Towards Automated Visual Assessment
of Progress in Construction Projects

Timothy C. Lukins *
School of the Built Environment
Heriot-Watt University
Edinburgh, EH14 4AS, UK

Emanuele Trucco †
School of Computing
University of Dundee
Dundee, DD1 4HN, UK

Abstract

Current assessment of progress in construction projects is a manual task that is often infrequent and error prone. Images of sites are extremely cluttered and rife with shadows, occlusions, equipment, and people - making them extremely hard to analyse. We present a first prototype system capable of detecting changes on a building site observed by a fixed camera, and classifying such changes as either actual structural events, or as unrelated. We exploit a prior building model to align camera and scene, thus identifying image regions where building components are expected to appear. This then enables us to home in on significant change events and verify the actual presence of a particular type of component. We place our approach within an emerging paradigm for integration in the construction industry, and highlight the benefits of automated image based feedback.

1 Introduction

1.1 Context and Motivation

We propose a system capable of automatically detecting changes on a building site observed by a fixed camera, and to identify such changes as parts of the construction plan. Within the construction industry, the pressure is to always seek to deliver on time and on budget. Regular progress monitoring on construction sites is a fundamental part of project management required to achieve this. It allows payments to contractors to be paid periodically on the basis of the amount of work completed, and to respond to the many unexpected events affecting work (e.g., weather, supplies, accidents). Monitoring is currently carried out by independent surveyors, who inspect a site to check advancements against project schedule. However, these traditional approaches are often infrequent, subjective and manually intensive,

Recent trends in construction aim to represent every aspect of a whole project in a single, integrated Building Information Model (BIM). A BIM can include various aspects of a project, giving rise to "*nD*-systems" [1], containing not only the structural,

---

*t.lukins@hw.ac.uk
†manueltrucco@computing.dundee.ac.uk

geometric and material properties captured by traditional architectural designs but also schedule, health and safety, and cost information [8]. The BIM can furthermore be based on assemblages of standardised *components*, organised in turn into a hierarchy of *work-packages* [11]. The latter identify manageable and distinct project stages, and are defined by project-specific criteria, e.g., contractor, site location, or structure. Such a model, coupled with a suitable, interactive GUI, provides managers with complete visibility over a project in time. Importantly, the expected state at any time can be visualised, that is: what should have been built by a given date.

This ideally would be directly related to progress by keeping the model aligned with reality, i.e., the actual status of work on site. Unfortunately this proves very difficult. Currently, information collected from site surveys provide rather general assessments. Automated monitoring techniques have been proposed for tracking components, including bar-code labelling, RF tagging and embedded sensors, but their costs and practical aspects limit applicability [13]. Photographs of a site are taken as a matter of course but mainly just for recording purposes. Memon *et al.* [12] is one of the very few construction studies investigating imaging for progress assessment. The authors use a commercial package to recover a 3D model of the structures on site; these are then compared to the CAD model of the planned bulding. The process is however entirely manual and only works for simple structures, and is not so scalable to the complexities found of real construction sites.

1.2 Objectives and Contributions

We are partnering with the University of Salford to investigate the potential of computer vision as a progress assessment tool within a BIM framework. In general, this involves assessing automatically which elements of a building have been completed, by when, and consequently what stage of progress has been reached. Our main contribution is indeed an initial system detecting changes in image sequences acquired by a single, fixed camera placed in an uncontrolled location, and identifying changes as planned, i.e., expected events in the construction plan, or not. This system differs from other reported vision research by focusing on detecting close-range structural change, by bringing together a number of different techniques as follows.

Firstly, the system estimates the unknown position of the camera with respect to the building by line-based 3-D model matching similar to [5]. However, implementations of algorithms such as Simplex and SoftPOSIT failed to align given the complexity of the building model and the high clutter of real-site images. We achieved reasonably good results with a powerful evolutionary optimizer, *Particle Swarm Optimization* (PSO) [7].

Secondly, the aligned images in the sequence are normalized photometrically to assuage the effects of uninteresting lighting changes. This is followed by identification of regions in which changes have taken place since previous photographs. We use *Simoncelli filters* [16] to get good estimates of time derivatives. Currently we focus and retain only those change regions in which changes are actually scheduled within the BIM around the time the image was taken.

Thirdly, image changes do not *per se* imply that the expected component has been put in place. For this reason changes must be classified as significant (part of the building plan) or not. In other words, the component expected must be recognized in the target region; if not, the change is deemed uninteresting. Recognition is performed by an *AdaBoost* implementation with simple, Haar-like filters as per [18].

1.3   Related Work

The advantages of an integrated IT framework for the construction industry have been laid out by Bjork [3]. Trucco and Kaka [17] have then first discussed the potential of computer vision for progress assessment within such a framework. Their work presented a prototype vision system dealing with localisation of specific structures on a construction site using an iconic image matching approach. To our best knowledge, no papers exist in the computer vision literature on the automatic assessment of progress on building site as part of an integrated construction model. However, work exists on various related topics between the intersection of CV and buildings in general.

One topic that forms the focus of related research is in reconstructing building models from images. Traditionally, photogrammetry is the tool often employed by construction to measure a building from images. Many of these techniques underpin the core approaches in CV which focus on the multiple view recovery of structure. Building on previous developments in [2], Dick *et al.* [6] present a powerful framework in which a dictionary of common building elements and their learned distribution is fitted to observed geometry. Exploiting structural knowledge constrains the combinatorics of the problem. Cantzler and Fisher [4] presented an algorithm exploiting geometric constraints to improve the quality of automatically reconstructed building models. Schindler *et al.* [15] adopt similar techniques in the recent trend for "urban modelling" of entire cities from video as a structure from motion problem. They use the vanishing points to categorise and match the linear features often found on buildings. Notice that all this work is concerned with reconstruction, not with measuring building progress, and that the final models do not generally have the level of detail to accurately measure individual elements of the structure.

A second topic is quality assurance, e.g., tolerance verification, a typical task of photogrammetry [19]. This is a research area that is also founded on interpreting 3D data. For example, Gordon *et al.* [9] present a complete system to match scans of buildings to original 3D plans, in order to find and highlight elements deviating from prescribed tolerances. This approach to detecting defects can visually show and compare the "as-built" state to the original model. Scanning technology implies that the results can be very accurate, but there are issues related to the sheer volume of data to be interpreted.

The work above concerns change detection in close-range imagery, within 100 metres of the sensor. A great deal of other research exists on observation and verification of buildings using remote, long-distance imagery, including LIDAR and visible-spectrum images. For example, Huertas and Nevatia [10] report a system detecting specific changes in the overall structure of buildings. They exploit linear structure and seek to discard false edges created by shadows. Generalised 3D models are fitted to the result; and discrepancies used to justify identify changes.

Finally, we observe that progress assessment hinges on *change*, defined either as departure from an *a-priori* model [9, 10], or as difference between images taken at different times. An important realization is that the more prior knowledge of the nature of the change is modelled, the more the task becomes akin to model-based detection and location. Finding significant changes between images relates to the vast research area of *change detection*, especially for surveillance applications. For reasons of space, here we point the reader only to the excellent survey by Radke *et al.* [14].

In summary, our work fills a niche in that several vision algorithms have been reported for automatic model building, model improving and quality control, but nothing seems to exist on the automatic assessment of progress on building sites.

2   System Overview

The key problem is that construction sites offer some of the worst instances of a cluttered scene. The complexities of the structure, especially while it is being built, will result in certain ambiguities and self-occlusion. Furthermore, changes to the true structure over time, which indicate the progress we are interested in, may be surrounded by a large number of additional spurious events. These can broadly be classified as uninteresting changes, and include effects that are *environmental* (light and shadows, rain, snow, etc), *partial* (longer cumulative periods of work, such as pouring concrete), and *occasional* (people and equipment moving or staying still for a period of time).

The task of spotting correct changes can be split down into a number of modules:

- Localisation - to align a model of the building components to a sequence of images.
- Detection - to spot changes occurring over time within the region of components.
- Verification - to confirm after a change if a component is indeed present.

To make the task easier, we can engineer the camera set-up and leverage the knowledge we have from the construction model. For example, we assume that input is based on a fixed, single camera so that the initial camera-model localisation need only be performed once. Given the complete and accurate construction model, which includes CAD descriptions of the building elements, we can partition the image into regions where components should occur. Observing over time the intensity variations within these regions allows us to detect potential changes for verification.

3   Processing Modules

3.1   Camera to Model Alignment

This is effectively a model-based fitting approach in which we use a standard Canny edge detector with further post processing to join up strong neighbouring collinear lines in the image (within 0.1 radians similar direction and 30 pixels separation). The lines are represented as segments with end-point and angle (x, y, ρ) in a $\Re^3$ space. We represent the camera pose by 6 parameters, and the model is then rendered from this viewpoint and the same edge detection algorithms run to create a further set of model segments.

Non-linear optimisation of the camera pose then proceeds using a Particle Swarm based approach [7] to minimise the Sum of Squared Distances between the two sets of segments. PSO is a stochastic technique that iteratively searches across the multidimensional problem domain using a "swarm of particles" that are each guided by their own best solution, and by knowledge of the current global best for the entire swarm. Each particles velocity is thus governed by a weighing between these locations (which are updated on discovering new optima) and by inertia and randomising terms. This allows the fitting to initially proceed from a large number of possible solutions before rapidly converging towards the overall best global solution. This approach is well suited to a complex search space created by unrelated structures.

Additional constraints in the camera location (between 5 and 100 metres away, and not below the ground plane) also restricts the search. The output of this process is to produce a set of component regions, or *template masks*, created from the projection of the model onto the aligned image.

3.2 Component Change Detection

The biggest factor influencing false detection of changes in the image are the sudden effects of lighting. To remove as much as possible these variations, the image is first converted to grayscale, and the pixels within $n \times n$ sub-blocks are normalised to have zero mean and variance of one - as shown in Figure 1. Since we are considering such relatively small areas (in terms of the overall image) this approach is justified for removing a considerable amount of localised brightness and shadowing.
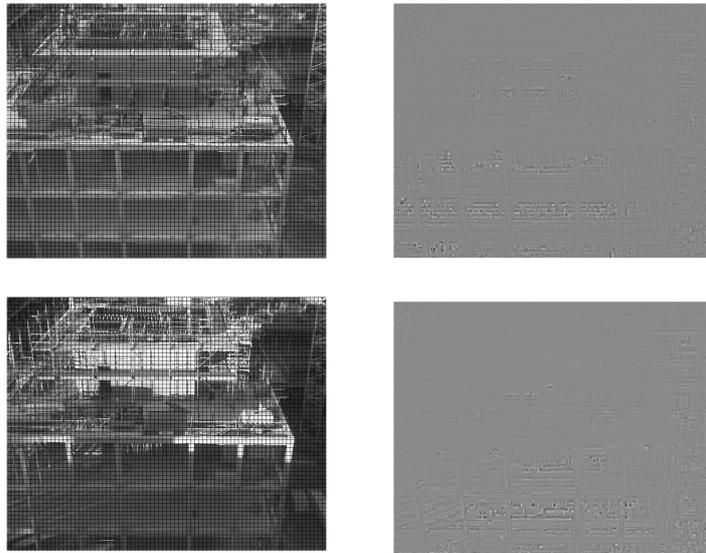
Figure 1: Gray-scale image *(top left)*, and same scene 20 minutes later with changed lighting conditions *(bottom left)*. Normalisation sub-blocks are then used to produce the respective corrected images shown retaining only low-frequency structure *(right)*.

We then wish to observe and mark sudden change occurring within particular regions with respect to time. For this, robust derivative estimation is achieved using a combination of $5 \times 5 \times 5$ two-stage low-pass (noise reduction) and high-pass (differentiate) 5-tap filters as developed by [16]. This is performed as a convolution first over the regions spatially using the smoothing kernel $[0.036, 0.249, 0.431, 0.249, 0.036]$ followed by the differentiation kernel $[-0.108, -0.283, 0.0, 0.283, 0.108]$ across time. We then calculate the mean for all the pixel temporal derivatives in the components template mask. A sensitivity threshold can then be set (as demonstrated experimentally in Section 4) to nominate peaks in the sequence as significant changes for further verification.

3.3 Verification of Components

We must now verify that detected changes are due to the appearance of scheduled components. We use a classifier constructed by the Adaboost algorithm based on a collection of "weak" classifiers [18]. We train the classifiers on a collection of 100 image samples

of the component (taken from other source images) which are further modified by varying degrees of rotation and intensity to mimic an increased training set of 1000 images. A further 2000 negative random samples from images of construction sites in which the component does not appear are also used. We employ the *Gentle* Adaboost algorithm based on a core set of non-rotated, linear Haar features to learn from this data.

The output of the classifier cascade when applied to an image is the location and scale of strong responses above a certain size as shown in Figure 2. This also highlights some of the Haar features learnt by the algorithm, showing the strong prevalence for "column like" edges and corners. Notice that in the output there are a number of false negatives, and also false positives in the incorrect matching to the smaller scaffolding poles at the top. However, what it important for the system is the presence of a column around the same time as change is detected. These locations can be directly compared to the distance from the centres of the component template mask in question. If verified, then the component is said to be present for the time of the originally detected peak.

Figure 2: Example output for column detection *(left)*, with Haar features used *(right)*.

4 Experimental Results

To illustrate this approach, we now describe a working implementation based on input from a real construction site. We use data from the construction of the new School of Informatics building at the University of Edinburgh. From this we consider a sequence of 1280×1024 images taken by a fixed camera every 20 minutes from 09:00 to 15:40 over 4 months (November '06 to February '07). We delete images taken over the weekend for a total of 1806 frames. As shown in Figure 3 this represents the construction of effectively three work packages of columns over increasing levels in the structure. Also shown is the 3D model which we we align on the basis of the ground floor components already in view. The resulting component template mask for 17 columns are further indicated in the figure. Notice we only consider those columns that are entirely in view, although it would be possible to also assess others which are not completely visible - especially when considering data from other cameras.
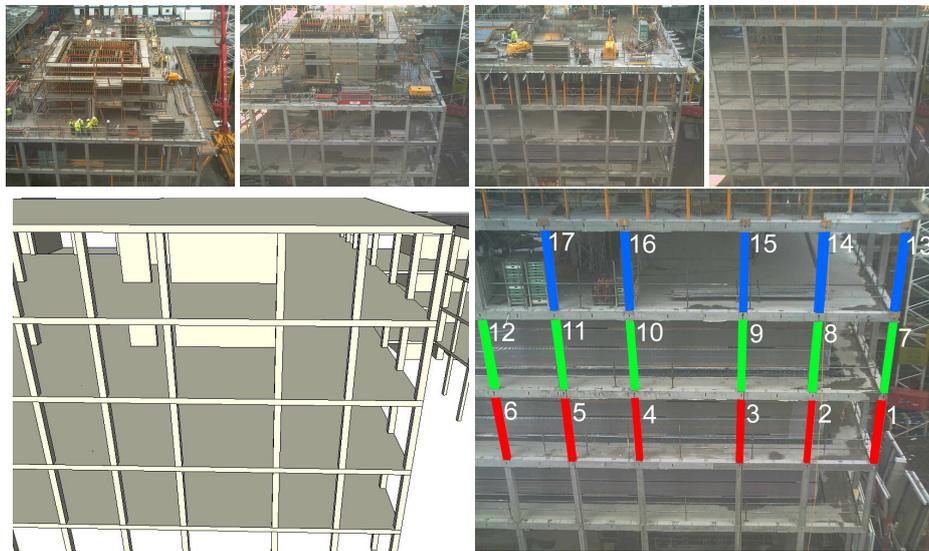
Figure 3: Input data sequence showing typical frames by month *(top)*, with aligned 3D model and resulting template masks for individual columns by work package *(bottom)*.

The alignment to perform this requires a degree of fine tuning to recover a good fit to the intrinsic parameters of the camera, particularly the focal length. This could be performed in advance by a calibration procedure. Optimisation of the pose proceeds based on the strong vertical and horizontal structure in the lower portion of the initial image. This is performed using 100 particles, taking 1000 iterations to converge to a reasonable solution. Following this, the template masks are generated from the projected outline of the final column positions. Each frame for a six-day period around expected construction time is loaded, converted to grayscale, and corrected for localised illumination using 64×64 blocks. The pixel values are then convolved with the smoothing and derivative filters. Taking the mean of the positive temporal derivative for each component template region reveals the plots as shown in Figure 4 for each period of construction, comparing the effects of illumination correction to the same process performed without.

The actual complexity and effectiveness of the graphs over these periods depends on the amount of overall work visible to the camera, and environmental factors during that time (i.e. periods of clear weather). For example, there is more activity and a large proportion of the rest of the site is still in view during the building of floor two. Notice that these can still cause numerous false negatives. However, illumination correction is able to remove many of these, especially the peaks caused by variation in the morning and evening (when the sun hits the site) as shown by the graphs on the right.

The performance of the final verification phase of the system is shown in Figure 5. ROC curves are established by comparing the correct timings of the columns against verification by Adaboost for detection thresholds ranging from 0.01 to 1.3 using both corrected and raw grayscale images. The best result is gained for corrected images with threshold 0.05, in which a total 16 of the 17 columns are correctly detected when they are actually installed.
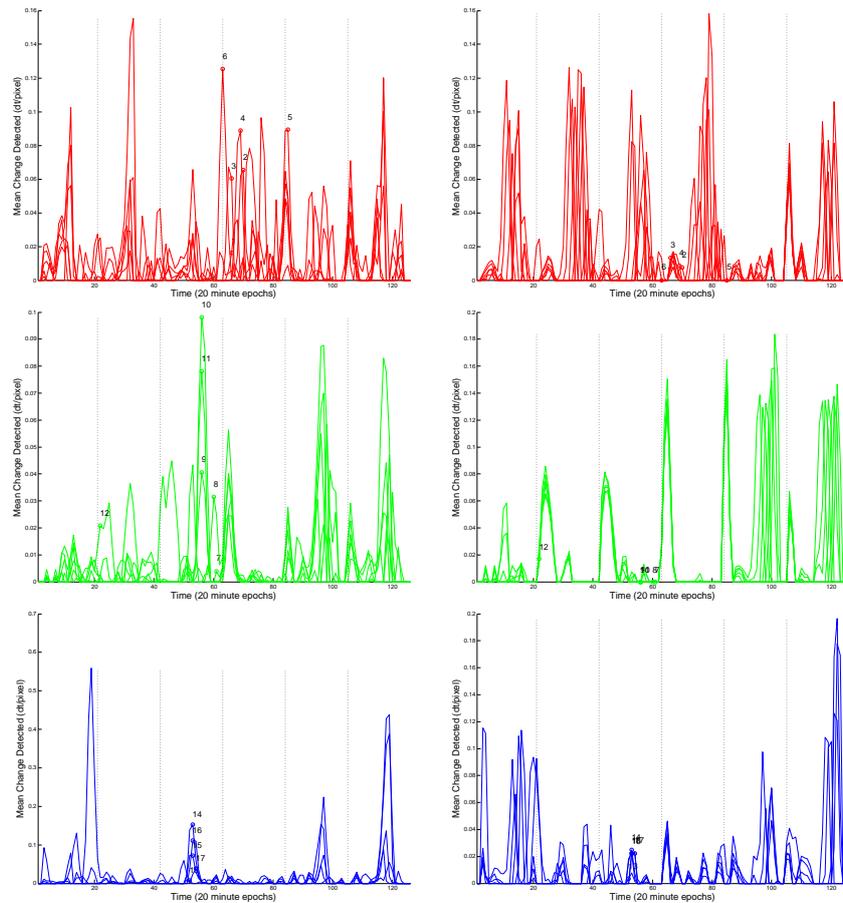
Figure 4: Change detection for 6 day period with illumination correction *(left)* and without *(right)*. Over the second floor *(top)*, third floor *(middle)*, and fourth floor *(bottom)* .

5   Conclusions

In this paper we have presented the application of techniques from computer vision to the task of reporting construction progress. We have shown how this is feasible by a combination of geometric model matching followed by statistical analysis of template mask regions to identify interesting changes for verification by an image based classifier. We have demonstrated how this approach can produce reasonable robust and reliable detection of key events during the construction process. The prospect of automated assessment represents huge potential for how large scale construction projects are managed.

However, a number of improvements are certainly possible. Firstly, in refining the model based fitting for image alignment, and to include input from other camera angles. This would enable us to handle occlusions in a more rigourous way (currently we ignore what we cannot see). Secondly, in making the detection algorithm more reliable by including texture and colour as indicators of change. In addition, more complex models of lighting could be used to remove illumination changes. Thirdly, in improving the verification classifier, especially for other types of component and differing materials.
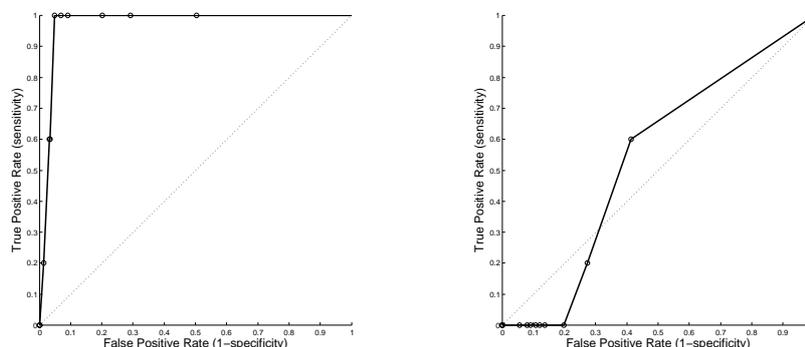
Figure 5: ROC curves for overall detection of 17 columns applying illumination correction *(left)* and without *(right)* with varying detection thresholds.

We are currently integrating this approach within an actual BIM framework in order to measure the true benefits for construction progress monitoring of an entire site. Even the simplest automatic confirmation of completion would enable further,more complex assessment (e.g. second floor activities require that the first floor be complete). In addition, individual work can be related to entire work-package progress, which will allow more meaningful and useful reports to be generated.

Acknowledgements

References

[1] G. Aouad, A. Lee, and S. Wu. nD modelling in construction: an integrated approach. *Journal of architectural engineering and design management*, 1(1), 2004.

[2] C. Baillard and A. Zisserman. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. In *19th ISPRS Congress and Exhibition*, Amsterdam, Netherlands, 2000.

[3] B.C. Bjork. Information technology in construction: domain definition and research issues. *International Journal of Computer Integrated Design and Construction*, 1(1):3–16, 1999.

[4] H. Cantzler, R. B. Fisher, and M. Devy. Improving architectural 3D reconstruction by plane and edge constraining. In *British Machine Vision Association Conference*, pages 43–52, Cardiff, UK, 2002.

[5] P. David and D. DeMenthon. Object recognition in high clutter images using line features. In *Internationa Conference in Computer Vision*, Beijing, China, 2005.

[6] A. R. Dick, P. H. S. Torr, and R. Cipolla. Modelling and interpretation of architecture from several images. *International Journal of Computer Vision*, 60(2):111–134, 2004.

[7] R. C. Eberhart and Y. H. Shi. Special issue on particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, 8(3), 2004.

[8] C. Fu, G. Aouad, A. Lee, A. Marshall-Ponting, and S. Wu. IFC model viewer to support nD model application. *Automation in Construction*, 15(2):178–185, 2006.

[9] C. Gordon, F. Boukamp, D. Huber, E. Latimer, K. Park, and B. Akinci. Combining reality capture technologies for construction defect detection: A case study. In *9th EuropIA International Conference (EIA9)*, pages 99–108, Istanbul, Turkey, 2003.

[10] A. Huertas and R. Nevatia. Detecting changes in aerial views of man-made structures. *Image and Vision Computing*, 18(8):583–596, 2000.

[11] Y. Jung and S. Woo. Flexible work breakdown structure for integrated cost and schedule control. *Journal of Construction Engineering and Management*, 130(5):616–25, 2004.

[12] Z. A. Memon, M. Z. A. Majid, and M. Mustaffar. An automatic project progress monitoring model by integrating autocad and digital photos. In *ASCE International Conference on Computing in Civil Engineering*, Cancun, Mexico, 2005.

[13] R. Navon. Research in automated measurement of project performance indicators. *Automation in Construction*, 16(1):176–188, 2007.

[14] R.J. Radke, S. Andra, O. Al-Kofaha, and B. Roysam. Image change detection algorithms: a systematic survey. *IEEE Transactions on Image Processing*, 14(3):294–307, 2005.

[15] G. Schindler, P. Krishnamurthy, and F. Dellaert. Line-based structure from motion for urban environments. In *3rd International Symposium on 3D Data Processing, Visualization and Transmission*, Chapel Hill, USA, 2006.

[16] E.P. Simoncelli. Design of multi-dimensional derivative filters. In *IEEE International Conference on Image Processing*, volume 1, pages 790–793, 1994.

[17] E. Trucco and A. P. Kaka. A framework for automatic progress assessment on construction sites using computer vision. *International Journal of IT in Architecture, Engineering and Construction*, 2(2):147–164, May 2004.

[18] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, 2001.

[19] T. Werner, F. Schaffalitzky, and A. Zisserman. Automated architecture reconstruction from close-range photogrammetry. In *CIPA 2001 International Symposium: Surveying and Documentation of Historic Buildings, Monuments, Sites, Traditional and Modern Methods*, Potsdam, Germany, 2001.