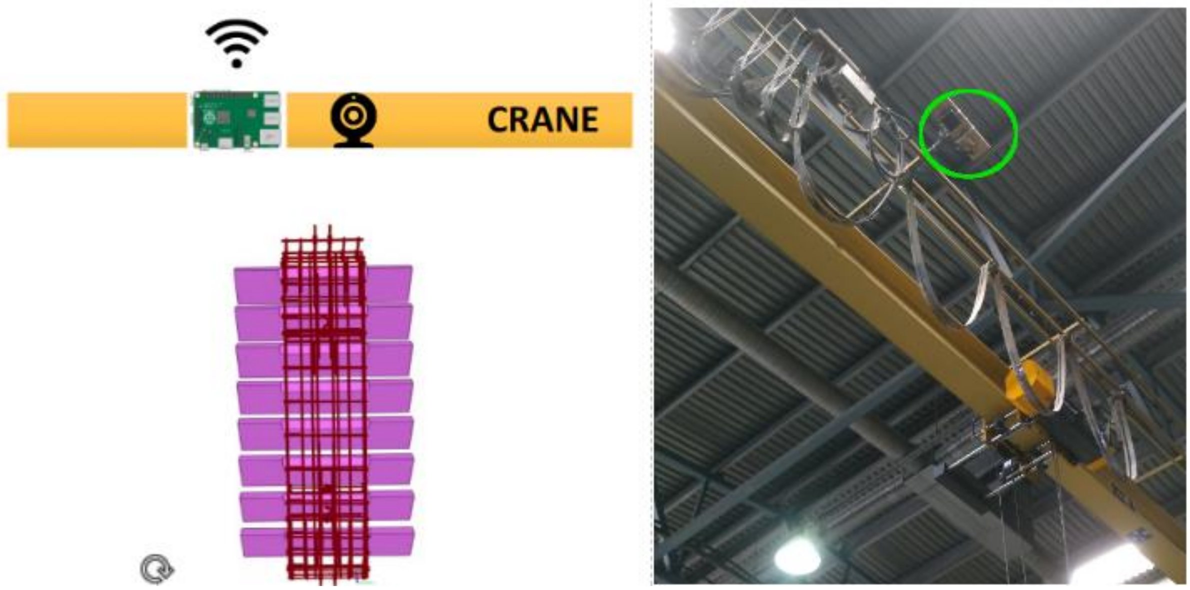
Fig. 11. A screenshot of the web application developed for the Rudus case study.

6.3. Data collection and validation

Camera setup

The camera system includes a mini PC for taking captured photos from camera and a depth camera for capturing photos in the factory (Figure 12). The camera was setup on a ceiling-mounted crane so that when the crane moved in the factory, the camera would be correctly positioned to take pictures of the element station (Figure 13).

Figure 12: The mini-PC and camera installed in the factory

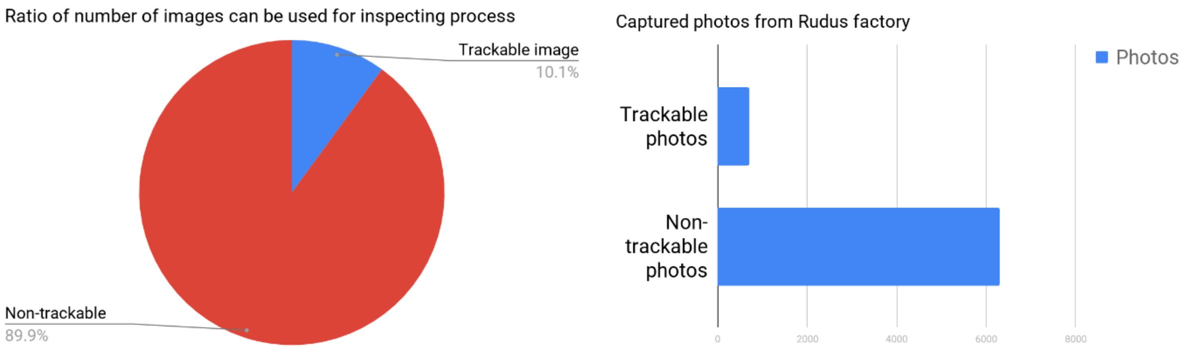


(A) **(B)**

Figure 13 Camera setup on indoor crane of the factory. (A) 2D sketch of camera setup in the factory. (B) Real setup of the camera system in the factory.

Collected datasets

We have collected and labeled ~6500 photos, but we used just 1340 photos for training and testing only due to budgetary constraints (670 photos contained the staircase, and 670 photos contained no staircase). Most of the raw images did not contain a staircase (Figure 14). In addition to the labeled photos, 290 000 unlabeled photos were automatically collected.



(A) **(B)**

Figure 14 collected and labeled dataset in Rudus case. (A) A pie chart for representing ratio of captured photos containing staircases of interest over captured photos containing no staircase of interest (Red one or non-trackable). (B) Bar chart represents the number of captured photos of each type.

We labeled manually 670 photos with a localized staircase and then labeled the rebars as illustrated in Figure 15. Three types of rebars, including A1-A2-MODIX , X5-Y7 and U3, are illustrated shown in Figure 16.

Figure 15. Manual labeling the data. The left image shows the way we labeled the staircase in the image. The right image shows the rebars labeling output.

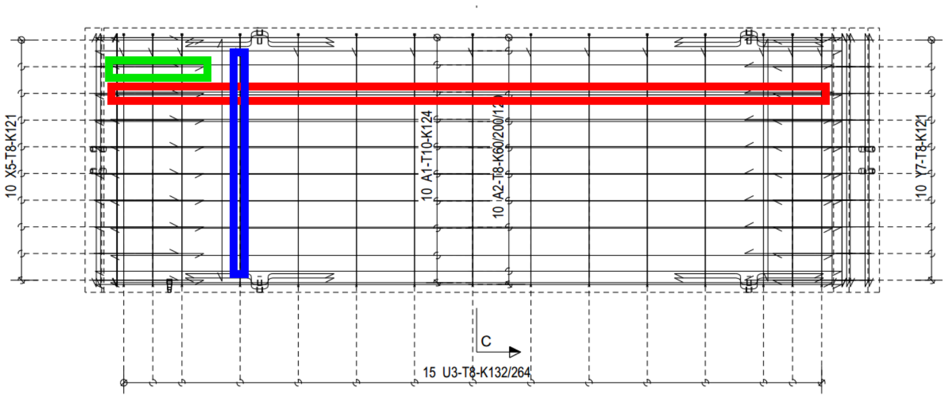
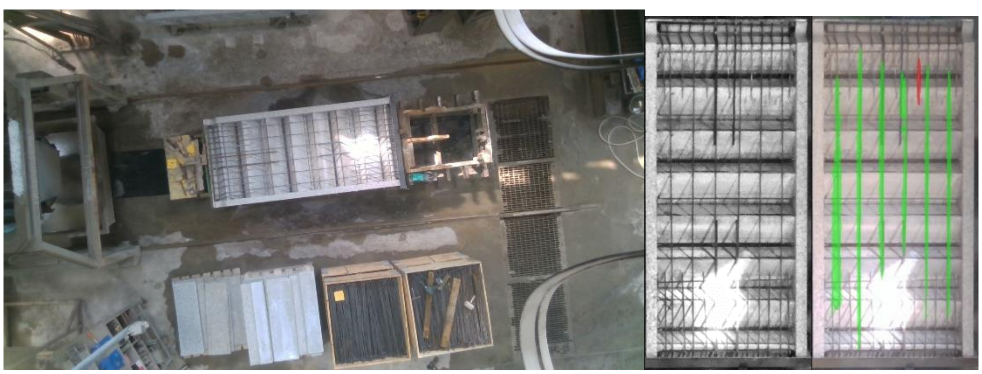


Figure 16. 3 tested types of rebars were A1-A2-MODIX (Red), X5-Y7 (Green) and U3(Blue).

Validation results

We were able to localize the staircase with a very high precision: *(*IOU 0.5: 1, IOU 0.7: 1, IOU 0.9: 0.96.) The rebar recognition resulted in acceptable F1 scores (A1-A2-MODIX: 0.79, X5-Y7: 0.69, U3: 0.63). Figure 17 shows the output from the algorithm.



(A) **(B)** **(C)**

Figure 17. Output from the algorithm. (A) captured photos. (B) Extracted and rotated staircase. (C) Detected rebars is highlighted with red, green, or blue color.

6.4. Challenges and learnings

We figured out three major challenges in setting up the camera in the factory. First, due to high-safety requirements in the factory, cameras system must be mounted on a wall, a ceiling or a crane instead of a standing tripod. So, rebars look quite small in captured images. Second, the camera system is mounted on an indoor crane of the factory, so several captured photos have been blurred while the indoor crane is moving. Third, the crane is not always above staircase of interest, so inspection process is not continuous. Then, we cannot track rebars installment step-by-step.

The current solution is to use deep learning methods for localizing the regions of the staircase of interest and recognizing its rebars. Unlike classical approaches, current supervised deep learning approaches requires a large number of high-quality labeled images. Due to budgetary constraints, we were able to use only 670 labeled images, which resulted in only acceptable level of rebar detection. With more images, the accuracy of the method could be solved.

7. AR Application for Progress and Quality Monitoring

We did an online survey in April 2019 to understand the needs and expectations of the AR application.

In the survey, participants were asked to rank the importance of the following 9 features:

- Detecting defects automatically
- Detecting flaws immediately after completing each work step
- Monitoring the progress of renovation (time spent at each stage)
- Checking what has been installed or removed at each location of interest
- Viewing the design and project schedule
- Sharing the flaws or progress information with your colleagues
- Generating work reports automatically
- Displaying assembly or writing PDF instructions
- Marking defects manually

Based on the survey results, we selected 3 key features to implement in the AR application. First is to monitor the progress of renovation, including detection of current stage and calculation and visualization of time spent in each stage. Second is to allow users to manually mark defects. Third is to allow users to share information with colleagues.

The application allows data collection and detects the work stage automatically from the picture. All pictures from a location are stored and shown in a timeline. The user can manually correct the classification of any incorrectly detected image which will be then used to further train the algorithm (Figure 18)

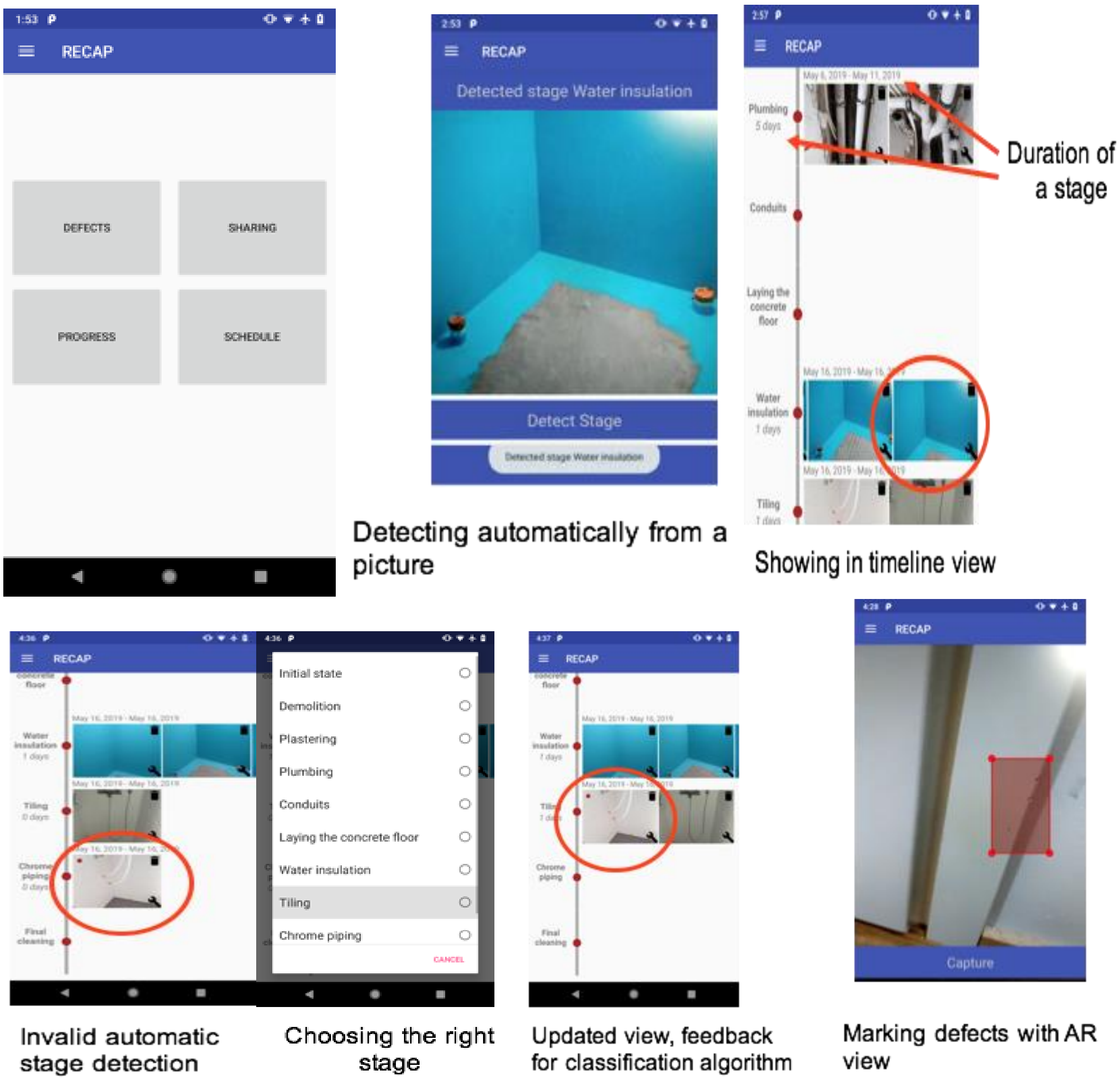


Figure 18. Functionality of the developed application

8. Conclusions

The goal of the research project, Reality Capture, was to investigate opportunities for automatic monitoring of progress, automatic quality checking of construction and visualization of progress utilizing Augmented Reality technologies. Due to limited research budget, we evaluated a few different use cases prioritizing use cases which have not gained a lot of attention in research literature and where data could be collected with inexpensive hardware.

The results show that all use cases were possible to solve technically, although none of the use cases were trivial. The bottleneck was data collection. The data for machine vision does not currently exist and each use case requires a lot of training material if deep learning methods are used. Because there are so many different tasks and location types to be detected, it is very difficult to do a full-scale progress monitoring system by deep learning alone. The same applies to quality detection. We were able to detect and locate quality issues but there are so few mistakes that it requires a lot of data collection to solve the problem for just one work type. We tried to analyze defect image libraries by companies, but images were taken inconsistently and the defects were not appropriately labeled or collected. The recent trend of using helmet-mounted 360 cameras may solve the data issue but requires a lot of manual labeling work. If data problem can be solved, these approaches could be used to evaluate progress in absence of 3D BIM models for locations with standard work, such as hotel rooms, bathrooms, kitchens etc. It is unclear how much progress could be deduced from more complex spaces.

The crane camera case showed promising results. It was possible to quite accurately detect the work phase by using Bag of Visual Words method with just a limited training sample. In future research, the same method could potentially be able to detect work phases inside the building with limited training data required as long as the visual modifications caused by the task are substantial. Even the crane camera implementation suffered from variable quality of data.

As an overall conclusion, machine vision techniques seem to be ready for utilization in construction projects, even when BIM models are not available. Future work should be carried out to investigate more use cases and focus more on economic ways to collect a lot of data, perhaps as a joint effort of construction industry. If data can be collected and labeled in a systematic way, even currently available machine vision techniques could have a lot of potential and help create a digital situational awareness of a construction project.