NOW YOU SEE IT: THE CASE FOR MEASURING PROGRESS WITH COMPUTER VISION

T. C. Lukins[1], Y. M. Ibrahim[2], A. P. Kaka[2] and E. Trucco[1]

[1]*School of Engineering and Physical Sciences, Heriot-Watt University, UK.*
[2] *School of the Built Environment, Heriot-Watt University, UK.*

Email: t.lukins@hw.ac.uk

Abstract:
The task of measuring the progress of construction is often a subjective process that is prone to error and frequently out-of-date information. The need to recognize completed work feeds into many aspects relating to cost control, scheduling and interim payments. Established photogrammetry techniques and advanced reconstruction tools are available for creating and comparing 3D models of the current site. However, these often involve intensive user interaction and have slow turnaround. In this work we propose the advantages of a fully automated approach using Computer Vision to provide timely and accurate feedback of site progress. We illustrate these benefits with a simplified test case highlighting some initial results based on a building component model. This framework has potential as a valuable aid for project management, enabling tighter control and greater efficiency.

Keywords:
Computer Vision, Progress, Project Control.

1. Introduction

1.1 Motivation and Background

Modern medium-to-large scale construction projects are complex affairs, requiring a great deal of planning and management. The core problem remains one of remaining on schedule – whether by organizing materials and tasks along identified critical paths; realizing and correcting errors to maintain quality; or imposing standard techniques and adopting recognized best practices. In parallel with these concerns, run the ever constant monitoring required to manage the costs of the project; enable payment of contracted work; and to stay within budget.

The tendency of the industry to adopt standardized systems and processes aims to address a substantial portion of these issues. This is based on the control offered by structured *object models*, and has resulted in much recent research into the automation of project planning and control. In essence, most such systems feature at their core a "4D model" involving the 3D design expanded over the time-line of the project. This facilitates the appearance, transition, and completion of individual components, as dictated by the schedule and the monitored status of construction work. In this way the overall costing and control of the project can be visualized and managed, since updates to the status of the components drive and feed report generation.

Such systems can formalize interoperability and transfer of construction data through the use of the IAI *Industry Foundation Classes*. These aim to provide the support for a new generation of tools in which the component model incorporates not only geometric structural information, but relates the instance within the entire project plan in terms of cost, scheduling, supply, contracting, services, and many other facets evolving over time. This methodology is now widely seen in new *nD systems* that allow for the first time a fully integrated approach to construction management. (Lee *et al*, 2003, Fu *et al*, 2006).

The basis of dividing work, or grouping components (for example, into associated work packages) is a common industry approach for project management. In essence these involve compete portions of the project that can be separately allocated and performed. They can involve a single component, or more commonly, support a whole range of preparatory work across the site. The process of division can be unique to individual contractors, and can be particular to their working practices. However, standard approaches can be recognized. Observing the completion of such packages serves as important landmark events in the project, and is a key feature of current methods for surveying site progress.

What is of particular interest is the ability to achieve the recognition of component completion on demand, and automatically, with minimal human intervention. Multiple benefits could be gained for the overall management of the project based on timely feedback. Even, for example, simply reporting a percentage complete for individual components would prove incredibly useful. In particular, the ability to measure progress would enable the accurate calculation of interim payments, to measure overall productivity to date, and ultimately support cost and schedule control.

Navon and Sacks (2006) noted that effective project control needs two kinds of information. The first is the plan that typically comes in the form of a list of activities to be performed, broken down in terms of a performance indicator. The second kind of information is a measurement of the actual performance based on the same performance indicator. While the first kind of information can be

obtained from experience, past records and the building model in question, the second kind can only be obtained through measurement. The traditional approaches to obtaining these measurements have been known to be subjective, costly and too infrequent to allow for effective project control. The way forward, according to Navon and Sacks is to automate the process.

A number of studies have been conducted in the area of automatic progress monitoring of construction projects. These studies employ technologies that range from laser scanning, radio-frequency tags, GPS, barcode, web applications or a hybrid of two or more technologies to monitor the progress of various aspects of construction. Navon (2007) gives an up-to-date review of these technologies. Some of these approaches can be very sound in monitoring various construction operations but their high costs are often a concern for practical applications. One of the most economical ways to track progress is by recording videos or taking photographs, and this is not a new approach in construction management. The challenge however, is to *automatically extract progress* information from the images and report on the status of the project at various levels of detail with minimal or no human intervention. This is the thrust of the present study.

This work is related to our initial investigation into automatic progress estimation using Computer Vision (Trucco and Kaka, 2004). In the rest of the paper we describe the use of alternative techniques, particularly focusing on the core desire to automatically recognize discrete construction components – which would ultimately help to assess the completion of entire work packages. By highlighting the issues involved, and giving an example simplified test case, we aim to justify the feasibility and further put forward the case for "closing the loop" on progress measurement by the means of Computer Vision.

1.2 Photogrammetric Approaches

The idea of using imagery to record and – more importantly - measure buildings is not a new one. The Prussian architect Albrecht Meydenbauer originally coined the term photogrammetry in 1867 as the process to produce topographic plans and elevation drawings from photographs. This key desire - the measurement of 2D or 3D objects from photographs - has grown into a standard and popular technique for the survey and archival of construction projects. When used to capture close-range aspects of buildings, the distinction is then often made that this is different from those alternative approaches that capture from greater distances associated with traditional "remote sensing". For example, recovering top-down models from aerial photography (Mikhail *et al*, 2001).

In many instances the staple techniques in photogrammetry involve recovering information about the position and properties of the camera used to capture the

images. Establishing these properties enable the process of *calibrated rectification*, by which 2D distances in the images can be extracted by the removal of perspective effects, and thus related to actual schematics. Comparing images from different images can be further fine-tuned using *bundle adjustment* to refine the relationships between multiple measurements. These processes are incorporated into many advanced tools such as Canoma, ImageModeler and PhotoModeler. However, the fact remains that these require substantial human interaction to selectively mark-up and correlate sample images – although the process can be automated up to an extent (Van den Heuvel, 2003).

A recently proposed example of using such tools to verify the progress of work has been to extend the current practice of regularly taking images of construction work (Memon *et al*, 2005). In this the authors directly compare a recovered "as-built" 3D model (generated by hand using PhotoModeler) with the AutoCAD design phase output. Through this they seek to adopt and unify the existing surveying of a site by photographs to an integrated database that tracks the schedule of work. This approach shows the potential that exists for better control and overall quality improvement for monitored construction. However, it suffers from the relatively slow turn around and technical expertise required to reconstruct the 3D geometry, even for relatively simply structures.

In many respect, a large amount of similar work in the field of augmented reality has also been carried out with the techniques developed from photogrammetry. The concern here is the reconstruction of a "reality model", i.e., a geometric representation of the current structure that can be merged and overlaid with the virtual model representing the final design, service features and final fittings (Klinker *et al*, 2001). Many additional techniques are employed in order to align the perceived world with the projection of the 3D model. It is however extremely challenging to be able to process the incoming video of the site in real-time, to relay the overlay to the as user as he or she moves around. The use of distinctive fiducial markers, which must be automatically discovered and their positing recorded in the scene, can greatly help in this process. However, this is necessarily an intrusive solution, requiring additional planning and work to place these markers over the site at visible and obtrusive locations.

Alternative modalities – such as laser scanning – have been the focus for other research in close-range architectural work. This is particularly suited to looking at the problem of defect detection since a great amount of detail can be captured (Gordon *et al*, 2003). The use of laser however is expensive and raises a whole set of new problems, especially with performing the capture, and in simply dealing with the vast volumes of data generated. This often involves some intensive post-processing methods for "cleaning" the data. For example, in detecting that the

million of so range points actually lie on the surface of a single wall, which can be represented in a far simpler way.

1.3 The Computer Vision Perspective

Many of the techniques developed by photogrammetry concerning the geometry of camera calibration and location, also lie at the core of Computer Vision (Hartley and Zisserman, 2004). The distinction is that Computer Vision aims to derive such information *automatically*, such that ideally no (or minimal) user interaction is required. This then involves further techniques for recognition, analysis and classification of what objects, or properties, are present within the scene. Fundamentally, most problems are concerned with extracting meaningful structure from the data (i.e. images), and to make sense of these based on previously learned knowledge or representations. This is a task that humans perform effortlessly, but one that is extremely hard for a computer to do well.

With regards to close-range images of buildings, most of the recent research in Computer Vision has been concerned with "reverse engineering" of architectural models. For example, (Schindler and Bauer, 2003) make extensive use of the presence of vanishing points and lines within architectural scenes, in order to detect and recover piecewise the constituent planes that make up the exterior of the building. Such work can produce very realistic looking models, particularly since the image can also be use texture the model and make it look more detailed and realistic, even though the underlying geometry is often coarse and ill refined.

This approach, by exploiting constraints commonly encountered in architecture (such as orthogonality, symmetry, repetition and parallelism), is exploited in a number of other developments. Importantly, it does not require an *a priori* knowledge of a model of the building, only the common "rules" that many human structures follow. The desire for automatic recovery is further driven by applications extending beyond single reconstructions of houses - towards entire towns and cities. This is motivated by the demands for content generation of urban models to accompany on-line mapping and satellite data - e.g. Google Earth. For example (Cornelis *et al*, 2006) describe a recent system that can recover the structure of entire streets from a fitted multi-stereo camera system.

An alternative view is that of matching known structure to the observations in the images: performing *model-based object recognition* (Trucco and Verri, 1998). These techniques have a long pedigree in Computer Vision and can be seen as the equivalent of camera *pose estimation* - since to find the location of the camera, and to find the model that fits the scene, reveal the same thing. Application of such techniques to images of buildings are most generally employed to provide an estimate of location, particularly in urban environments when satellite and other

services can fail due to “canyon” effects resulting in lack of coverage. The work of (Klinec, 2004) is an example of this where they use a comprehensive 3D city model to match to visible building images and so recover the exact position of the user. Again, such work intersects with that of Augmented Reality, when attempting to compensate for motion and real-time model based tracking (Reitmayr and Drummond, 2006). These systems also rely on fusion of geo-referencing and orientation sensors in order to achieve a good initial estimate.

The culmination of the work by (Dick *et al*, 2004) offers an approach that bridges the middle ground, by having at its core a generative model of individual “Lego-like” primitive geometric components (for doors, windows, wall, etc). This is coupled within a probabilistic framework that takes account of the distribution of such components over a building – as learned from training 3D data and experienced architects. Each primitive is further parameterised in order to allow for different sizes and lengths. The objective is to then find a composite architectural model and set of parameters that best explains the visible scene.

The previous related work to this research (Trucco and Kaka, 2004) offers an alternative image based “iconic vision” approach that is used to first help locate images of particular components within the site. This is based on using a suitable metric to find the distance between two images – one of the prototype components, and the other querying for a particular scene. In adopting a statistical approach, resilience to viewpoint and lighting can be gained. The idea is that it could then act as an indexing stage prior to comparison to a CAD model, in order to scan and provide an initial estimate of location. This work is related to the further topic of *change detection* as approached by Computer Vision to establish and statistically interpret the differences between sets of images (Radke et al, 2005).

2. Challenges with Interpreting Images of Construction

The great advantage to using images is that they are extremely quick, easy, and non-intrusive to capture. However, deriving information on structure from images is intrinsically a hard problem, since it is often ill defined and requires additional assumptions and prior knowledge. In particular, images of construction sites contain high amounts of *clutter*, i.e. elements confusing an automatic recognition system; including shadows, reflections, occlusions, different materials, equipment, and people. All this can make the discovery of the true structure of a building extremely difficult.

In general, there are two possible routes for capturing progress images of the site. One possibility is to have access to fixed security and web-cameras directed at the work in progress. The disadvantages here lie with the inflexibility to adapt to

changes in the structure – particularly as portions become occluded. This requires substantial forethought and planning in deciding where to place the cameras and supporting network infrastructure for complete coverage. However, such an approach offers the possibility of constantly on-demand images, taken from a known location, and can be integrated with site security systems.

The alternative is to have regular surveys conducted on foot. The advantages here are that higher resolution and better quality images could be captured from any given location, especially in response to changes. And yet, this flexibility can lead to problems in reconciling the location of the capture, although, following a strict protocol and utilizing further geo-referencing technologies could alleviate this. A combination of both approaches could also be performed.

Having then gathered suitable data, and given the architectural model for expected structure, two further approaches are then possible for matching. One way is to recover a complete 3D model from the images first and then try to match them to the model. Given that the recovery of structure from images is such a hard problem, and so prone to error, this is only really possible for large-scale structure. The subsequent matching of 3D models is also difficult given the complexity of the representation.

The other approach is to try and re-establish the 3D camera pose by matching the back-projection of 2D structure to features in images. This is effectively model-based recognition, and is reliant mainly on the ability to reliably extract matching *features* (e.g., lines or points) from the image data. Since this operation is performed on the data itself there is no intermediate step that can introduce errors and increase computation. The matching is still difficult, since it is non-linear to perform and must be able to take account of potential *outliers* in the discovered features (elements of the scene can not be explained by the model).

A number of techniques can then be introduced when it comes to matching the model to scene features. In particular, *reducing complexity* by employing methods from computer graphics – such as back-face and occlusion culling - to remove those elements of the model that cannot be seen from a proposed camera position. Furthermore, the relationships between individual component locations can be exploited to provide additional *contex*t when matching. For example, knowing that a column is part of a colonnade (consisting of a number of very similar components) would mean that all columns, or their pediment, must also be used in order to identify the individual component. The matching process can then become analogous to a search, in which the reliability of finding one element leads to support for the location of its neighbours.

The accuracy with which this match can occur can be defined by the *resolution* of the model and the images. The ideal of being able to identify individual fittings

and small interior details requires additional captures which generally lead to increasing ambiguity. For example, a picture of an interior doorway and surrounding wall gives very little information regarding its location or identity within the model. Ultimately, the question of scale is dictated by the requirement for what is useful for the purposes of management and estimation.

Once components have been discovered within the scene, the task then switches to being able to confirm their existence. Initially, through the process of matching, it is hopefully established if particular components are actually present at all. This issue is further related to combing matches from multiple viewpoints, since a particular component may only be visible from certain angles. In the case of visibility from a number of views, then increased reliability can be gained by combining estimates. However, ambiguity must again be resolved in order to confirm that the component is self-consistent across all views, and to establish those portions of the site where no information is available.

Given that a component can be seen, and that the matching verifies that there is structure in the scene that supports its presence – then additional processing must occur to establish its status. In particular, the *texture or colour* of the identified region can be analysed, since in many cases this is indicative of the current stage in the construction lifecycle. For example, a column may be shuttered, cast, rendered or painted. Seldom is a component simply "filled up" like an empty glass. Progress is really an assigned percentage to the visible work, as seen across the entire work-package. Learning to accurately estimate and recognize such variations associated with each stages of a components lifecycle, depending on locally used materials and conditions, is a particular challenging problem.

3. A Simple Example

3.1 The Data

This simplified test case represents two phases of construction of a holiday house in Turkey. Geometric information for instances of columns, slabs, walls and a base, are provided as shown in Figure 1 below. The assemblages of these components describe the difference gained by adding the first floor and some of the exterior brick walls. Accompanying these models are a number of images taken from a variety of angles around the site – at reasonably close range such that the majority of the entire building is in view.

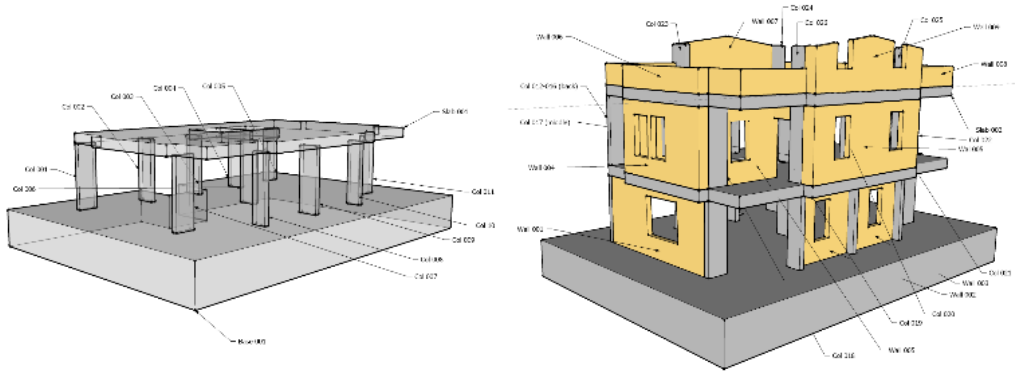

Figure 1: Individual labelled components for two phases of construction.

3.2 Extraction of Scene Structure for Matching

We first apply a standard Canny edge detector to the input image generating a map of high-contrast image points, or contour elements. From this we join the individual edges into line segments if they are of a certain size (at least 30 pixels) and if they can be joined with other edges within a certain angle (<0.05 radians) and gap (<=15 pixels). Final results are around 2000 individual segments – as defined by the co-ordinates of the two end points as show in Figure 2. Notice however that shadowing can create additional edges, and that regions of similar contrast can results in none. Further refinement of these segments could be achieved by only accepting those that lie towards a vanishing point, or can be matched via a Hough transform, or limited to those that occur start at a corner point. These could possibly further reduce the amount of clutter and reveal true structure.

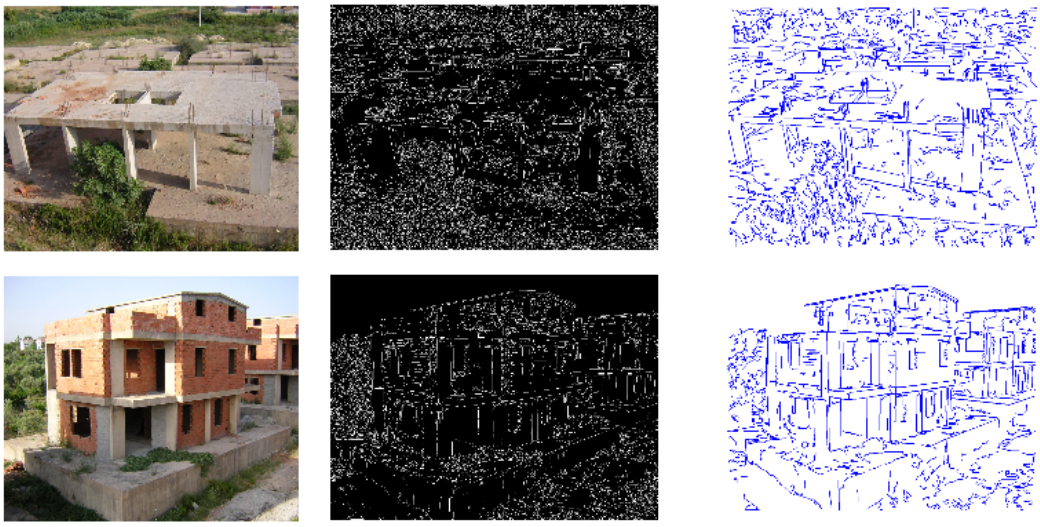

Figure 2: Images (left), edges (middle), and extracted line segments (right).

3.3 Location of Components in the Scene Structure

Having then extracted the dominant line segments in the scene, we can then attempt to match to the 3D data. Our approach is a model based fitting approach to pose estimation, which exploits the context provided by fitting all components, but removing those edges from the model that are not visible from the suggested camera pose. This distance between back-projected edges and the discovered line segments is derived by first representing each line as an end point (*x*, *y*) and angle (*theta*). All edges over a certain length (15 pixels) are split into smaller segments and each segment is represented twice – since each line can be said to have two directions (David and DeMenthon, 2005). The final distance is the Sum Squared Error between matching closest model and image segments.

We then use a standard *particle swarm optimisation* (Clerc and Kennedy, 2002) technique to attempt to minimize this distance. This is a suitable non-linear technique that combines a good degree of freedom to generate possible poses – yet will constantly converge toward the best overall solution. We aim to further help this minimization by constraining the camera angle, such that it is never too close to the model, and that it never looks completely away from the centre. These safeguard against degenerate solutions in which no lines would match and would only confuse the algorithm. The initial starting pose (which must be relatively close to the image) is shown in Figure 3.

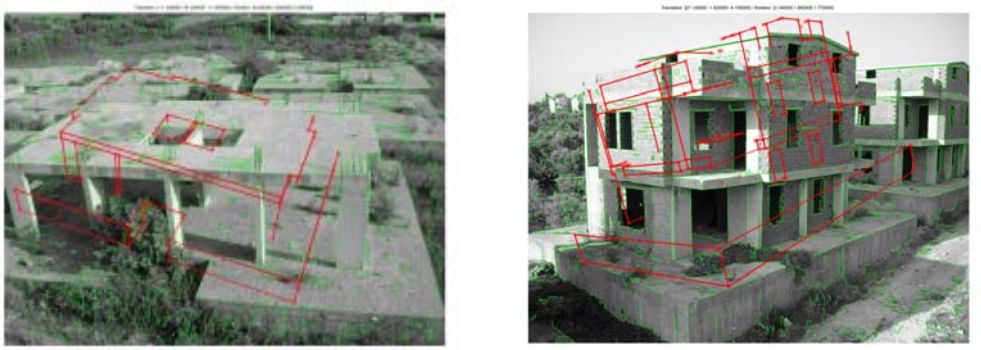

Figure 3: Initial model poses projected onto image and discovered line segments.

3.4 Evaluation of Individual Components

Once the optimisation has converged the model has been aligned to the image line segments. From this global fitting we then further attempt to optimise each unique component in turn. The idea here is that any serious deviation from the individual pose solution gives and indication that *local* structure is sufficiently present in the image to justify its presence. Conversely, any component that cannot be supported by the data (without the support of the surrounding context of other components) will result in a radically different pose. If the camera location or angle varies by more than 5% we thus reject that component. The final output for found components is shown in Figure 4 below.

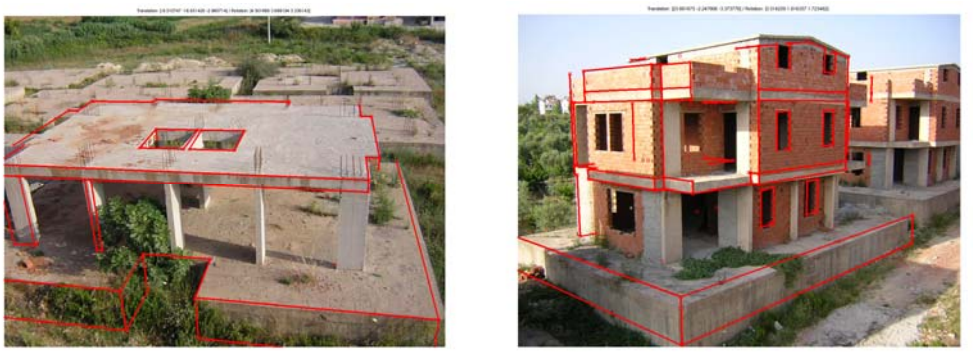

Figure 4: Final pose and components found.

4. Conclusion

In this paper we have proposed the possibilities and advantages for estimating progress of construction by automatically locating components in images of the

site. We have summarized how the trend to achieving this has been driven by recent work focusing on control for construction management, standard component frameworks, and the ability to leverage design phase 3D models. Having presented a review of the current state of the art, and the challenges posed by the complexity of construction sites, we showed how techniques in Computer Vision could potentially provide a solution. While this has potential, many problems still remain for not only finding components, but in then accurately measuring their level of "completeness". The main benefit offered by this work is for a possible means of maintaining on-demand productivity metrics at a finer granularity than currently supported by occasional site survey. Ultimately such information can be used for quicker release of interim payments on the basis of completed work-packages, and to enable integrated cost and schedule control.

For our future work we aim to build more robustness around the prototype framework we propose here for finding components. Currently a considerable amount of parameter adjustment and initial pose alignment needs to be performed to guarantee a solution. Ideally, we would like to remove all dependencies on user interaction, or the need to specifically tailor the system to a particular site. We also wish to push the limits toward much more realistically complex and useful test cases in order to truly verify the potential for this approach. This will lead us to tie in the status of multiple components to estimation of entire work package progress. Furthermore, there are other possible uses for Computer Vision analysis of regularly recorded images of the site. These include route/activity analysis – showing current "hot spots" of work with the potential to modify layout for increased productivity and safety. Alternatively, attempting to actually recognize what the materials and equipment is on site (as opposed to the structure) would be possibly useful to identify deliveries and provide an up-to-date inventory for management of resources.

Acknowledgments:
Turker Bayrak provided the holiday house images and original model. This research is funded by the UK EPSRC grant EP/C535200/1.

References:

Cornelis N., Leibe B., Cornelis K. and Van Gool L. (2006), 3D city modeling using cognitive loops, Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission, Chapel Hill, USA.

David P. and DeMenthon D. (2005), Object recognition in high clutter images using line features, Proceedings of the International Conference in Computer Vision, Beijing, China.

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

Clerc M., and Kennedy J. (2002), The Particle Swarm-Explosion, Stability, and Convergence in a Multidimensional Complex Space, IEEE Transactions on Evolutionary Computation, 6 (1): 58-73.

Eade E.D. and Drummond T.W. (2006), Edge Landmarks in Monocular SLAM, Proceedings of the British Machine Vision Conference, Edinburgh, UK.

Fu C., Aouad G. Lee A et al (2006), IFC model viewer to support nD model application. . Automation in Construction, 15 (3): 178-185.

Gordon C., Boukamp F., Huber D., Latimer E., Park K. and Akinci B. (2003), Combining reality capture technologies for construction defect detection: A case study, Proceedings of the 9th Int. Conference on E-Activities and Intelligent Support in Design and the Built Environment, Istanbul, Turkey.

Hartley R. and Zisserman A. (2004), Multiple View Geometry in Computer Vision. Cambridge University Press.

Hu J., You S. and Neumann U. (2003), Approaches to large-scale urban modeling. IEEE Computer Graphics and Applications, 23 (6): 62-69.

Klinec D. (2004), A model based approach for orientation in urban environments, Proceedings of the XXth Conference of ISPRS '04, Istanbul, Turkey.

Klinker G., Stricker D. and Reiners D. (2001), Augmented reality for exterior construction applications, Augmented Reality and Wearable Computers.

Lee A., Marshall-Ponting, G. Aouad et al (2003), Developing a vision of nD-enabled construction. Construct I.T. report, construct I.T. centre, UK.

Memon Z.A., Majid M.Z.A. and Mustaffar M. (2005), An automatic project progress monitoring model by integrating AutoCAD and digital photos, Proceedings of the ASCE International Conference on Computing in Civil Engineering, Cancun, Mexico.

Mikhail E.M., Bethel J.S. and McGlone J.C. (2001), Introduction to Modern Photogrammetry, Wiley.

Navon R. (2007) Research in automated measurement of project performance indicators. Automation in Construction, 16 (1): 176-188.

Navon R. and Sacks R. (2006) Assessing research issues in automated project performance control (APPC), In press: Automation in construction.

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

Reitmayr G. and Drummond T. (2006), Going Out: Robust Model-based Tracking for Outdoor Augmented Reality, Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Santa Barbara, USA.

Schindler K. and Bauer B. (2003), Towards feature-based building reconstruction from images, Proceedings of 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Plzen-Bory, Czech Republic.

Trucco E. and Kaka A.P. (2004), 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.

Trucco E. and Verri A. (1998), Introductory Techniques for 3-D Computer Vision, Prentice Hall PTR.

Van den Heuvel F.A. (2003), Automation in Architectural Photogrammetry: Line-Photogrammetry for the Reconstruction from Single and Multiple Images. NCG, Netherlands Geodetic Commission, Delft.