From registration to recognition of indoor construction states using on-site videos and 4D building models

Christopher Kropp¹, Christian Koch², and Markus König³

1) Ph.D. Candidate, Computing in Engineering, Ruhr-Universität Bochum, Germany. Email: christopher.kropp@rub.de
2) Dr.-Ing., Prof., Intelligent Technical Design, Bauhaus-Universität Weimar, Germany. Email: c.koch@uni-weimar.de
3) Dr.-Ing., Prof., Computing in Engineering, Ruhr-Universität Bochum, Germany. Email: koenig@inf.bi.rub.de

Abstract:
Interior finishing works cover a big budget part of every construction project. Current practice is lacking procedures and methods for frequent onsite progress monitoring, so that unexpected and unrecognized delays during this phase of construction result in a decreased overall project performance.

To address this problem, this paper synthesis previously presented parts of a method that recognizes the actual state of construction activities from as-built video data and as-designed 4D BIM data using computer vision algorithms. Under this method, two main steps are required: First, the video frames are registered with the underlying 4D building model, meaning that the camera position and orientation of each video frame is determined within the coordinate system of the building model. During this step an iterative registration procedure is proposed that combines relative and absolute pose estimation. Second, once the camera poses are known, the relevant construction activities and elements from the 4D building model are projected onto the image space to determine regions of interest, which are then taken as input for computer vision based activity state recognition methods.

Compared to previously presented conceptual work, this paper puts a strong emphasis on experiments and test results. As the overall method consists of several consecutive steps, each single process is first tested individually, before the combined procedure and the general applicability of the method is evaluated by means of two exemplary types of construction activities. All experiments show very promising results and reveal the method’s potential to support automatic indoor construction progress monitoring.

Keywords: Building Information Modelling, progress monitoring, computer vision, object detection, video registration.

1. INTRODUCTION
Building Information Modeling (BIM) is increasingly gaining attention in the digital management of construction projects. It is seen to be useful when collecting, integrating and linking several data items for different life-cycle phases, such as the planning, the design, the construction as well as the operation and maintenance. In the context of construction management, 4D building models are commonly used to visualize and analyze the construction program and to support on-site progress monitoring by linking activities of the schedule with corresponding building elements.

Progress monitoring in current practice is commonly characterized by weekly or daily paper-based progress reports that are manually collected by field personal (Roh et al. 2011), resulting in infrequent and low-quality progress updates followed by the very low potential of adequate schedule adjustments. In particular, interior construction works usually cover a significant percentage of the entire project’s budget. This means, a higher degree of automation in indoor progress monitoring would result in a reasonable amount of cost savings.

With the aim of increasing the degree of automation in this area, several research initiatives have proposed the use of sensing technologies. For example, Radio-Frequency Identification (RFID) (e.g. Razanvi & Haas 2011 ), Ultra-Wideband (UWB) (e.g. Cheng et al., 2011), laser scanners (e.g. Bosché et al. 2014), and images and videos (e.g. Golparvar-Fard et al. 2015, Hamledari et al. 2017) are used to capture the as-built state of construction work.

Each of these technologies have their inherent advantages and disadvantages. When looking at the major sensor requirements for automated BIM-based progress monitoring, four aspects need consideration:

- **Registration**: As-built sensor data needs to be mapped to and registered with the as-planned BIM model schedule activities.
- **Activity coverage**: As-built sensor data needs to allow for construction state detection, e.g. RFID cannot help determine the painting status of a concrete wall.
- **Spatial coverage**: As-built sensor data needs to handle permanent and temporary construction obstacles, which requires high location accuracies, e.g. to distinguish whether materials are already mounted or still stored next to final installation place.
• **Infrastructure independent**: As-built sensor data needs to be collected without the need for additional infrastructure, e.g. tagging, sender, receivers.

According to Kropp et al. (2018), Table 1 summarizes the suitability of different sensing technologies with regard to the previously defined aspects. In the context of indoor construction progress monitoring, vision-based methods appear to have high potential of increasing the degree of automation as this technology covers all these aspects.

### Table 1. Rating of requirement aspects for different sensing technologies

<table>
<thead>
<tr>
<th></th>
<th>AutoID/RFID</th>
<th>WiFi/UWB/GPS</th>
<th>Laser scans</th>
<th>Unregistered images</th>
<th>Registered, unordered images</th>
<th>Registered image sequences/videos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration</td>
<td>+</td>
<td>+</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Activity-coverage</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Spatial-coverage</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Infrastructure independent</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

### 2. RELATED WORK

Recent research work addressing the vision-based automation of construction progress monitoring can be categorized into data acquisition, image-to-image registration, image-to-model registration and progress recognition. This section presents an overview of existing approaches and technologies that are related to the authors’ work.

#### 2.1 Data acquisition

Several surveys have presented an overview of available sensing and automation technologies for construction progress monitoring (Yang et al. 2015), listing RFID, laser scanning as well as image and video capture systems. The limitations of RFID and laser scanning systems are presented above (see Tab. 1).

Image and video capture systems as the basis for progress monitoring are characterized by low hardware costs, short acquisition time, the ease of collection, and portability. Almost no special knowledge is needed when taking photos of a construction site. Common camera systems may consist of either a monocular or a stereo sensor. The perspective can be fixed and known, or uncertain. Next to taking single photos, videos have been recently used to capture as-built scenes (e.g. Dai et al. 2013). Occlusions may occur when sensing the site using cameras; however, these can be handled by appropriately traversing the indoor scene.

#### 2.2 Image-to-image registration (relative registration)

There are mainly three approaches when trying to estimate the (relative) camera motion based on subsequent images within a scene: Visual Odometry (VO), Structure from Motion (SFM) and Visual Simultaneous Localization and Mapping (VSLAM).

VO has been first presented by Nister et al. (2004), who describe an algorithm for estimating the trajectory of a camera that traverses through a scene by extracting and tracking visual features. SFM reconstructs the camera motion and the structure of the scene based on advanced feature descriptors and robust matching of unordered image sets (Golpavar-Fard et al. 2015). Bundle adjustment is performed to minimize the re-projection error. VSLAM considers both the estimation of the camera position and orientation, and the structure of the scene (map). This is why it is sometimes referred to as real-time or online SFM.

#### 2.3 Image-to-model registration (absolute registration)

In contrast to relative registration (image-to-image), absolute registration (image-to-model) is concerned with the registration of as-built image data and the corresponding as-design model. The main idea is to identify correspondences of different geometric primitives (points, lines, planes) in order to determine translation, rotation and scale of captured images in reference to the 3D design model.

The first category of approaches assumes cameras to be fixed and their position and orientation as known on site. Lukins et al. (2007) have used a fixed camera to classify changes in the scene appearance as structural events related to the building model using pose estimation methods. They report about failed alignments due to building model complexity and cluttered scenes. The approach introduced by Ibrahim et al. (2009) obtains prior knowledge of building components and their occupancy within a scene from a 4D building model registered to the camera. However, the authors state that fixed cameras on a construction site can hardly be used for tracking progress due to inflexibility in response to changing structures.
The second category is able to deal with unordered images taken from different unknown positions and orientations of the camera. These images are either manually registered to a 4D BIM model or they are used to firstly reconstruct a point cloud (see SFM) that is then automatically registered with the 3D model using the iterative closest point (ICP) algorithm. A similar approach is presented by Kim et al. (2013) that deals with the fully-automated registration using re-sampling of the point cloud and the 3D building model to provide a common data layout and avoid initial guess initialization of the ICP algorithm.

The third category uses direct correspondences between model features and their projections to the images, also called the Perspective-n-Point/Line (PnP/PnL) problem. A set of three correspondences is the smallest subset of correspondences used in solving the problem. Chen has presented a general solution for the PnL problem (Chen 1991). The approach of Přibyl et al. (2015) is based on lines only and is declared as very robust to outliers, but requires at least nine line correspondences, whereas Xu et al. (2016) considers points and lines. However, in case the correspondences contain many outliers, the accuracy is negatively affected (Přibyl et al. 2015).

2.4 Progress recognition

Several approaches have focused on vision-based recognition of construction components and activities. Roh et al. (2011) has presented an object detection method especially designed for indoor purposes of progress monitoring. Kropp et al. (2013) have presented an approach to recognize structural finishing components, exemplified by the recognition of heating devices. A cascaded material state determination approach has been presented in Kropp et al. (2014), which has been extended by Hamledari et al. (2017) to detect electrical outlets.

Ibrahim et al. (2009) have presented approaches that support work package progress inspection by available 3D models and a fixed camera. Those authors agree that this approach faces challenges by heavy clutter, uncontrolled and unforeseeable lighting variations, dynamic target appearance and frequent occlusions.

Chen et al. (2011) have used UWB to fuse the camera images with three dimensional space, while depth images are considered in the approaches of Khosrowpoura et al. (2014). These approaches involve a lot of time applying the camera system on one single task. Workers need to be observed permanently by the camera to guess the current activity, without inferring the activity state.

2.5 Problem statement and objectives

As of today, there are no methods available that sufficiently assist to continuously automate construction progress monitoring. Existing methods mostly deal with outdoor construction sites. Furthermore, these methods are neither directly applicable indoors nor validated for interior sites. Moreover, they are predominantly concerned with accurate geometry reconstruction and subsequent as-design comparison, but lack the capability of activity state recognition and project delay prediction.

To overcome the mentioned limitations, the authors address the following main research objectives (Kropp et al. 2018):

- Leverage of 4D BIM information to enable smart methods for video based progress monitoring;
- Robust registration of acquired as-built data with the underlying 4D BIM model;
- A recognition approach that is able to consider a vast set of construction activities contained in 4D BIM models.

3. METHODOLOGY

The proposed overall concept for automated progress monitoring is presented in Fig1, and regards all consecutive processes starting from data acquisition towards activity state recognition. The as-built construction site data is captured with video cameras, the as-built video frames are registered with the 4D BIM model, and the status of construction activities is determined by analyzing the content of video frames sequences. The identified activity states can then be used as an information resource for assisted decision making, which may result in re-scheduling.

3.1 Registration

Registration is the procedure to determine the pose (position and orientation) of the camera that captured a video frame relative to the coordinate system of the as-design building model. Figure 2 depicts the proposed method. The initial registration is performed manually (only once for the first video frame) by navigating (moving, rotating, scaling) through the digital model in order to match the camera view with the model view.
Once the initial pose is estimated, the automatic fine pose estimation employs an image-to-model registration algorithm to determine a more accurate pose. Subsequent frames are directly fed into the fine pose estimation and, in case of a successful estimate, further forwarded to the activity state recognition step. If the fine pose estimation fails, e.g. in case of less visual features, rough motion estimation runs an image-to-image registration algorithm.

**Fine pose estimation (absolute or image-to-model registration)**

Based on the expected construction state and a rough estimate of the camera pose from the previous video frame, the geometry of the 4D building model is analyzed and line segments are extracted. These are then projected onto the image space and define a region of interest (ROI) that is scanned when extracting image lines. The problem of matching model lines and image lines is solved by incorporating a modified RANSAC algorithm that uses model lines intersections as additional point features to improve accuracy (Xu et al. 2016). Taking the line and point correspondence subset as the input for a PnX algorithm, the model parameters of the camera pose can be determined. Detailed information on the line extraction and line matching algorithm can be found in Kropp et al. (2018).

**Rough motion estimation (relative or image-to-image registration)**

In contrast to the fine pose estimation, which derives camera poses absolutely to the building model geometry, rough motion estimation predicts poses based on relative image-to-image registration.

Two distinct frames $i$ and $j$ are considered to estimate transformation and scale. It is assumed that the absolute
pose \( P_a = R_a|c_a \), composed of the absolute rotation \( R_a \) and the absolute camera center \( c_a \), from previous fine pose estimation as well as its relative counterpart \( P_r = R_r|c_r \), composed of the relative rotation \( R_r \) and the relative camera center \( c_r \), from image-to-image registration are known for both frames. Consequently, the ratio of the Euclidean distances between the absolute \( c_a \) and the relative \( c_r \) camera centers yields the scale factor from the relative to the absolute coordinate system

\[
s_{ij} = \frac{|c_{ai} - c_{aj}|}{|c_{ri} - c_{rj}|}.
\]

For a frame \( k \) for which an absolute pose is not available and only \( P_r \) is known, the relative motion from the pose of frame \( j \) is determined as

\[
P_{r\Delta} = R_{r\Delta}|c_{r\Delta} = P_{rj}^{-1} \cdot P_rk.
\]

The relative motion is then scaled to the absolute motion

\[
P_{a\Delta} = R_{a\Delta}|s_{ij} \cdot c_{r\Delta}.
\]

Knowing the absolute motion from frame \( j \) to frame \( k \), the absolute rough pose is defined as

\[
P_{a\text{rough}_k} = P_{a\Delta} \cdot P_{aj}.
\]

### 3.2 Recognition

The recognition step of the framework contains two sub-steps prior to the actual activity state recognition step. First, relevant construction objects (e.g. walls, heating devices) and their geometry in the image space are determined by search space reduction. Afterwards, relevant image areas are rectified in order to reduce the recognition problem to the two-dimensional space.

Without considering design knowledge, any construction object could possibly appear anytime in every image part with any rotation, scale and translation. However, knowing the camera pose within the digital building model and the anticipated construction activities, relevant building objects are projected onto the captured video frames resulting in regions of interest (ROI). These regions are further processed in the next step.

In order to make the appearances of objects from different perspectives comparable, the ROIs from the previous step are “normalized” by rectification. The rectification process projects these ROIs onto a common image plane thus approximating the original image to a unified view that ideally would be identical for all input images.

State recognition is the actual process of recognizing objects or materials within the ROIs identified within the previous steps. Here, a vast amount of existing algorithms can be applied for this purpose. Most commonly, objects and materials are detected using machine learning methods and classifiers. An overview of existing methods is presented in [28].

### 4. EXPERIMENTS AND RESULTS

The goal of the experiments is to evaluate not only the individual steps of the presented frame, but also the entire framework as one unit. On-site videos have been captured with a tablet PC at a resolution of 1920x1080 pixels. More detailed experiments and results can be found in Kropp et al. (2018).

#### 4.1 Registration

Exemplary for the registration step of the framework, this section is dedicated to testing the fine pose estimation (image-to-model registration) method. Camera pose deviations are simulated on a set of defined ground truth camera poses. The simulated deviations include rotation and translation of the camera. Figure 3 depicts the applied maximum deviations, namely -20°/+20° rotation (Fig. 3a-c, rotation of x-, y-, z-axis) and -1m/+1m translation (Fig. 3d-f, translation of x-, y-, z-axis).

Using a test set of 510 images, the goal of this experiment is to recover the actual camera pose based on the divergent input camera pose. Figures 4a and 4b show the results of this experiment plotting the mean point distance of the endpoints of the model lines between the simulated pose and the recovered pose. For translation, a mean error of approx. 0.1m has been measured, while for rotation a mean error of approx. 0.4 m has been achieved.
4.2 Overall workflow

To evaluate the entire workflow, a video consisting of 808 frames (approx. 32 seconds) have been recorded at the renovation of the IC building on the Ruhr-Universität Bochum campus using a monocular rear camera of a tablet PC. This video captures a non-present radiator and installed drywall panels. The used methods for radiator recognition and state classification for dry wall installation are described in more detail in Kropp et al. (2018). Figures 5, 6 and 7 depict screenshots of the developed software prototype. The first row, from left to right, presents the registration result, namely, the camera view with highlighted ROIs for walls and radiator, the design model view, and the camera pose in the design model. The second row depicts the recognition results for the left wall, the radiator and the right wall depending on the size of the ROI in the video frame (visibility).
5. CONCLUSIONS

This paper presented a framework for vision-based indoor construction progress monitoring starting from video registration towards activity state recognition. This framework is based on information in the as-designed 4D BIM model and video frames that are merged and processed to derive monitoring information about construction tasks. Individual frames are first registered with the 4D BIM model. Then relevant as-designed construction objects are
mapped onto respective video frames to create regions of interest, which are subsequently fed into activity state recognition methods. The experiments presented in the paper show that the automation of video-based progress monitoring is possible. However, due to the complexity of the framework, there are many points for optimization and further improvement, for example, towards real-time registration and recognition, and the use of depth information in modern visual sensors.

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge the financial support by the German Research Foundation (DFG) for this work under the grants KO 4311/4-1 and KO 3473/8-1.

**REFERENCES**


