Crane camera image preparation for automated monitoring of formwork and rebar progress

Antti Aikala*, Mustafa Khalid, Olli Seppänen and Vishal Singh
Department of Civil Engineering, Aalto University, Finland
* email: antti.aikala@aalto.fi

Abstract

Crane cameras are emerging as a cost-effective technology for enhancing situational awareness at construction sites. Crane camera images in combination with point clouds can facilitate automated progress monitoring of sites, which can lead to significant time and cost savings in comparison to manual inspection methods. In this work, we focus on preparing crane camera images for monitoring the progress of formwork and rebar used for laying concrete slabs. Preparation is needed, because the view angle and position of the target slab are changing between images because of crane boom rotation. Furthermore, vertical distance from the camera to the topmost slab changes as the building rises. Our target is to rotate, crop and rescale the original images in such a deterministic way that the slab under construction is always shown in the result images at the same position, orientation and scale. In order to achieve that, we need to know the view angle and the positions of both the camera and the slab. Camera positions can be read from image metadata and, because the crane rotation axis position is known, the boom angle can also be calculated. We noted that the GPS information extracted from the image metadata typically has deviations that could adversely affect the transformation, and that occasionally GPS values are too erroneous to be used at all. We present experimental results from data collected at a large construction site in Helsinki, Finland.

Keywords: Progress monitoring, crane cameras, image transformation, Intersection over Union (IoU)

1. Introduction

Effective control of construction site activities requires frequent monitoring of the site. Many methods have been proposed to monitor construction sites, including handheld cameras Golparvar-Fard et al., 2015, Unmanned Aerial Vehicles (UAVs) and laser scanners. Recently, a new technique to acquire as-built data has emerged, which is the use of crane cameras. The major advantage of crane cameras over other methods is that the data acquisition can be fully automated with minimum cost and effort. Furthermore, crane cameras are non-intrusive to site operations. The photogrammetry software provider Pix4D has developed a crane camera solution for automatically taking and storing 2D overview images, uploading images into a cloud service and generating daily georeferenced 3D point clouds. Although the 3D point clouds can provide the current floor under construction, it is the 2D images that provide the richest information about the state of the slab. Additionally, although there is only one 3D point cloud generated per day, there are multiple 2D images. Thus, construction progress can be monitored with higher temporal granularity using 2D images, and we will focus in this paper on the analysis of 2D images.

One useful type of information that can be inferred from crane camera images is the progress of formwork and rebar. However, in order to effectively perform progress analysis, crane camera images need substantial preprocessing. Two key issues need to be addressed by the preprocessing step: first, the changing *scale* of images as the building is built and increases in proximity to the crane camera and second, the changing *orientation* of the images as the crane rotates about its axis. Both issues can adversely affect the detection of slab edges, which can in turn degrade the formwork and rebar analysis. Additionally, keeping track of the correct scale can allow us to leverage size features (such as thickness) to distinguish rebar from formwork, while knowing that the correct orientation can allow us to leverage the vertical orientation of rebar.

In this paper, we present our method of executing the preprocessing step, including the assumptions we make about the data. We also explore the most important factors for successful preprocessing, evaluate the quality of our preprocessing and discuss possible ways of improving the task. This paper is organized as follows: Section 2 discusses related work over image-based formwork and rebar progress monitoring. Section 3 describes image and other data available for the preprocessing and our preprocessing method. Section 4 presents the results and discussion, quality of input data and evaluations of area detection. Section 5 lays out the conclusions and future work.

2. Related Work

Tuttas et al. (2016) presented a comparative analysis of data acquisition using crane cameras, UAVs and hand-held cameras. They pointed out that crane cameras are the best solution when daily images of building slabs are required, since the data acquisition can be fully automated and the cameras provide a nadir view of buildings. Turkan et al. (2014) presented a method to label column, floor and wall objects in the as-designed model as being in either “built”, “formwork” or “rebar” state. Taking a Terrestial Laser Scanned (TLS) point cloud registered with the BIM model, they noted that the percentage of object recognition varied differently with respect to a Euclidean distance threshold for different states of construction. This method does not capture the percentage completion of formwork and rebar, which is better done using 2D images.

Braun, et al. (2015) proposed a method to combine temporal information of the construction state with geometric matching of the as-built point cloud and as-planned model using precedence relationship graphs. However, this method is also not capable of inferring the formwork and rebar progress of floor slabs. Goldapar-Fard et al. (2015) showed how unordered images and time-labelled 3D plans can be used in progress analysis. One important part in their method is the detection of occlusions. However, the 3D voxel side length is 6 cm that is probably too coarse for detection of rebars which have a typical thickness of only 1 cm. Analysing the percentage completion of formwork and rebar on floor slabs requires a level of detail not present in 3D point clouds, to which previous works have typically limited their progress analysis. However, the details like rebars are clearly visible in 2D images.

The global positions of the crane cameras vary according to the horizontal angle of the crane jib and the installation position of the camera in the jib (Pix4D, 2017). Furthermore, the height difference between the camera in the jib and the target slab affects to the scale of the slab images (Tuttas, et. al, 2016). For the work progress analysis of the slab work, it is desirable to get the nadir view images with fixed orientation and scale that allows direct comparison of the slab area in the original images with the corresponding 2D floor plan. Assuming that the position of the camera and slab are known, this image preparation can be carried out with the geometric transformation using, for example, open source OpenCV package (Bradski, et al., 2008). To the best of our knowledge, this is the first work that deals with the preparation of crane camera images for formwork and rebar progress analysis.

3. Methodology

3.1 Data description

Images used for this analysis were provided by a crane camera solution by Pix4D implemented at a large construction site in Finland. Two cameras were installed on the boom of a tower crane located centrally on the site. The cameras were located approximately 13 and 37 meters from the rotation axis of the crane, respectively and their viewing direction was towards the ground. The vertical height of the crane boom, and thus of the cameras, was about 100 meters from sea level. Images used for this work were taken between 5th July and 2nd November 2018, in total 5361 images from 39 working days. In addition to the 2D images, georeferenced daily 3D point clouds were available on the cloud. The height of the building could be inferred from these point clouds.

The areas of interest were two horizontal floor slabs from a residential building. We labelled the slabs A and B as shown in Figure 1. The two slabs were built at a different pace, so the floor number being built was different for each slab. Selection of the areas of interest was based on visibility of the areas in the pictures, because these areas were closest to the crane and were thus viewed relatively often by the crane cameras. The images contained some useful metainformation for the analysis, such as time, GPS position and camera number. On the other hand, we knew from the BIM model and site plans the positions of the slabs and the crane. Figure 1 illustrates the north direction, positions of the crane, camera movement lines and areas of interest (slabs A and B) on a map that has been rotated in the orientation of the local site coordinate system.

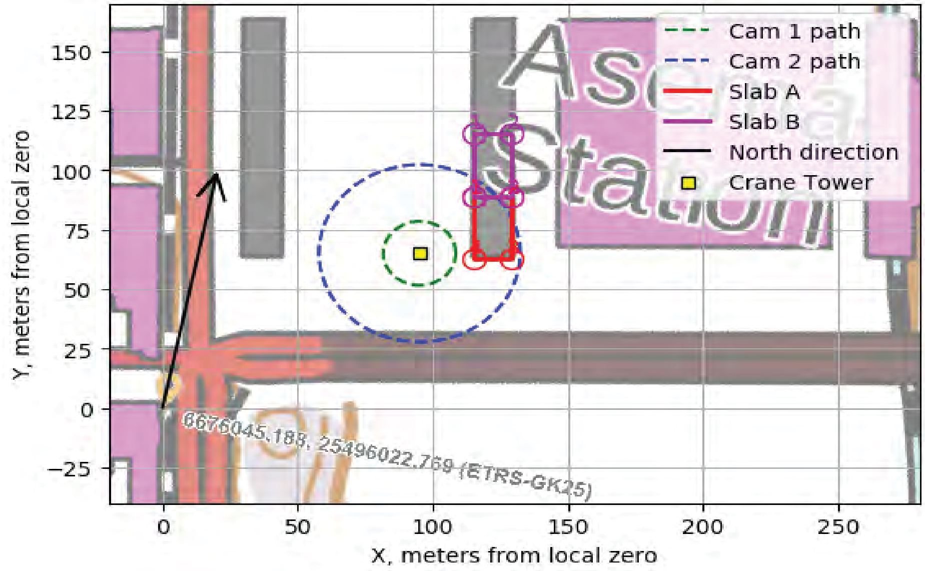


Figure 1. Background map in the local site coordinate system and the key elements for the analysis

Since the cameras were in nadir orientation, the viewing directions of the camera did not change when the crane moved. This simplified the processing, because the sensor could be assumed to be parallel to the horizontal slab plane. Only the positions and yaw angles of the cameras changed according to the rotation of the crane boom. Additionally, it was assumed that the GPS values in the image metainformation were reasonably accurate. Furthermore, we assumed the images were undistorted, since no distortion was visually discernible. For the distorted case, calibration would be needed. The focal length per detector width was approximately 28 mm for 35 mm film.

3.2 Preprocessing algorithm

The preprocessing algorithm selects the desired slab area from the crane camera images. It uses camera and area positions to determine where in the original image pixel area is the area of interest.

A simplified flowchart of the preprocessing algorithm is presented in Figure 2. The inputs are shown on the left side of the flowchart. Inputs are grouped into **Static** and **Varying** data. Static data contains inputs which can be set when the system is set up at the beginning of the construction process. Varying data depends on the target area and the image to be converted, its time value, the actual camera position and the pixel values of the image. Subfunctions are shown as yellow circles. Each subfunction provides one main result, shown in the text box next to the function. The final result of this algorithm is a transformed image. In the flowchart, the lines show data flows. All inputs of one function have the same colour.

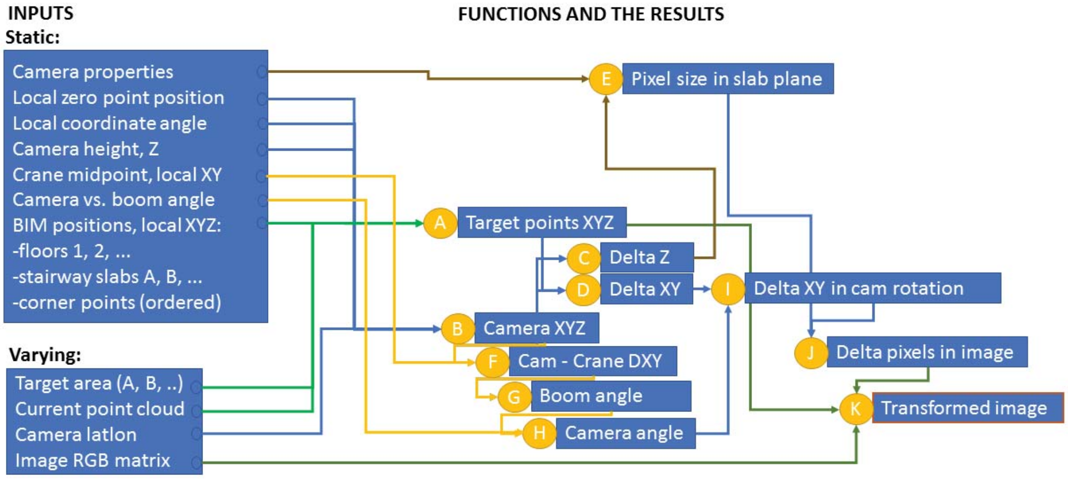


Figure 2. Flowchart diagram for automatic 2D image rotation and cropping.

The inputs to the algorithm are explained in the following. *Camera properties* define the camera's focal length for 35mm film and the number of pixels. *Local zero point position* and *local coordinate angle* are used for coordinate changes from Latitude and Longitude values to local coordinate XY values. For example, in our test case, the site coordinate zero point position and the angle between coordinate systems can be seen from Figure 1. The angle was 11.5 degrees and the local zero point was in Graus-Krueger (ETRS-GK25) coordinates, which we converted to latitude-longitude values (60.1971945, 24.9282963). *Camera vs. boom angle* is the angle between the direction of the boom and the rotation of camera. Since the camera installation was fixed, this value is constant and was estimated from the images to be 180°. Positions of the areas of interest were extracted from the design model (BIM model) and stored as a table; each floor slab has multiple corner points in the form of X and Y values. Heights, i.e. *Z* values, were also read from the BIM model and specified for each floor separately.

Varying inputs are defined by the area of interest under consideration and the particular image being converted. In addition to the actual image data matrix of colour pixel values, each image has some metainformation such as time, GPS position, camera index value, etc. With the time value, we could refer to the correct point cloud in order to obtain the current height of the slab being analysed.

We now define the functions depicted in the flowchart:

Function A uses the BIM model to determine the area of interest corner point locations in local coordinates. **Function B** converts image latitude and longitude GPS values to local (X, Y) coordinates. It first converts all latitude and longitude coordinates to UTM easting and northing coordinates (E and N), then it subtracts the local zero point position from the image coordinates and rotates the difference values in order to obtain the image position in local coordinates (XY_{Image}):

$$XY_{Image} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \times \left(EN_{Image} - EN_{zero\ position}\right) \qquad (1)$$

where α is the angle between local Y direction and global North, EN_{Image} is the image's global *(E, N)* position and $EN_{zero\ position}$ is the global position of the local zero point *($X = 0$, $Y = 0$).*

Functions C and **D** calculate the vertical difference and horizontal differences between the slab's corner points and the camera. **Function E** estimates the size of one pixel in the slab (formula 2), which is needed for scaling real world metric differences to differences expressed as pixels. It uses vertical difference from camera to slab and camera properties, such as focal length, sensor width and corresponding number of pixels. Note that we have square pixels, but lengths were calculated for both dimensions, image height and width, separately, allowing the method also to be used with special sensor types.

$$L_{px} = \frac{Z * L_{sensor}}{\boldsymbol{f} * N_{px}} \qquad (2)$$

where L_{px} is the slab surface length of one pixel, Z is the vertical difference between camera and slab, L_{sensor} is the length of the sensor in the direction of L_{px}, f is the focal length and N_{px} is the number of pixels in the sensor in the direction of L_{px}.

Final conversion from metric value differences to pixel index difference can be made when the rotation of the image is known. **Function F** calculates the horizontal difference from the crane boom's rotation point to the camera and **Function G** calculates the boom rotation angle in the XY plane, which has a value of 0 when the boom is in direction X from the crane. **Function H** gives the camera image rotation in the XY plane. **Function I** uses camera rotation information and changes horizontal differences from the local XY coordinate directions to directions of the camera sensor. Then **Function J** scales metric differences to pixel differences.

Function K combines corner point differences in pixels to the East and North directions and the rotation of the camera, and returns corner point pixel differences as image index values. Note that the camera centre point is always in the middle of the image, for the nadir orientation. Finally, **Function L** uses corner point indices to generate the final image. The scaling of the transformed image can be fixed when the original slab size is known. This size is calculated from the slab corner XY coordinate positions. The result of function L is a *rotated*, *cropped* and *rescaled* image.

4. Results and Discussion

In this section, we will present an evaluation of the accuracy of the *camera positions* as the crane rotates, which are simply the GPS values obtained from the camera every time an image is taken. Then, we will *evaluate the match* between the slab area detected by our algorithm and the ground truth, which is taken manually by selecting the slab area.

4.1 Accuracy of camera GPS values

Ideally, as the crane rotates about its axis, the camera positions should trace a circle whose centre point is in the crane midpoint and whose radius is the location of the camera on the crane boom as measured from the crane mid-point. The crane position was given, but the camera location on the boom was estimated as the median of Euclidean distances between the camera GPS values for multiple images and the crane midpoint. Some images did not contain GPS values (we could not determine the reason for this). Two images contained clearly false GPS values. However, reasonably accurate positional data was available for 2086 images taken with camera 1 and 2009 images taken with camera 2 with positions near the crane, which made it possible to estimate the GPS accuracy. Figure 3 shows camera 1 and 2 image positions in local coordinates. In addition, it shows the estimated paths of the real camera locations. Estimated circle path radius values were 13.4 m for camera 1 and 37.2 m for camera 2.

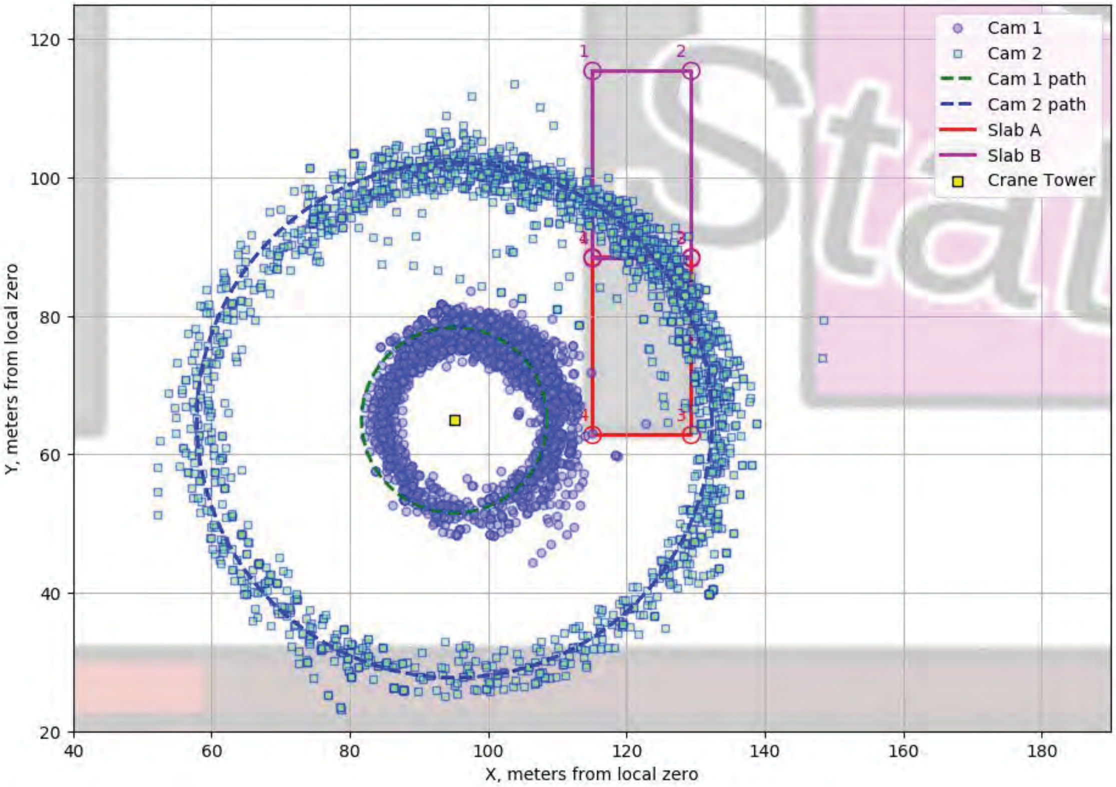
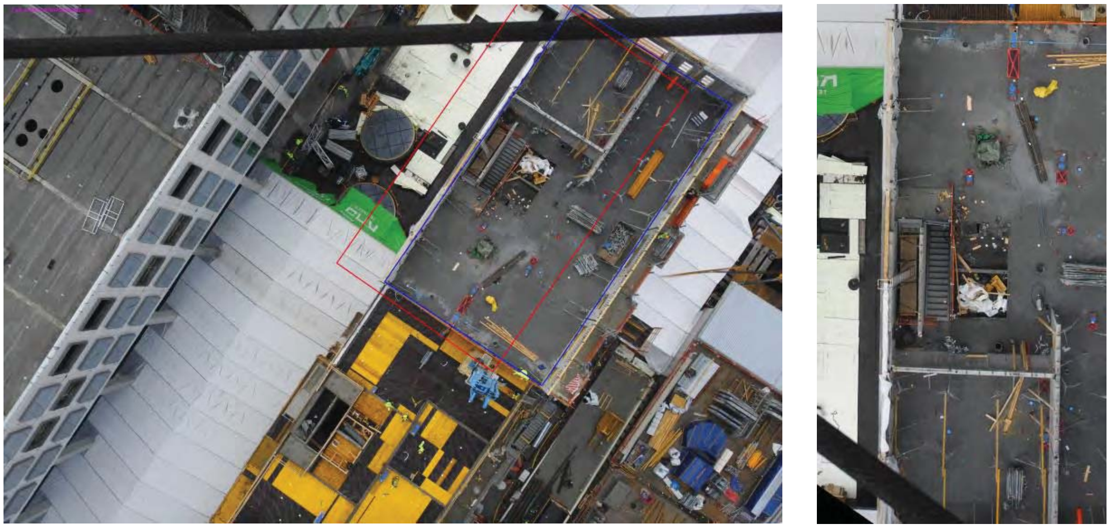


Figure 3. Crane camera image locations and cameras' path on the background map

Camera 1 image position data indicates that the actual crane midpoint might be about 2 meters away (in the X direction) from the crane coordinates available to us (which we assume to be the crane midpoint). However, the possibility of this error is not discernible in the results for camera 2. With the available crane midpoint, the standard deviations for image distance from the centre point are 2.4 and 2.9 meters for cameras 1 and 2, respectively. If we instead take the crane midpoint to be 1.8 meters further in the X direction, the deviations would be 1.9 and 2.9 meters for cameras 1 and 2, respectively. However, regardless of the crane midpoint, the camera position (GPS) values deviate significantly from the reference circle, especially considering that the width of the area of interest is only 10 meters. Furthermore, despite lying on the reference circle, the camera position values might still be offset from the actual position of the camera. This type of error can lead to the calculation of an incorrect rotation angle for the orientation-correction part of our algorithm.

4.2 Evaluation of the slab area selection



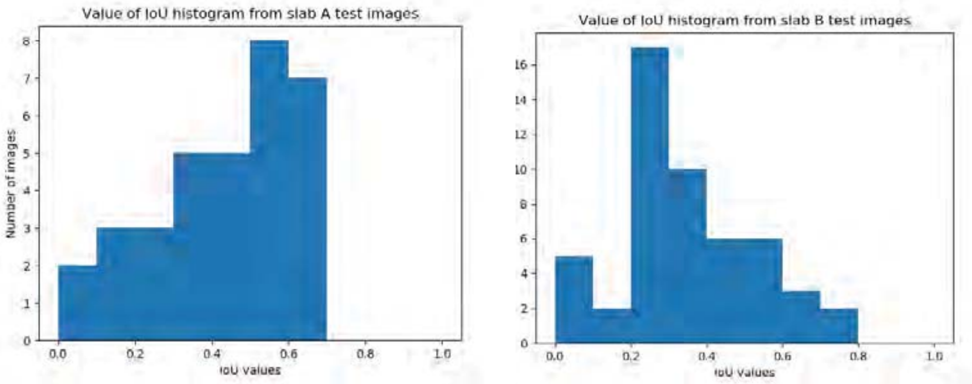
(a) (b)

Figure 4. (a) A sample image from our dataset (taken on 6th July 2018 with Camera 2). The red box shows the estimated boundaries of the slab and the blue box shows the ground truth (selected manually). (b) The detected area is selected, rotated and scaled according to local coordinates

Figure 4 shows the result of a complete implementation of the algorithm. Figure 4 (a) shows the estimated slab boundaries. To evaluate the match between the estimated and ground truth boundaries, we used the metric *Intersection over Union (IoU)* (Weidner, 2008), in which the intersection of the estimated and reference areas is divided by their union. Thus, an IoU value of 1 means perfect overlap and a value zero means no overlap. For the Figure 4 result, this value is 0.57. For the estimated area *S* and the reference area *R*:

$$IoU = \frac{\mathrm{S} \cap \mathrm{R}}{\mathrm{S} \cup \mathrm{R}} \tag{3}$$

To evaluate IoU, we manually specified 33 images from slab A and 51 from slab B. Our main selection criteria for the images was that the slab should be completely visible in the image. Histograms of IoU values plotted against number of images are shown in Figure 5.



(a) (b)

Figure 5. Results of the algorithm for slabs A and B

The histograms show that average IoU values lie around 0.4. Considering that the IoU value of

Figure 4a detection was 0.57, we can conclude that the results are not adequate as such. Thus, some additional methods are needed to improve the results.

5. Conclusions and future work

In this paper, we presented a method for removing the rotation effect of crane camera images and thus enhancing further analysis of the images. The idea was to use camera and area positions for rotation, scaling and cropping of the input images. We compared the results of the algorithm with a reference set and found that the IoU was on average 0.4. We noted that the error in image GPS positions has a significant adverse influence on the slab detection accuracy. However, despite the error, the algorithm is able to select and rotate the approximate area of interest, which can be expanded according to a pre-specified threshold to ensure that the entire slab is selected.

If we acquire the accurate crane mid-point position and camera installation points in the boom, we could accurately infer the camera's circular path as the crane rotates. With this information, we could adjust image positions to the nearest position in the path. This would decrease errors in the rotation angle calculated by our algorithm. The algorithm could also be combined with a simple edge detection technique, which could fine tune the rotation of the image and thus improve the slab detection accuracy. Additionally, we intend to explore the use of 3D voxel traversing to identify occlusions and thus remove regions in the image that are not part of the slab.

Acknowledgements

We thank YIT for providing the data for this study. The work was supported by the Reality Capture research project funded by Business Finland, Aalto University and a consortium of 5 companies.

References

Bradski G. and Kaehler, A. (2008) Learning OpenCV

Braun, A., Tuttas, S., Borrmann, A., & Stilla, U. (2015, 6). Automated Progress Monitoring Based on Photogrammetric Point Clouds and Precedence Relationship Graphs. Proceedings of the 32nd International Symposium on Automation and Robotics in Construction and Mining (ISARC 2015). International Association for Automation and Robotics in Construction (IAARC). doi:10.22260/isarc2015/0034

Eickeler, F., & Jahr, K. (n.d.). PREDICTION AND SIMULATION OF CRANE BASED CAMERA CONFIGURATIONS FOR CONSTRUCTION SITE MONITORING.

Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. (2015, 1). Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models. Journal of Computing in Civil Engineering, 29, 04014025. doi:10.1061/(asce)cp.1943-5487.0000205

Pix4D. (2017, March 27). Crane camera site surveying. Retrieved from https://www.pix4d.com/blog/crane-camera-site-surveying

Turkan, Y., Bosché, F., Haas, C. T., & Haas, R. (2014, 4). Tracking of secondary and temporary objects in structural concrete work. Construction Innovation, 14, 145-167. doi:10.1108/ci-12-2012-0063

Tuttas, S., Braun, A., Borrmann, A., & Stilla, U. (2016). Evaluation of acquisition strategies for image-based construction site monitoring. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 41, 733-740.