

PREDICTION AND SIMULATION OF CRANE BASED CAMERA CONFIGURATIONS FOR CONSTRUCTION SITE MONITORING

Felix Eickeler and Katrin Jahr

Technische Universität München · Arcisstraße 21 · 80333 München · Germany · felix.eickeler@tum.de ·
katrin.jahr@tum.de

Abstract. Undetected postponements of particular construction steps can add up over time and lead to massive delays of the whole construction process. To recognize delays early, the construction progress can be controlled using digital site monitoring, especially photogrammetric analysis. A resources-friendly, continuous approach towards photographic site monitoring lies in using crane cameras. However, badly chosen camera parameters can lead to insufficient imagery. In this paper, we predict suitable parameters for crane cameras. To this end, we perform a parametric study by simulating crane camera data of an existing project using a structure from motion pipeline.

1. Introduction

Reacting to short-term changes and hurdles on construction sites as well as ensuring optimal building progress have been subject to the experience of the workforce on site and the information at hand. Interruptions and postponements of individual construction steps or the whole construction process might not be observed properly, involving the risk of recognizing delays only after countermeasures could be taken.

Over the last decades, computer aided design and digital support tools have been increasingly adopted by the construction industry, both improving the process of construction progress planning and extending the toolkit for construction progress monitoring. Regarding construction progress monitoring, methods relying on photogrammetric analysis have proven an effective tool (Son and Kim, 2010; Xu et al., 2015; Omar and Nehdi, 2016). Using point clouds, different building parts can be recognized and compared to the work schedule. In combination with the digital model of the building, a dependency tree containing structural relevant parts can lead to the recognition of hidden building elements (Tuttas et al., 2016). In order to generate high quality point clouds, a significant number of consecutive photographs covering the monitored area is needed. Common approaches to obtain sufficient amounts of photographs include the usage of drones and crane cameras (Tuttas et al., 2016; Pix4D, 2017). An advantage of crane cameras is the permanent presence on site, allowing for automatic, steady monitoring. However, in recent works the quality of point clouds generated using crane cameras lacked details in vertical direction due to low resolution and the absence of different viewing angles. Additionally, special attention has to be paid to the viewing angle of the crane cameras. In the course of construction, the distance between the crane camera and the observed object changes. In early construction phases, the construction most likely is flat and therefore far away from the crane jib—to provide good quality images, the camera angle should be acute. In later construction phases, the construction rises and approaches the crane jib. To cover the whole area of interest, the camera angle should be obtuse. For fixed focal length cameras, a trade-off between image quality and coverage of the monitored area is necessary.

In this paper, we present an approach to simulate different crane camera parameters and the resulting point clouds by means of an existing construction project. In the first part of the paper, we introduce the workflow of our simulation and the tools used. In the second part, we take a look we present a case study and establish the relevant camera parameters. Finally, we summarize our findings.

2. Overhead Camera Simulation

Our first step in optimizing the quality of the point cloud that originated from images recorded by crane based cameras is to generate a virtual model of the process itself. Although different software tools for 3D reconstruction are available, we frame the reconstruction pipeline with a digital model of a construction site and a physical camera model that renders the site to valid input images. This upstream addition grants us the ability to feed different configurations into the simulation and run these configurations in a virtual setup. As a closed loop process, the whole chain can be used to evaluate the configuration and identify the effect of different parameters in our combined capture and reconstruction pipeline. The detailed items of the process are shown in Figure 1.

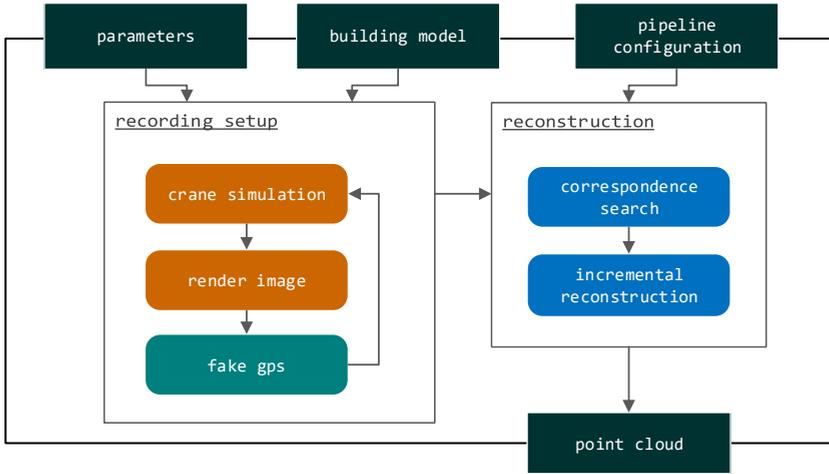


Fig. 1: Process of the simulation of the recording setup. Information is shown in dark green and cornered edges. Orange was modeled in Unity subsection 2.2, pine as a standalone program and blue corresponds to reconstruction software colmap.

2.1 Identification of Parameters

We created a digital model of our recording setup by identifying the parameters of importance and omitted all sensor specific parameters as well as component vibration and weather conditions. For this paper we used a single crane setup and assigned a parameter to every degree of freedom (see Figure 2). While the crane is sweeping over the building, images are being taken according to the angular resolution defined ($\Delta\gamma$). For the simulation, the crane sweep starts at the minimal angle (γ_{min}) and continues to the maximal angle (γ_{max}). This defines the rotation of the crane and should be sufficient to capture the whole building. For each camera defined, the position on the boom (a_{camera}), the camera orientation (α_{camera}), and the field of view (ω) given as focal length, is defined. The overall crane boom height (h_t) is specified as the vertical difference between camera and ground. In our model, the floor of the excavation is named zero. Additionally the camera parameters are described by the resolution of the resulting image, and the focal length of the lens. The imaginary detector is a standard sized 36x24 mm sensor.

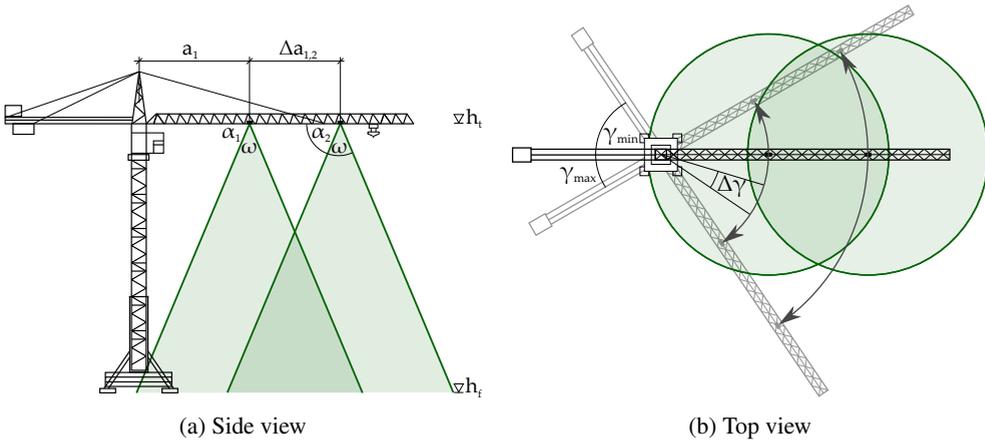
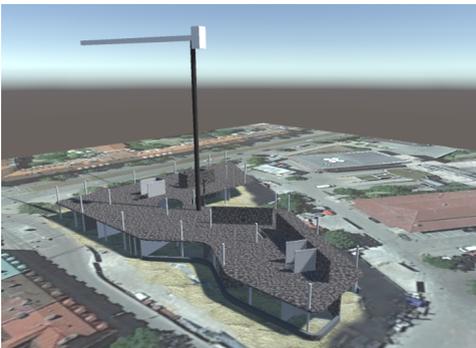


Fig. 2: Schematic representation of a tower crane with installed crane cameras depicting the examined crane and camera parameters

2.2 Crane Simulation And Image Creation

We use the all-purpose game engine Unity to simulate the described crane movement and render the images. Unity is able to import many spatial formats such as *.fbx and *.obj which can be extracted from common CAD software. In addition we choose Unity because the render pipeline can be customized easily to mimic the behaviour of a real world camera due to the high quality of built in shaders. We use a unidirectional light source and omitted the crane in the rendering to avoid the influence of shadows.

The crane and the cameras are placed according to the input parameters (see subsection 2.1). The crane sweep is chopped by the angular resolution. Every differential step is modeled as a single frame in Unity. For every cameras existent, a image is rendered. An example for the initial setup can be found in Figure 3.



(a) Experimental setup in Unity

```

<OverheadCameraConfiguration angularResolution="5" lowerBoundary="-90"
upperBoundary="150" craneHeight="48" configNumber="6">
  <Cameras>
    <Camera focallength="35">
      <Resolution horizontal="4672" vertical="3104" />
      <Spatial x="19" y="0" z="0" />
      <Orientation x="0" y="0" z="90" />
    </Camera>
    <Camera focallength="35">
      <Resolution horizontal="4672" vertical="3104" />
      <Spatial x="43" y="0" z="0" />
      <Orientation x="0" y="-30" z="90" />
    </Camera>
  </Cameras>
</OverheadCameraConfiguration>

```

(b) Input parameter as XML

Fig. 3: Images showing the initial setup of our simulation. We used a 3D representation of the TranslaTUM, an addition to a university hospital which was built in 2016.

To avoid repeating textures we assigned a large random noise, black and white texture to the concrete material. This step is essential to the simulation since the process of feature matching during the reconstruction (Section 3) would assume the same identity.

We further reduced the crosstalk during reconstruction with only pairing spatial corresponding images. Since we know the camera position, we convert the local coordinates of our simulation to global GPS coordinates. This is done by linearization of the arc length. These coordinates are provided to the reconstruction pipeline as metadata (Figure 4b) of the images.



(a) Image from a crane based camera



(b) GPS-Coordinates of an image set

Fig. 4: Resulting images from the crane simulation.

3. Reconstruction

Each image set that is generated from a crane sweep is the base of a 3D reconstruction. We use the open source **colmap**, a structure-from-motion program, to generate a dense point cloud (Schönberger and Frahm, 2016). The reconstruction process can be decomposed into two major and six minor sub-steps:

correspondence search:

1. **Feature Extraction:** To extract features in the images, a SIFT-algorithm is used. The recognized key-points are stored in a local database.
2. **Feature Matching:** Correspondences between the detected features in different images are detected. To specify which images should be compared to one-another different strategies can be selected.

incremental reconstruction:

3. **Sparse Reconstruction:** The images are registered and new points are added solving the Perspective-n-Point problem (Fischler and Bolles, 1981). Camera models can be provided to minimize the error. For each reconstruction, an initial image pair is used to reference the registered images in the point cloud.
4. **Calculate depth and normal maps:** Based on the image registration as part of the MVS(Multi-View-Stereo), depth maps are calculated (Schönberger et al., 2016).
5. **Fusing Maps:** The depth maps are fused into one dense point cloud.
6. **Meshing the point cloud:** The surface is estimated using a Poisson reconstruction. (Kazhdan and Hoppe, 2013)

We used a script to control every sub step of the point cloud generation with colmap. Different **pipeline configuration** can be modify the reconstruction parameter, such as resolution and the number of images used for the dense reconstruction. After the initial step and a SIFT-Feature matching in 5 octaves, the image was reduced to a fixed size for the whole remaining process. We used spatial matching on the provided GPS coordinates. Therefore only pictures in proximity are paired and their features matched. Since the whole process of reconstruction is computationally intensive we used the build-in CUDA functionality to speed up the cycles.

4. Parametric Study and Verification

To evaluate the described approach we ran several simulations of a known construction site. The tested parameters were part of a factorial experiment and used the same pipeline configuration. With a factorial experiment it is common to vary multiple parameters at once. The basic idea is, that instead of walking through all possible parameter configurations, one is to create a statistical analyzable set of data. The samples needed to permit any reliable forecasts are drastically reduced from volume of n to the surface of n . It is of essence that multiple parameters and their relations are varied in a non repeating pattern. In the analysis step the results are compared and the effects of the parameters are discussed.

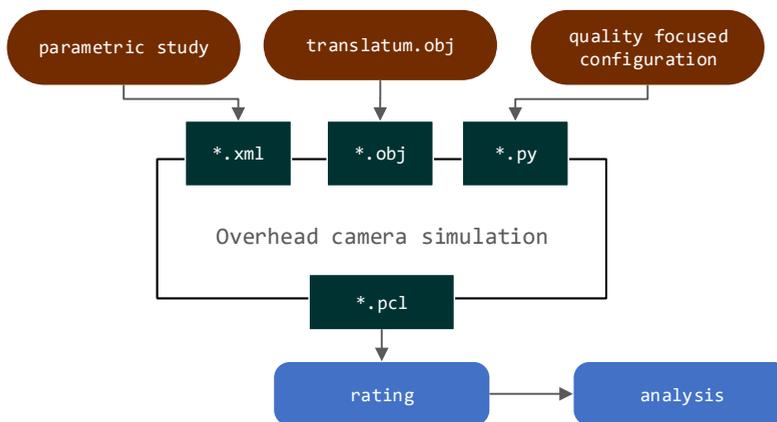


Fig. 5: The extended process to evaluate the simulation

We rated the different point clouds based on different indicators:

The main focus was to review the point cloud and evaluate the quality of the mesh. Since previous crane camera based point maps have been lacking detail in the vertical direction, this fact was intensively monitored and rated separately. Other than that, we looked at the overall quality of the matching and the representation of the columns on the roof of the 3d model to review reconstruction details. For further evaluation we took three additional metrics into account. One being the number of residuals and the second being the number of observations that were registered in the sparse reconstruction step. The third metric is the number of reconstructed points in the dense point cloud. While this alone is not a criterion for quality, it is a good indicator for the details that were captured during the reconstruction process.

Instead of taking distances and angles as two absolute parameters, we introduced the relative position of the cameras to one another. Together with the relative angle of the cameras, we can capture the

stereo view orientation. These differences are presented in the parameter table as ($\Delta\alpha_{12}$ and Δa_{12}). Each camera will be either on the left or the right side of the boom.

4.1 Factorial Sample Set

The test set contained 10 initial configurations. All these configurations comply with certain restrictions:

1. Cameras will always face each other (or be neutral). Thus we always have an overlap of matching recognizable features in the images.
2. No camera angle will be steeper than 30° .
3. The distance between two cameras has to be at least 16m.
4. The maximal and minimal rotation angles (γ_{min} , γ_{max}) will surpass the outline of the building by at least 10°
5. The angular resolution $\Delta\gamma$ is 8° or higher. While running the first tests, we realized that with a lower angular resolution than 8° no initial image pairs could be found.
6. The focal length is related to a 36x24 full frame camera and converted accordingly. We used fictional 24mm and 35mm lenses.
7. As the crane is positioned inside the excavation, the crane height from the ground is approximately 30m. Therefore the distance between the slab and the camera is always 24m.
8. At some point of the sweep both cameras need to capture parts of the 3D model

Since the number of free parameters was still very high, we used a fixed crane height ($h_t = 48$ m) and boundaries (γ_{min} , γ_{max} , $\Delta = 150^\circ$) for each combination. The concrete list of parameters is shown in Table 1. In this small test set we varied 5 of the six possible parameters with three value steps on each camera. This results in a wider spread in the differential values (see Camera 2).

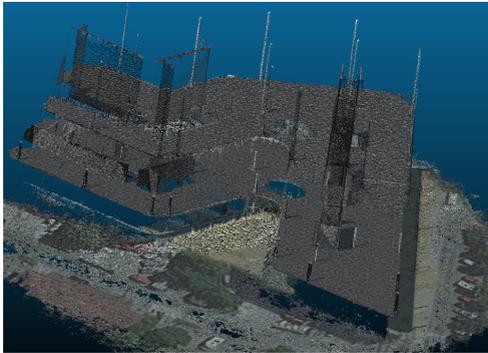
Table 1: Parameters for 10 test cases

Shared parameters			Camera 1		Camera 2	
<i>Nr.</i>	$\Delta\gamma[^\circ]$	<i>Focal length [mm]</i>	$a_1[m]$	$\alpha_1[^\circ]$	$\Delta a_{12}[m]$	$\Delta\alpha_{12}[^\circ]$
1	5	35	11	30	24	-45
2	5	35	11	15	40	-30
3	5	35	19	15	32	-15
4	5	35	19	0	24	-30
5	5	35	3	0	40	-30
6	5	35	11	15	32	-15
7	3	35	11	0	32	-45
8	3	35	19	30	24	-75
9	8	35	3	15	24	-45
10	8	35	11	0	24	-45

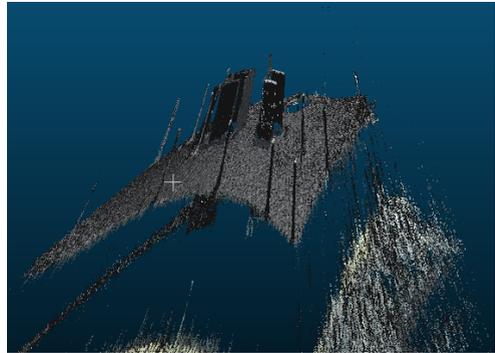
4.2 Results and Analysis

While all of our samples could technically be reconstructed, some of the point clouds were heavily distorted. The negative results are 3, 4, 6 and 8. In every faulty sample the outer camera is only capturing the outer rim of the building. A special case is number 5 since it is containing both, a damaged and a correct reconstruction (Figure 6a). The overall high quality samples, 1, 2, 9, 10, also provide superior vertical resolution. Small reconstruction errors can be found at the corners and showed parts (Figure 6c, first light grey staircase). In all resulting test samples, the viewed spot of the underlying texture is the same. The flatness of that satellite image is handled correctly in samples with lower relative angles (smaller than 0°). The reconstructed space is closely related to the angular start and stop condition of the sweep (missing details in the profile of the V).

We ran a regression analysis to identify how the metrics influencing the overall quality of the reconstruction. A significant correlation exists between the number of monitored points and the observations. While the number of points increases the quality, the relative number of observations has a negative influence.



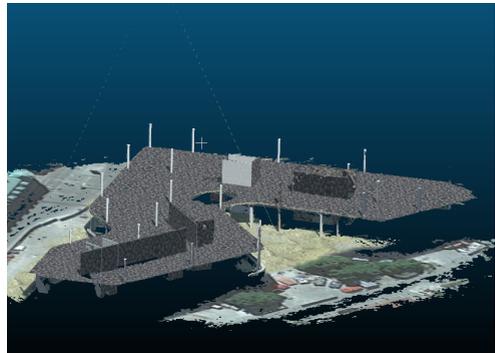
(a) Twin image, sample 5



(b) Heavy distortion, sample 4



(c) Meshed point cloud, sample 1



(d) Side view, sample 1

Fig. 6: Point clouds generated from Table 1

Increasing the radius and changing to wide angle lenses will tend to most of the disconnection issue. We also noticed that experiments with higher angular resolution will decrease the number of disconnects. This will however increase the reconstruction time. In all reconstructions shadows have a high impact on the result. This correlates to our real world experiences with images from construction

sites taken on sunny days. In the simulation we could add a diffuse light to increase the performance. Since the input images for the simulation are of high quality, the process of reconstruction can be optimized for better vertical resolution by increasing the image size during the point cloud generation. This again will lead to longer reconstruction periods and is therefore mostly suited for real world images.

5. Summary and conclusions

In this paper, we presented an approach to predict suitable parameters for crane based cameras by using a structure from motion pipeline on simulated imagery. We composed a set of test cases with varying camera parameters and evaluate them. With the findings of the regression analysis, we can predict the quality of suitable parameters without relying on the dense reconstruction. This reduces the computing time manifold and makes a larger scale parameter study viable. Additionally we can relate the orientation of the cameras and a fixed spatial reference during the rotating motion to an increase in quality.

In future studies we would like to deploy totally randomized textures to further reduce the crosstalk during the reconstruction. Another topic for future work is the evaluation on using zoom lenses, the interaction of cameras installed on different cranes and using more than two cameras per crane. Finally, we want to run tests in an experimental setup on a real construction site.

Bibliography

- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- Kazhdan, M. and Hoppe, H. (2013). Screened poisson surface reconstruction. *ACM Transactions on Graphics*, 32(3):1–13.
- Omar, T. and Nehdi, M. L. (2016). Data acquisition technologies for construction progress tracking. *Automation in Construction*, 70:143–155.
- Pix4D (2017). Crane Camera Site Surveying. <https://pix4d.com/crane-camera-site-surveying/>. Last visited on 2017/06/17.
- Schönberger, J. L. and Frahm, J.-M. (2016). Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113.
- Schönberger, J. L., Zheng, E., Frahm, J.-M., and Pollefeys, M. (2016). Pixelwise view selection for unstructured multi-view stereo. In Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III*, pages 501–518. Springer International Publishing, Cham.
- Son, H. and Kim, C. (2010). 3d structural component recognition and modeling method using color and 3d data for construction progress monitoring. *Automation in Construction*, 19(7):844–854.
- Tuttas, S., Braun, A., Borrmann, A., and Stilla, U. (2016). Evaluation of acquisition strategies for image-based construction site monitoring. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. ISPRS Congress, Prague, Czech Republic*.
- Xu, Y., He, J., Tuttas, S., and Stilla, U. (2015). Reconstruction of scaffolding components from photogrammetric point clouds of a construction site. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 401–408.