DEVELOPING INTEGRATED PROJECT HISTORIES BY LEVERAGING MULTI-SENSOR DATA FUSION

Semiha Kiziltas[1], Anu Pradhan[1], and Burcu Akinci[2]

ABSTRACT

When individuals estimate the production rates of construction activities in a project, they frequently refer to the past production rates achieved in a similar project. Typically, the data represented in historical documents and databases do not provide detailed information depicting the conditions under which activities were executed (i.e. contextual data). Activity-specific contextual data would be helpful for estimating a production rate for an activity since it will enable comparisons of favorable/unfavorable conditions observed in a past project with the conditions expected to occur in an upcoming project. Since current data collection process does not consider estimators' needs from a previous project, most of the relevant contextual data are not collected and when they are collected, they are typically acquired by different parties and stored in dispersed documents and databases. It is time-consuming and tedious to integrate the data stored in dispersed archives manually and make comprehensive analysis when needed. Advances in reality capture technologies (such as equipment sensors, smart tags, laser scanners) provide opportunities to collect some of the data required by estimators digitally and more comprehensively and accurately due to less reliance on manual data collection. However, the relevant data are still stored in dispersed databases. To improve decision-making of estimators, the different types of data colleted from these multiple sources need to be fused and represented in an integrated project model. This paper describes the need for and proposes an approach for fusing data collected from multiple sources to generate integrated project histories to support estimator's decision making.

KEY WORDS

project model, automated reality capture technologies, sensors, decision making, estimation.

INTRODUCTION

Estimators, when they estimate production rates for a new bid, search for actual production rates achieved in similar previous projects and stored in historical documents and databases to make their estimates realistic. Selecting a reliable production rate entails understanding under which conditions the actual production was achieved and assessing how similar the upcoming project's conditions to the conditions observed in previous projects. One of the problems associated with utilization of current project historical documents and databases during estimation of activity production rates is the limited data represented in these documents for estimators. Contextual data, describing the conditions under which an activity

1 Graduate Student, Civil and Envir. Engrg. Department, Carnegie Mellon University, 5000 Forbes Ave. Pittsburgh, 15213, USA, +1 412/268-3781, FAX 412 /268-7813, semiha@cmu.edu, pradhan@cmu.edu

2 Assist. Prof., Civil and Envir. Engrg. Department, Carnegie Mellon University, 5000 Forbes Ave. Pittsburgh, 15213, USA, +1 412/268-2959, FAX 412 /268-7813, bakinci@cmu.edu

was executed, is not captured and stored in detail in current historical documents. Another problem is that even the limited collected data is not stored in a single source but stored in multiple locations, so whenever they need to be utilized, it is quite time consuming to integrate them to get meaningful information for estimators.

Currently, data collection at construction sites does not incorporate the needs of estimators when the project is being executed; hence the opportunity to collect important data (such as type of excavated material for an excavation activity) that will be required later by estimators is missed. Similarly, even the relevant data is collected, it does not get to be integrated for future referral of estimators. Rather they are collected and stored discretely in multiple locations (e.g., daily data kept on time cards, soil conditions kept on reports, resources utilized kept in logistics systems) and it becomes cumbersome to integrate them later when estimators need to have a more comprehensive view of the project.

There is a need for a formalism to enable identification of estimators' activity-specific contextual data requirements from a past project history and to identify how to collect and integrate such data in a project model for future usage of estimators. Advances in current reality capture technologies (such as on board equipment sensors, smart tags) and existing available databases (such as weather database) can enable collecting some of the data required by estimators. However, the raw data collected via some of these technologies needs to be processed further and fused to be in a format useful for estimators. It is quite time consuming, tedious and difficult to manually integrate data collected from such multiple sources. Hence, there is a need for automated multi-sensor data fusion.

Based on a case study conducted on an highway construction project, this paper explains the limitations in current practice of creation and utilization of past project historical documents, highlights the need to develop integrated project histories that incorporate information about products, processes and the related activity-specific contextual data. In addition, the paper describes an approach for fusing data collected and stored at multiple sources to create integrated project histories to be utilized by estimators in the future.

MOTIVATING CASE STUDY

A case study is being conducted on a forty month highway construction project with an estimated cost of twenty-three million dollars, and a schedule including approximately 750 major activities in the CPM schedule. The scope of the project is 5.7 miles of roadway, including construction of three pre-cast reinforced concrete box culverts. The research team focused on the bulk excavation activity underneath the proposed highway section since this activity gets affected from a variety of different factors (e.g., unexpected weather conditions, equipment break-downs, soil conditions) considerably.

The original schedule included forty activities related to bulk excavation and among which fifteen of them had production rates different from the estimated production rates. Table 1 provides an example group of production rates observed for different bulk excavation activities. These fluctuations in production rates were identified from the project's cost report; however that report did not contain detailed contextual data that would help in understanding the reasons behind such fluctuations.

Table 1: Alternative production rates for bulk excavation activity

Phase No	A	B	C	D
Phase no description	Excavator:800cy/hr Trucks: 100 ton Haul distance: 4000'-5000'	Excavator:800cy/hr Trucks: 100 ton Haul distance: 5000'-6000'	Excavator:800cy/hr Trucks: 100 ton Haul distance: 6000'-8000'	Excavator:800cy/hr Trucks: 100 ton Haul distance: 8000'-10000'
Estimated quantity (cy)	254,852	387,252	171,841	59,993
Actual quantity to date (cy)	76,233	76,972	81,118	14,840
Actual prod. rate (cy/mhr)	0.66a*	0.73a	0.45a	0.76a

A study on literature on production analysis of excavation activities shows that there are various factors that might affect the production rate. These factors are type of excavated material, degree of excavation difficulty, water table level, underfoot conditions, capacity and number of excavation and hauling units, cycle time duration of hauling and loading units, haul-road gradient, length, width and conditions, factors (such as traffic) that affect waiting time of equipment, depth of cut, slope, space constraints in the load area, weather condition (e.g. Kannan 1999). Among these factors, we have focused on a sub-group of factors, identified based on extensive interviews with senior estimators from two companies, to further analyze the given situation. Some of these sub-group factors are already incorporated in Table 1 and others are stated in Table 2.

In Table 1, a phase number defines capacities of excavating/hauling equipment and the hauling distance predefined at the estimating stage to enable estimators understand the context of how an activity is being executed. While this information is useful, it does not clearly explain why there are such differences observed at the job site. For example, as seen in Phase C and D columns of the table, the actual production rates were fluctuating considerably although same construction equipment was used and the hauling distances were close to each other. To understand the reasons of fluctuations, we looked at possible sources where some more information on the contexts under which these activities were executed can be obtained. Table 2 provides group of contextual data/factors focused in this research and sources from which such data were accessed for this project. Due to the scarcity of the data stored in the data sources of the project, to explain the fluctuations, other data sources were investigated such as on-board equipment sensors (OBI) utilized on trucks and available

* The actual production rates are not included here as they constitute confidential data. Each production rate used in this table is described relative to the average baseline estimate, which is shown as a cy/mhr.

databases to fill the missing data such as weather from weather database and soil properties from USGS (US Geological Survey) database, as shown in Table 2.

Table 2: Sources of contextual data investigated in the project

Contextual Data	Data source in the project	Collected at job site?	Alternative data source
Actual depth of cut	Survey points	No	OBI
Hauling distance	Daily time cards	No	OBI
Hauling/Loading unit capacities	Logistics database, time cards	Yes	
Type of excavated material	Not available	No	USGS
Activity shift (daytime versus nighttime)	CPM and two week look ahead schedules	Yes	
Weather condition	Not available on a daily basis	No	Weather database

By exploring in detail the data stored in the sources listed in Table 2, it was observed that the way the data was collected was different from the way it was stored. It was observed that, on the CPM schedule, bulk excavation was divided into multiple zones on the planned roadway route to be executed on two shifts (i.e. night shift and day shift) with two different crews on each shift. Data collection was based on locations, whereas recording was based on phase numbers. This resulted in having the production data bundled under the same phase number without differentiating based on other contextual data, such as when it was executed (i.e. night/day shifts), where it was executed (i.e. zones that define soil type and depth of cut).

To have a better understanding of the situation in more detail, we looked at two specific days where two bulk excavation activities with the same phase number were executed on 6/7/2005 and 6/15/2005. As shown in Table 3, on both days the same crew, same equipment and same hauling distances were used on two different zones. The production was twice as much on 6/15/2005; however, the reasons behind this large difference were not captured. Weather database showed that it was raining on 6/15/2005 (0.04 inch), however the production was almost twice as much the one observed on 6/7/2005, during which the weather was normal. Thus, it is evident that factors as depth of cut, soil conditions might also be playing role in the variation of the production rate. It was observed that depth of cuts in two locations were close, however, the soil types were defined to be represented differently on USGS database hinting that it might be one of the reasons leading to the difference. However, it is not possible to define the exact soil type based on the data available. When the truck payload data was investigated, the empty travel times and loaded travel times were 20-25% less than the average times over the same hauling distance for the day with higher production suggesting that there might be some congestion at the site resulting in lower production rate on 6/7/2005. As demonstrated above, such kind of analyses cannot be done without being able to collect and fuse data from a variety of sources.

Table 3: Contextual data required on two days

		Date: 6/7/2005	Date: 6/15/2005
	Production	z cubic yard	1.64z cubic yard
Contextual Data	Actual depth of cut	31 feet	26.5 feet
	Hauling distance	Phase A	Phase A
	Hauling/Loading unit capacities	800 cy/hr excavator	800 cy/hr excavator
	Type of excavated material	LOB[+]	LOD
	Activity shift	Day-time, 8hrs	Day-time, 8hrs
	Weather condition	Normal	0.4 inches of rain

As demonstrated above, potential factors affecting productivity of excavation activity were either collected/known during construction and never get to be aggregated to cost report level (such as weather, soil conditions, sensor data for equipment production); or not collected and missing (such as actual depth of cut, hauling road conditions). In order to utilize the data collected from such multiple sources effectively, one needs to represent them in an integrated way so that one can have a holistic understanding of what is happening at a job site and be able to analyze how things vary across various dimensions.

Based on these observations, we are currently conducting research on 1) identification of contextual data that needs to be collected at jobsites for future decision making, and 2) fusing various types of data coming from variety of sources on a job site to enable the utilization of such project data constituting the history of a project in cost estimation of future projects.

OVERALL VISION

Generation and representation of a project's history in an integrated way can address the first research goal described above and provide a better support for estimating activity production rates. A multi-sensor fusion approach to integrate the dispersed data sources for supplying the need of project history model addresses the second research goal.

First part of the envisioned approach leverages the information in a product and process model, within which an activity knows to which components it is related; which construction method and correspondingly, which resources will be utilized during its execution. The user selects an activity to be constructed to identify estimators' needs. To identify activity-specific

[+] Shows the designated soil type in USGS soil database.

data needs of estimators, the envisioned system provides the user with templates to enable identification of important (defined by the user) contextual data required to understand the conditions under which the activities are executed.

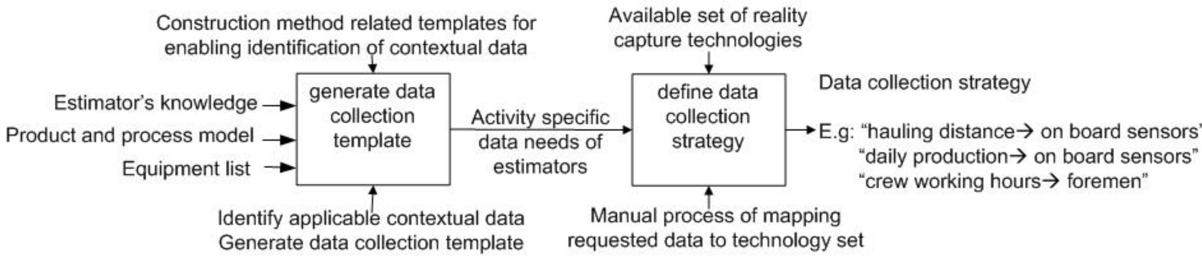

Figure 1: IDEF0 diagram- Identification of estimators' needs and data collection strategy

In this research, contextual data is defined as factors affecting the production rate of activities grouped as design related, construction site related, construction method related and external (such as weather) factors. The user of the system is defined as estimators, whose data needs will be identified and then collected by the site personnel. The system involves mechanisms for reasoning about project model to define factors applicable to selected activities, which are related to the product elements it acts on and which construction method is applied to. These reasoning mechanisms identifies the construction method applied to a selected activity, extracts the applicability facets as "component", "action", "resource type" and the contextual data related to these extracted facets for estimators. The output is activity-specific list of data that needs to be collected at a job site. This set of needs then will be used to identify available data capture technologies to be used for data collection. The output of this second process is a data collection strategy within which each data requirement is matched to a set of data capture technology or agents (such as foremen).

Using the data collection strategy, the project team can collect the data needed, once a project starts. The second part of the approach (as shown in Figure 2) focuses on the project execution phase and targets integration of the data based on the location and the time it was executed on the job site using data fusion techniques.

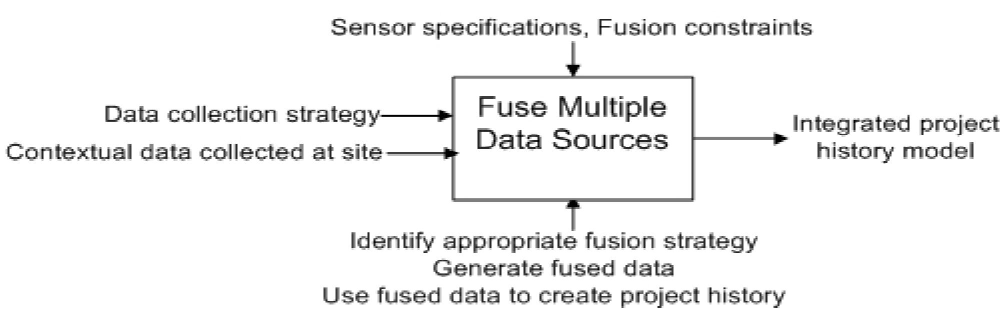

Figure 2: IDEF0 diagram-Data fusion for creating integrated project history models

The idea is to transform the collected raw data such that it is meaningful for estimators and fuse them to be stored in an integrated project history model. In order to fuse multiple data sources, the appropriate fusion strategy needs to be identified. The data coming from multiple sources can be fused based on either activity (temporal fusion) or location information (spatial fusion), or using both the information (spatio-temporal fusion). Once the fusion strategy is identified, the collected data can be fused utilizing the sensor specifications and fusion constraints (availability of hardware or software systems). The output of this process is an integrated project history model to be used in future estimates.

METHODOLOGY

Creating an integrated project history model to support estimating decisions requires first identification of contextual data needed by estimators. Previous studies on productivity analyses provide specific lists defining what needs to be collected on sites that can be considered as contextual data (Kannan 1999, Liberda 2003). These studies either defined these needs at a "project-level" (e.g., Liberda 2003;) or an "activity-level" (e.g. Kannan 1999, Staub-French 2003). The former identifies factors affecting productivity general to a project rather than specific to activities; whereas the latter includes factors specific to given activities. Some researchers at "activity level" focused only on design features affecting the productivity of activities (e.g., Staub-French 2003), whereas others identified a general set of factors affecting production rate of specific activities (e.g. Kannan 1999). Based on interviews, conducted with senior estimators from two companies and background literature, factors effecting productivity of activities and activity specific data that estimators would like to see in project historical documents are identified and combined under four groups as design related factors (e.g., depth of cut for an excavation activity), construction method related factors (e.g., equipment types and capacities), construction site related factors (e.g., type of excavated material) and external factors (e.g., weather). This initial list is used as a basis to develop a vocabulary for enabling identification of contextual data by estimators.

Once it is known what to capture on construction sites, the next challenge is how to collect and make it useful for estimators. With the advents in reality capture technologies, a number of sensor fusion systems applicable to various phases of construction projects were identified to enable automated data collection (e.g. Akinci et al. 2006, Navon 2002). All such research studies exploit the benefits of multi-sensor data fusion to enhance the decision making processes during various construction phases. The challenge is to map the data requirements to a set of available technologies for collection and then to fuse the already captured data from multiple sources.

For the stated challenge, we have started to develop and evaluate a system architecture for data fusion purposes (Figure 3). This architecture is based on Dasarathy's fusion functional model (Dasarathy 1997), where the entire fusion processing is categorized into three general levels of abstraction, the data level (sensor fusion), the feature level (feature fusion) and the decision level (decision fusion). In our architecture (Figure 3), we have three sub-fusion system pertaining each fusion level to avoid creating a monolithic fusion architecture, which is generally not recommended by software architecture community due to its single point of failure and other maintenance issues (Bass et al 2003). In addition, such an architecture facilitates identifying types of problems for each fusion level, and enables recognizing commonality among problems and candidate solutions (Hall et al. 2004).

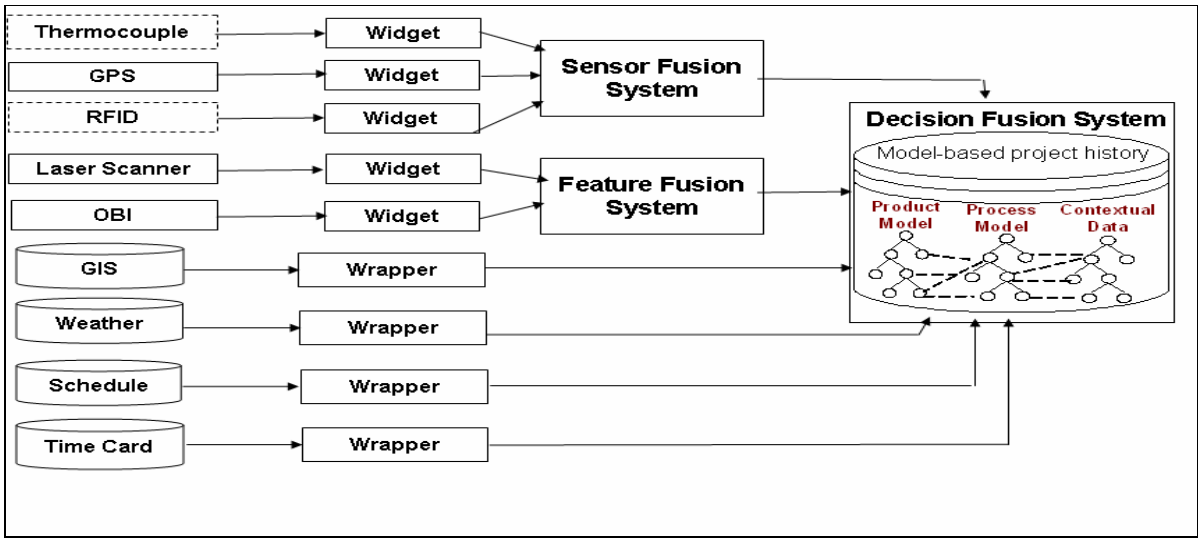

Figure 3: System Architecture for Fusion Process

In sensor fusion, the raw data from multiple sensors such as GPS and temperature sensor, which are measuring the same physical phenomena, are directly combined. However, some sensors such as laser scanner cannot measure feature and its attributes directly. In our case, laser scanner is used to scan an excavated area (i.e. feature) to measure the excavated volume and depth of cut (i.e. feature attributes). In such scenario, the feature and its attributes are explicitly extracted from the sensors' data. However, the difference between sensor fusion and feature fusion levels are not based on sensor-types, but on the context under which the fusion occurs. For instance, if a temperature sensor is used to assess the strength of the concrete, it is considered as feature fusion since the temperature value doesn't measure the strength of concrete directly. In decision level fusion, the information collected from both sensor and feature levels are integrated and analyzed to achieve a decision. For example, the information related to various factors affecting the production rates, such as weather, depth of cut, excavated volume are obtained from either sensor or feature fusions, these information are then further fused and represented in an integrated project model to support estimator's decision making.

Figure 3 shows an example of a set of sensors related to excavation activity executed on highway projects. In the current case study project, we are utilizing most of the sensors/databases, except the ones listed inside dashed boxes in Figure 3. In sensor level fusion for our case study, the location information is obtained from GPS devices. The location based data collection is required for excavation activity, since location specific parameters (such as soil conditions, depth of cut) needs to be collected. Data collected from GPS devices are corrected for possible errors, such as multi-path problem (GPS signal bouncing off a reflective surface prior to reaching the GPS receiver antenna [Trimble 2005]).

Such sensor specific software programs or hardware devices utilized for increasing accuracies of data collected are referred to as sensor widgets in Figure 3.

Similarly, for feature fusion level, laser scanner can be used to obtain the excavated volume and depth of cut. In addition, equipment OBI is used to obtain payload information, and other truck performance information. The equipment OBI employs a number of sensors such as pressure sensor, real-time clock to measure payload and cycle times, respectively. The cycle times and payload information are used to calculate the performance measure, such as productivity of a truck operation.

Once the sensor and feature level fusions are performed, it is possible to measure the features and their attributes. However, this information does not necessarily assist estimators to understand the factors affecting the productivity. For instance, in order to know the productivity of excavation at different stations, the location information of different stations obtained from GPS and the payload data obtained from equipment OBI need to be fused together. In addition, to understand the effect of soil type on payload productivity, the soil type information (obtained from soil database) needs to be fused along with location and payload information. Such 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 identical across multiple domains, but differ among domain in decision level fusion (Hall et al 2001).

To facilitate decision level fusion in our approach, we are evaluating location-based fusion (geo-spatial fusion) technique based on station-based linear referencing system, which is used for assigning and finding the location of any point along a network by specifying the direction and distance from a known given point on the network (Brennan and Harlow 2002). There are many kinds of linear referencing methods, such as milepost, reference post, and engineering station (FGDC 2005). Initially, the engineering station referencing method has been adopted in our approach, as most of the datasets (such as schedule, on board sensor data) were based on station numbers in the case study project. A future extension to this approach is to include both location and activity-based information (spatio-temporal fusion).

As a first step in geo-spatial fusion, we performed dynamic segmentation, a process of transforming linearly referenced data stored in a tabular structure into a geometric feature, on given datasets (schedule, excavation model, weather, etc). The schedule data is transformed into line feature (utilizing From_Station and To_Station information) along a road centerline. Similarly, the excavation activity and soil profile are transformed into polygon features. The selection of proper geometric features for given data is based on the scope of application. If area is important, as in the case of excavation data and soil data, polygon feature is suitable. The second step in our geo-spatial fusion was to perform overlay analysis, where the two or more separate spatial datasets (points, lines or polygons) are combined to create a new output dataset (Maguire et al 2005). In our approach, the soil and location-based payload data were overlaid against each other to understand the effect of soil type on payload productivity. The results obtained from such an overlay analysis can be presented to an estimator in appropriate views to help him better understand the fluctuating payload productivity based on soil types.

CONCLUSION

A detailed case study highlighted some of the limitations of creation and utilization of past project histories for future decision making of estimators. A major problem emphasized in the case was the scarcity of the contextual data required for helping estimators to select a production rate by understanding the conditions under which it was achieved. This problem was attributed to the manual data collection processes that do not consider the needs of estimators proactively during project execution and data storage processes that do not store the data in an integrated way even if they are collected.

Based on this, the paper presented the need for creating integrated project histories along with the methodologies to formalize (a) enabling identification and capturing of contextual data and (b) fusing multiple data sources. In an initial implementation, data from multiple sources have been fused leveraging geo-spatial fusion process. Data coming from these sources were further refined to get the required data by estimators. This initial implementation showed that approach is viable and useful in creating integrated history models for the purpose of supporting the decision making of estimators. Future work will include incorporation of spatio-temporal fusion and component-based fusion to support activities other than excavation.

ACKNOWLEDGEMENTS

The project is funded by a grant from the National Science Foundation, CMS # 0448170 and supported by Trumbull Corp. NSF's and Trumbull Corp.'s supports are 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 and Trumbull Corporation.

REFERENCES

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

Bass, L., Clements, P. and Kazman, R. (2003). "Software Architecture in Practice," 2nd Edition, Addison Wesley, Boston.

Brennan, P. and Harlow, M. (2002). "Linear Referencing in ArcGIS," ESRI Press.

Dasarathy, B. (1997). "Sensor Fusion Potential Exploitation-Innovative Architectures and Illustrative Applications", IEEE Proceedings, Vol. 85, No. 1, 1997.

FGDC (2005). "Framework Introduction and Guide," < http://www.fgdc.gov/framework/handbook> (last accessed February , 2006).

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

Kannan, G. (1999). "A Methodology For The Development Of A Production Experience Database For Earthmoving Operations Using Automated Data Collection," PhD, Faculty of the Virginia Polytechnic Institute and State University.

Liberda, M., Ruwanpura, J.Y., and Jergeas, G. (2003). "Construction productivity improvement: A study of human, managerial and external factors." ASCE Construction Research Congress, Hawaii March 2003.

Maguire, D. , Batty, M. and Goodchild, M. (2005). "GIS, Spatial Analysis, and Modeling," ESRI Press.

Navon, R. and Goldschmidt. E. (2002) "Monitoring labor inputs: automated-data-collection model and enabling technologies", Automation in Construction, 12 (2002) 185-199.

Trimble (2005). "Differential GPS," < http://www.trimble.com/gps/dgps.html> (last accessed October 2005)

Staub-French, S., Fisher, M., Kunz, J., and Paulson, B. (2003). "A generic feature-driven activity-based cost estimation process." Advanced Engineering Informatics, 17, 23-39.