Capturing and Representing Construction Project Histories for Estimating and Defect Detection

Burcu Akinci, Semiha Kiziltas, and Anu Pradhan

Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213
{bakinci, semiha, pradhan, LNCS}@cmu.edu

Abstract. History of a construction project can have a multitude of uses in supporting decisions throughout the lifecycle of a facility and on new projects. Based on motivating case studies, this paper describes the need for and some issues associated with capturing and representing construction project histories. This research focuses on supporting defect detection and decision-making for estimating an upcoming activity's production rates, and it proposes an integrated approach to develop and represent construction project histories. The proposed approach starts with identifying the data needs of different stakeholders from job sites and leverages available automated data collection technologies with their specific performance characterizations to collect the data required. Once the data is captured from a variety of sensors, then the approach incorporates a data fusion formalism to create an integrated project history model that can be analyzed in a more comprehensive way.

1 Introduction

The history of a construction project can have a multitude of uses in supporting decisions throughout the lifecycle of a facility and on new projects. Capturing and modeling construction project history not only helps in active project monitoring and situation assessment, but also aids in learning from the trends observed so far in a project to make projections about project completion. After the completion of a project, a project history also provides information useful in estimations of upcoming projects.

Many challenges exist in capturing and modeling a project's history. Currently, types of data that should be collected on a job site are not clearly identified. Existing formalisms (e.g., time cards) only consider a single view, such as cost accounting view, on what data should be captured; resulting in sparse data collection that do not meet the requirements of other stake-holders, such as cost estimators and quality control engineers. Secondly, most of the data is captured manually resulting in missing information and errors. Thirdly, collected data are mostly stored in dispersed documents and databases, which do not facilitate integrated assessment of what happened on a job site. As a result, there are not many decision support systems available for engineers to fully leverage the data collected during construction.

This paper provides an overview of findings from various case studies, showing that: (1) current data collection and storage processes are not effective in gathering the

I.F.C. Smith (Ed.): EG-ICE 2006, LNAI 4200, pp. 15 – 22, 2006.

data needed for developing project histories that can be useful for defect detection and cost estimation of future projects; (2) sensing technologies enable robust data collection when used in conjunction with a formalized data collection procedure; however, the accuracy of the data collected using such technologies is not well-defined. Findings suggest a need for a formalized approach for capturing and modeling construction project histories. An example of such an approach, as described in this paper, starts with identifying different users' needs from project histories, provides guidance in collecting data, and incorporates a framework for fusing data from multiple sources. With such formalism, it would be possible to leverage project histories to support active decision-making during construction (e.g., active defect detection) and proactive decision-making for future projects (e.g., cost estimation of future projects).

2 Motivating Vignettes from Case Studies

Four case studies were conducted in commercial buildings with sizes ranging from 3,345 m^2 to 12,355 m^2, where laser scanners and temperature sensors were used in periodically collecting data from the job sites to actively identify defects [1]. In addition, we conducted a case study on a 19 km highway project, during which we identified a set of issues associated with data collection [2]. Currently, we are conducting another study on a 9 km of a roadway project, where we are trying to understand how to leverage and fuse the data collected by equipment on-board instrument (OBI) and other publicly available databases (e.g., weather database), to create a more comprehensive project history to support estimators in determining production rates in future projects [3]. Below summarizes some findings from these case studies.

The need for and issues associated with data collection to support multiple decisions:
Current data collection at job sites seems to support mostly the needs of construction schedule and cost control, yet data from job sites are useful for other tasks, such as active defect detection and situation assessment in current projects and estimating the production rates of activities in future ones. Such varying uses place different requirements on data collection from sites. For example, highly accurate information on geometric features of components is necessary to identify defects [4], versus general information on processes and the conditions under which processes occur (i.e., contextual data) is helpful for estimating a production rate of a future activity [3].

The case study findings showed that most of the data collected on site focused on the resources used for a given task, but the contextual data was rarely collected [2]. Similarly, the available sources of data and related databases did not provide detailed information that can be used to assess why certain production rates were observed to be fluctuating when the same activity was performed in different zones and dates [3]. As a result, the collected data was not useful in helping estimators in picking a production rate among alternatives for a new project.

Issues with manual data collection and utilization of sensing technologies:
Current manual data collection processes utilized at job sites do not enable collecting required data completely and accurately. One of the cases showed large percentages of missing data describing the daily productivities of activities and the conditions under which such daily productions are achieved [2]. Even when collected, the quantity

information was described based on some indirect measures (e.g., number of truck-loads of dirt moved out); resulting in inaccuracies in the data collected and stored.

Utilization of sensing technologies (e.g., laser scanners, equipment OBIs), reduces the percentage of missing information. However, there can still be inaccuracy issues, if a sensor is not well calibrated and its accuracy under different conditions is not well defined. In certain cases, data collected from such sensors needs to be processed further and fused to be in a format useful for decision makers.

Issues with and the need for fusing data from multiple sources:
Currently, data collected on job sites is stored in multiple dispersed documents and databases. For example, daily crew and material data are kept on time cards, soil conditions are described on reports and production data are stored on databases associated with equipment OBIs. To get a more comprehensive understanding of how activities were performed, one needs to either fuse data stored in such various sources or rely on his/her tacit knowledge, which might not be accurate. In a case study, when an engineer was asked to identify reasons for explaining the fluctuations in the excavation work, he attributed it to fog in the mornings and the soil conditions. When the data collected from equipment OBIs merged along with the data collected in time-cards, the soil profiles defined by USGS and the weather data, it was observed that the factors identified by him did not vary on the days when there were large deviations on production [3]. While this showed the benefits of integrating such data to analyze a given situation, the research team observed that it was tedious and time-consuming to do the integration manually. For instance, it took us approximately forty hours to fuse daily production data of a single activity with the already collected crew, material, and daily contextual data for a typical month.

3 Vision and Overview of the Approach

We have started developing and implementing an approach that addresses the identified issues and needs based on the case studies. This approach consists of two parts focusing on: (1) formalizing a data collection plan prior to the execution of construction activities (Figure 1); and (2) fusing the data collected and creating integrated project histories to support decision-making (Figure 2). So far, our research has focused on using such formalism to support defect detection on construction sites and to support estimation of production rates of activities occurring in future projects. Hence, the corresponding figures and subsections below highlight those perspectives.

3.1 Formalization of a Data Collection Plan

The data requirements of decision makers from job sites need to be incorporated prior to data collection. The first step in doing that is to understand what these requirements are, and how they can be derived or specified (Figure 1). Since our research focus has been to support defect detection and cost estimation, we identified construction specifications and estimators' knowledge of factors impacting activity production rates as sources of information to generate a list of measurement goals.

The approach leverages an integrated product and process model, depicting the as-design and schedule information, and a timeframe for data collection, as input. Using this, it first identifies activities that will be executed during that timeframe and the corresponding measurement goals, derived based on specifications and factors affecting the production rates of those activities. Next, measurement goals identified for each activity are utilized to identify possible sources for data collection with the goal of reducing manual collection. Sources of data include sensors (e.g., laser scanners, equipment OBIs) and general public databases (e.g., USGS soil profiles, weather database). With this, the approach generates a data collection plan as an output.

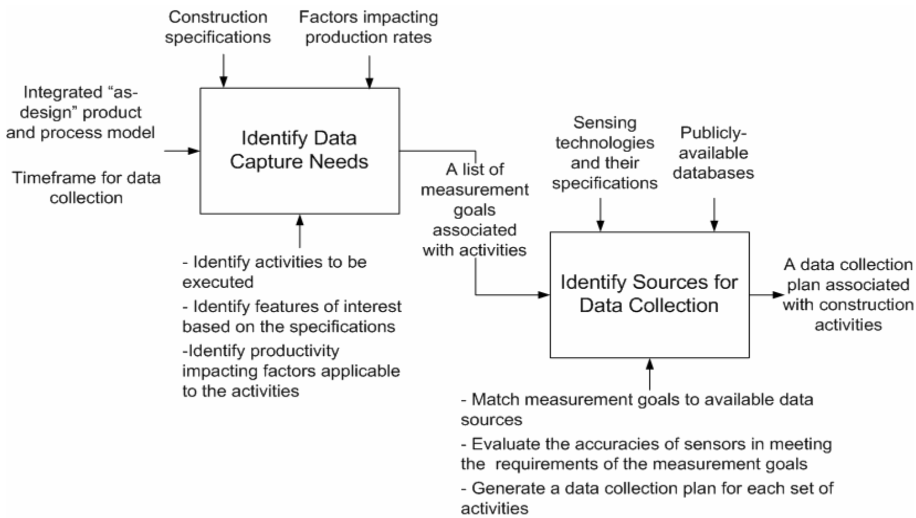

Fig. 1. An approach for generating a data collection plan

Construction specifications provide information on the expected quality of the components by specifying the features of a product to be inspected and the tolerances to deduce the required accuracy for measurements. Specifications can be represented in a computer-interpretable way and can be used to automate the generation of measurement goals for a given set of components [4].

Creating a project history model to be utilized by estimators requires not only product related data, but also contextual data representing the conditions under which the activities were executed, so that estimators can understand what happened in the past, compare it to the current project's conditions and make a decision accordingly. Therefore, within the scope of this research, project history can be defined as as-design project information augmented with activity-specific as-built project data. As built project data are enriched with contextual data, which are captured and stored on a daily basis. We built on the factors identified in the literature (e.g., [3]) and have further extended it based on findings from a set of interviews conducted with several senior estimators in heavy-civil and commercial construction companies. Table 1 provides examples from an initial list of factors identified for excavation, footing and wall construction activities. The factors identified can be grouped under the categories of *design-related factors*, *construction method-related factors*, *construction site-related factors* and *external factors*. Based on these factors, it would be possible to generate a list of measurement goals for each activity, map them to a set of sensors and publicly-available databases that would help in collecting some of the data needed. As a result it would be possible to generate a data collection plan associated with each group of activity to be executed.

Table 1. Initial findings on the factors affecting productivity of activities

Factor Groups	Specific examples of factors for excavation, foundation an wall construction activities
Design-Related Factors	Depth of cut, height, length and width of components
	Shape of cut/shape of component
	Total quantity of work for the entire project
	Number and sizes of openings in walls
	Existence of steps on walls and footings
	Rebar/Formwork to concrete ratio, rebar size
Construction Method-Related Factors	Type and capacity of equipment, number of equipments
	Crew size and composition
	Stockpile dirt vs. haul off
	Method of forming and type of bracing used, formwork size
	Material characteristics, such as concrete strength, soil type
Construction Site-Related Factors	Site access constraints and space availability
	Moisture content of soil
	Length, grade, direction, width of haul roads
External Factors	Time of year, weather, project location

3.2 Data Fusion and Analysis for Creating and Using Project Histories

Once a data collection plan is generated, it can be executed at the job site to collect the data needed. The next step is to process the data gathered from sensors and databases and fuse them to create an integrated project history model that can serve as a basis to perform analysis for defect detection and cost estimation (Figure 2). Different components of such an approach are described below.

3.2.1 Utilization of Sensors for Data Capture

Many research studies have explored utilization of sensors on construction sites for automated data collection (e.g., [1,5,6]). In our research, we have explored the utilization of laser scanners, Radio Frequency Identification (RFID), Global Positioning System (GPS), thermocouples, and equipment OBIs to capture data on job sites for supporting active defect detection and estimators' decision-making. The selections of these technologies are based on the lessons learned from past and ongoing research projects within our research group [1,6]. Our experiments demonstrated that such technologies can enable the capturing of some of the data needed for project history models in an automated way [6].

Our experiments also showed that the behaviors of such sensors vary greatly at job sites; hence manufacturers' specifications might not be reliable in varying situations. For example, we observed that the reading ranges of active RFID tags was reduced by about 1/4th or 1/5th of the specified ranges when they were used to track precast elements [6]. Similarly, laser scanner accuracy varies considerably based on its incidence angle and distance from the target object [7]. While in most cases, the accuracy and the reliability of the data were observed to be better than the manual approaches, it is still important to have a better characterization of accuracies of sensors under different conditions (e.g. incidence angle) when creating and analyzing project history models. Currently, we are conducting experiments for that purpose.

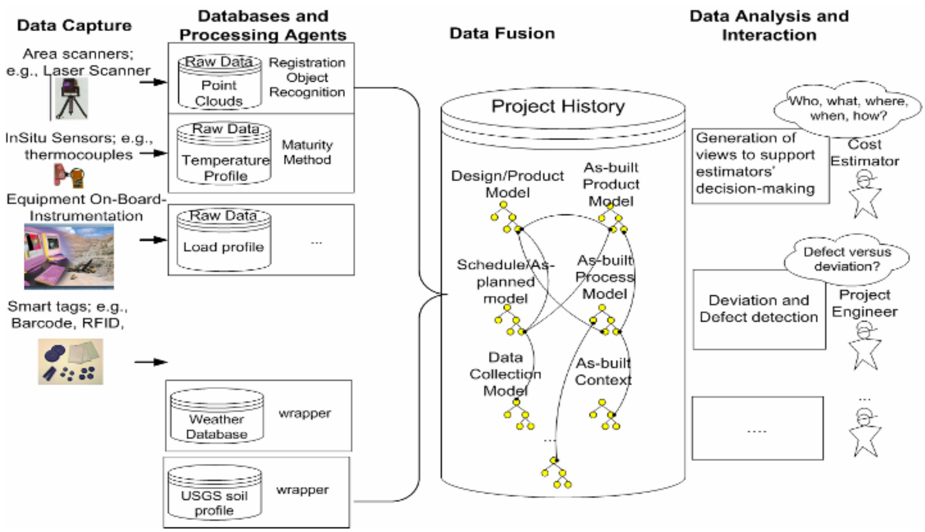

Fig. 2. An approach for data fusion and analysis for creating and using project histories

3.2.2 Formalization of Fusing Data from Multiple Sources

Data collected from multiple sources need to be fused to have a more comprehensive assessment of a project. We have started to develop and evaluate a system architecture for data fusion purposes, based on Dasarathy's fusion functional model [8], where the entire fusion processing is categorized into three general levels of abstraction as, the data level (*sensor fusion*), the feature level (*feature fusion*) and the decision level (*decision fusion*).

In *sensor fusion*, the raw data from multiple sensors, which are measuring the same physical phenomena, are directly combined. For example, the data collected from GPS and RFID readers can be directly combined after initial corrections to track the location and ID of components respectively [6]. However, some sensors, such as laser scanners, cannot measure a component and its geometric features directly and hence, the data collected needs to be processed further and fused with other data at a *feature and component level*. In one of our research projects, laser scanner is being used to detect geometric deviations, i.e. length, height and width of building components [1]. Since laser scanners provide point cloud data, the components and their features needed to be explicitly extracted from point clouds using 3D computer vision techniques [1]. The sensor and feature level fusions are done with appropriate processing agents (Fig 2).

The third level of fusion described in [8] is the *decision level fusion,* where the data fused at the sensor and feature levels are further integrated and analyzed to achieve a decision. We are leveraging different models such as-built product/process model and data collection model for decision level fusion (Fig 2). Decision level fusion is challenging compared to sensor and feature level fusions, since the formalisms used in sensor and feature level fusions are well defined and can be identical across multiple domains. However formalisms for decision-level fusion differ among domains since they need to support different decisions [9].. As discussed in Section 3.1., different tasks require different sets of data being collected and fused. Hence, the decision-level fusion requires customized formalisms to be developed to enable the integration and processing of the data to support specific decisions. In our approach, decision-level fusion formalisms are designed to generate the views (e.g. from the estimator's perspective) that are helpful in supporting decisions to select a proper production rate. These formalisms are not meant to perform any kind of predictions or support case-based reasoning.

3.2.3 Formalisms for Data Interaction and Analysis to Support Active Defect Detection and Cost Estimating

In this research, we have explored project history models to support defect detection during construction and in estimating production rates of future activities. An approach implemented for active defect detection leverages the information represented in as-design models, construction specifications, and the as-built models, generated by processing the data collected from laser scanners. It uses the information in specifications to identify the features of the components that are of interest for defect detection and compares the design and as-built models accordingly. When there is a deviation between an as-design and an as-built model, it refers to the specifications to assess whether the deviation detected exceeds the tolerances specified. If it exceeds the tolerances, then it flags the component as a defective component [4].

In supporting estimators' decision-making, we have been focusing on identifying and generating views from integrated project history models, so that estimators can navigate through the model and identify the information that they need to determine the production rates of activities in future bids. Initial interviews with several estimators from two companies showed that estimators would like to be able to navigate through production data in multiple levels (e.g., zone level, project level) and in multiple perspectives (e.g., based on a certain contextual data, such as depth of cut), and be able to compare alternatives (e.g., comparing productions on multiple zones) using such a model. These views will enable estimators to factually learn from what happened on a job site, and make the estimate for a similar upcoming activity based on this learning. We are currently implementing mechanisms to generate such views for estimators.

4 Conclusions

This paper describes the need for capturing and representing construction project histories and some issues associated with it for cost estimation and defect detection purposes. The approach described in the paper starts with identifying some data capture needs and creating data collection plan for each activity to satisfy those needs. Since several case studies demonstrated that manual data collection is inaccurate and unreliable, the envisioned approach focuses on leveraging the data already stored in publicly-available databases and data collection through a variety of sensors. Once the data is captured from a variety of sensors, they should be fused to create an integrated project model that can be analyzed in a comprehensive way. Such analyses include defect detection and situation assessment during the execution of a project, and generation of information needed for estimators in determining the production rates of future activities.

Acknowledgements

The projects described in this paper are funded by two grants from the National Science Foundation, CMS #0121549 and 0448170. NSF's support is gratefully acknowledged. Any opinions, findings, conclusions or recommendations presented in this paper are those of authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Akinci, B., Boukamp, F., Gordon, C., Huber, D., Lyons, C., Park, K. (2006) "A Formalism for Utilization of Sensor Systems and Integrated Project Models for Active Construction Quality Control." Automation in Construction, Volume 15, Issue 2, March 2006, Pages 124-138

[2] Kiziltas, S. and Akinci, B. (2005) "The Need for Prompt Schedule Update By Utilizing Reality Capture Technologies: A Case Study." Constr. Res. Cong., 04/2005, San Diego, CA.

[3] Kiziltas, S., Pradhan, A., and Akinci, B. (2006) "Developing Integrated Project Histories By Leveraging Multi-Sensor Data Fusion", *ICCCBE,*, June 14-16, Montreal, Canada.

[4] Frank, B., and Akinci, B. (2006) "Automated Reasoning about Construction Specifications to Support Inspection and Quality Control" , *Automation in Construction*, under review.

[5] Kiritsis, D., Bufardi, A. and Xirouchakis, P., "Research issues on product lifecycle management and information tracking using smart embedded systems", Advanced Engineering Informatics, Vol. 17, Numbers 3-4, 2003, pages 189-202.

[6] Ergen. E., Akinci, B. Sacks, R. "Tracking and Locating Components in a Precast Storage Yard Utilizing Radio Frequency Identification Technology and GPS." *Automation in Construction*, under review.

[7] Axelson, P. (1999). "Processing of Laser Scanner Data – Algorithms and Applications," *ISPRS Journal of Photogrammetry & Remote Sensing*, 54 (1999) 138-147.

[8] Dasarathy, B. (1997). "Sensor Fusion Potential Exploitation-Innovative Architectures and Illustrative Applications", *IEEE Proceedings*, 85(1).

[9] Hall, D. L. and Llinas, J. (2001). "Handbook of Mulitsensor Data Fusion," 1st Ed., CRC.