# Summary for CIFE Seed Proposals for Academic Year 2019-20

<table>
<thead>
<tr>
<th>Proposal number:</th>
<th>2019-09</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposal title:</td>
<td>Human and Robot Interaction for Dynamic Updating of Building Information Models</td>
</tr>
<tr>
<td>Principal investigator(s) and department(s):</td>
<td>Kincho H. Law, Civil and Environmental Engineering</td>
</tr>
<tr>
<td>Research staff:</td>
<td>Max Ferguson, PhD Candidate</td>
</tr>
<tr>
<td>Total funds requested:</td>
<td>$92,584</td>
</tr>
<tr>
<td>Project URL for continuation proposals</td>
<td><a href="http://cife.stanford.edu/Seed2018%20Dynamically%20Updating%20BIM">http://cife.stanford.edu/Seed2018%20Dynamically%20Updating%20BIM</a></td>
</tr>
<tr>
<td>Project objectives addressed by proposal</td>
<td>Usable, Buildable, Operable</td>
</tr>
<tr>
<td>Expected time horizon</td>
<td>&lt; 2 to 5 years &gt;</td>
</tr>
<tr>
<td>Type of innovation</td>
<td>Incremental, Breakthrough</td>
</tr>
</tbody>
</table>

## Abstract (up to 150 words)

**The problem:** Having an accessible and constantly-updated digital model of a facility or construction site could lead to significant improvements in safety, energy efficiency and worker productivity. While recent advances in machine learning have made such a system feasible, the interactions between people and an automated system remain important for real-time modeling of a facility.

**The proposed solution:** We will extend our Object R-CNN framework for dynamically updating building information models using computer vision. This framework allows people and automated systems to update the model of a facility. A web-based interface to observe and correct changes to the model will be developed so that an accurate digital model can be maintained.

**The proposed research approach:** Our efforts will focus on improving our current algorithms for updating digital building models using a stream of RGB-D images, and exploring how this dynamic model can be made available and correctable to human stakeholders.

---

1 The PI(s) must be academic council member(s) at Stanford.
2 For this and the next points, delete the answers that don’t apply to your proposal.
Engineering or Business Problem
Automation in construction, maintenance and facility management promises to bring significant productivity gains to a number of industries. During the past year, we have shown that mobile robots can automatically collect information from a work site environment using RGB-D cameras and LIDAR [1, 2], potentially enabling significant advances in automation, safety, and remote project management. Nevertheless, state-of-the-art computer vision algorithms are not yet able to handle the complexity and diversity of objects that may exist in a facility or a construction site. To this end, we believe that a combination of human supervision and modern machine learning techniques are required to maintain an accurate dynamic model of the facility or construction site. Through human supervision, the capabilities of a machine learning system will continuously and progressively improve, eventually leading to a system that can handle the geometric complexity and diversity of a real construction site. The core research objective is to explore how a stream of sensor data from a mobile robot or drone be combined with human supervision to form a real-time semantic-rich model of the environment in real-time, as illustrated in Figure 1.

For robots to operate in dynamic and unstructured environments such as construction sites or medical facilities, they must be able to infer or obtain a spatiotemporal, semantic-rich model of the environment. Such a model may include static information about the facility, such as the position of door and walls. To facilitate safe and efficient automation, the model should also contain dynamic information about the environment, such as the position of other robots in the facility. The position of movable objects, like furniture or safety cones, might also be of interest. As a concrete example, we consider a cleaning robot which depends on a real-time model of the building to plan out cleaning activities. As an example, suppose the computer vision algorithm on the cleaning robot fails, and incorrectly predicts that there is a safety cone blocking its path along the hallway. To enable cleaning to continue, a human operator can manually remove the safety cone from the virtual building information model, reflecting the real scene at the site. Our research will focus on how human supervision can be integrated with our existing computer vision algorithms, allowing continuous learning and improved fault tolerance.

Maintaining a real-time digital building model can be useful for managers and automated systems, throughout the entire building lifetime. During construction, the model can be used to enable greater automation, as well as providing a real-time source of information for construction managers and related professionals. In the operations phase, the spatiotemporal information can be

![Figure 1. The primary research problem involves automatically updating a detailed geometric model using sensor information from a mobile robot. Contrary to previous research, we plan to investigate how human supervision can be used to assist this process in complex environments.](image)
used to ensure that automated processes are operating in a safe and efficient manner. By incorporating human supervision, the ability of the automated data-capture system can evolve and improve with time, making it useful throughout the building lifecycle. Based on the above, real-time building information can influence the following aspects of a facility: (a) Buildability, by facilitating automated construction processes (b) Operability, by enabling the safer operation and efficient management of mobile robots in the facility. (c) Sustainability, by automating the collection of data for decision making related to improving productivity during construction, and efficiency in facility operation. Additionally, real-time semantic information about the building can be used for controlling and inspecting productivity, quality, and safety, which alleviates risks and opens new pathways for owners, operators, designers and builders.

Theoretical and Practical Points of Departure

The proposed research project builds on the success of our previous CIFE project, titled: “A Framework for Updating Building Information Models with Mobile Robots”. In this project we have developed an algorithm for updating building information models using RGB-D images [1]. Objects such as safety cones and trash cans are automatically detected using RGB-D images and added to a geometric model of the environment. The dynamically generated model is exposed to human operators and automated systems through both a web-based user interface (UI) and an application programming interface (API). The system is also capable of extracting object attributes, such as the color, size, and material of each object. One of the notable limitations of our system is that it does not provide a way for a human operator to rectify incorrect object predictions, or train the model to detect new objects. While this does not limit the quality of the research outcome, it does significantly limit the applicability of this system to real-world problems.

In recent times, there has been significant development in automatically extracting semantic information from 2D images [3–6] and 3D point clouds [7, 8]. Methods developed at CIFE have focused on semantic segmentation of large point cloud datasets [9] as well as real-time object detection and model updating [1]. Companies such as Doxel and Einsite are beginning to adopt this computer vision technology for various applications in the Civil Engineering domain. Commercial tools such as Revit, ArchiCAD, and Vectorworks are commonly used to create, update and edit building models. However, these tools are not designed to support automated real-time updates from robots or building sensors. That is, we are yet to see any tools that allow a human operator to directly interact with a real-time building information model, let alone a geometric model that is being actively updated by machine learning algorithms. Additionally, it is not clear how these next-generation BIM tools will work, and what limitations they may have.

In real-time BIM, models are constantly updated using some form of sensor data. Researchers have proposed using real-time BIM to assist crane navigation during building construction [10]. A good discussion of real-time BIM is provided by Hwang and Liu, but they emphasize that the integration between real-time systems and BIM remains very low [11]. Some recent commercial offerings from architecture, engineering and construction (AEC) software vendors lean towards the integration of real-time data with BIM [12]. However, there does not appear to be any offerings which are tailored to low-latency real-time building information modelling. To the best of our knowledge, there are no existing tools or well-defined methodologies that tightly integrate machine learning systems with human supervision, as we are proposing.

The concept of active learning systems is currently popular in the machine learning field [13, 14]. Researchers are interested in developing systems which continue to improve at a given task, either with or without human supervision. A similar approach has been applied by many major
corporations to develop commercial systems which constantly improve, such as web search, product recommendations and advertising [15]. However, due to the increased technical complexity of computer vision and spatial modelling, there has been little investigation into the use of active learning in BIM.

In short, there is a lack of methods that address the problem of automatically updating a spatiotemporal model with machine learning and human supervision. Moreover, the potential impacts on the industry from obtaining such knowledge have not been quantified or fully explored.

**Research Methods and Work Plan**

This project will build on the theoretical framework for dynamic building information models, outlined in our previous project. The previous framework and accompanying implementation provided a method for updating building information models using streaming data from a mobile robot. We will introduce two new concepts to this framework: supervision and observation. Supervision is the idea that a human operator or an automated system observes the changes to the geometric model, and provides corrections to the underlying predictions. Observation is the idea that, a person or an automated system, watches the geometric model and uses the real-time data for the purpose of decision-making, automation, or improved safety.

The proposed works can be organized into three main categories: (1) improving our computer vision algorithm for automatically identifying common construction site objects; (2) exploring how people and automated systems can interact with, modify, and utilize a dynamic geometric model; and (3) validating the proposed framework with field tests. Each category is now described in more detail:

1. Extending the Object R-CNN algorithm to synchronize machine observations with a semantic-rich geometric model of the environment. In this stage, we will extend our Object R-CNN algorithm to update a geometric model based on observations from an RGB-D camera. Contrary to our previous approach which only adds objects to the model, we will now focus on updating, editing and removing objects from the model. Depending on feedback from industry partners and the Technical Advisory Committee, we will focus on detecting common construction site objects, such as safety cones and signs, or common building objects such as furniture.

2. Exploring how people and automated systems can interact with the real-time, updated geometric model. The computer vision algorithm will undoubtedly make errors when generating a semantic-rich model from point cloud data. We will explore how an operator can interact with the automatically generated model, and possibly correct misclassifications. Additionally, we will explore how automated systems can use the dynamic model for path planning and collision avoidance (for robotic applications).

3. Validating the proposed framework using field tests. To validate the proposed framework, we will conduct a number of real-world tests. For demonstrative purposes, we will use our mobile robot to maintain and publish a dynamic model of the Y2E2 building for the latter half of the project. We will also explore how our algorithms and platform can be connected to existing data collection systems to enable real-time system information.

**Expected Results: Findings, Contributions, and Impact on Practice**

The biggest anticipated outcome of this research is a novel and elegant dynamic building information framework that will better facilitate bidirectional communication between mobile robots and building information models. The main components of the framework are shown in
This framework will be an ideological extension of the traditional BIM ontology, designed to better support automation in the construction site and modern facility. By incorporating human supervision, we believe that the resulting framework and algorithmic implementations will be beneficial to a wide range of stakeholders, from those in AEC to robotics. Additionally, we expect that this contribution will be of interest to parties in both industry and academia. The framework will provide a platform for future research in real-time building information modelling. Looking further ahead, the framework represents a step towards increased automation in a modern facility or construction site.

If successful, the secondary outcome will be a computer vision algorithm capable of automatically updating a geometric model, based on streaming data from a mobile robot or fixed camera. An interface will be developed to allow a human operator to interact with the computer vision predictions and make corrections where necessary.

**Industry Involvement**

The proposed work aims to contribute to productivity gains during the construction and operation phase of the building lifecycle, and hence it will be of interest to stakeholders in construction, facility management, operations, and maintenance. This proposal is prepared partially based on our conversations with CIFE member companies (Hilti, Glodon, Bouygues Construction) as well as companies such as Einsite and Doxel. We will continue to seek expertise, feedback, guidance and collaborations from these companies. We believe that the development of supervised computer vision algorithms will be highly beneficial to many existing and emerging AEC technology companies. We also expect that the research will be particular applicable to stakeholders who are interested in using robots to automate construction, operations, or maintenance. We aim to showcase our algorithms and framework using data collected from the field through collaborations with CIFE members and other companies (such as Einsite).

![Diagram of information flow](image)

Figure 2. Flow of information from mobile robots to the building information model. Research will focus on how computer vision algorithms and human supervision can be used to maintain a detailed and dynamic model of the built environment.
Additionally, the research team has successfully demonstrated machine learning in the earthquake engineering, energy, manufacturing and IoT domains and has received best research paper awards from ASCE, ASME and IEEE conferences on the subject. Research student has recently won the Engineering Category of the Pacific Earthquake Engineering Research Center’s Hub ImageNet Challenge [16]. Working with the National Institute of Standards and Technology (NIST), the research team has also been actively involved in developing standard representations of machine learning models. Specifically, the Gaussian Process Regression model has been accepted in PMML Version 4.3, supported by the Data Mining Group (DMG). The research team is currently working with NIST, SoftwareAG and IBM towards a standard representation for Deep Learning (CNN) model. Direct interactions with companies and standards organizations should provide the research team ample opportunities working with a broad spectrum of industries.

**Research Milestones and Risks**
The following four milestones, which correspond with the research outline described above, will be used to measure the progress of the proposed research:

1. Development of an interface allowing a human operator to correct predictions from our current computer vision algorithm (supervision).
2. Extension of the Object R-CNN network integrated with human supervision to predict new objects and new object attributes
3. Validation of the dynamic building information framework and computer vision algorithm, using a mobile robot in the Y2E2 building. Validation of the algorithms and framework using data from a construction project.
4. Submission of papers to peer-reviewed journals and conferences, final report preparation, and industry demonstration of prototype platform

A timeline of the proposed research is provided in Figure 3.

The risks are primarily associated with the inter-dependency of different milestones; for example, the validation of building information framework is affected by the accuracy and degree of maturity of the object recognition algorithm. To mitigate such risks, multiple strategies will be employed including:

- Researchers will focus on developing the different components of the system in parallel, relying on valid sets of assumptions about the structure of the input data at each milestone
- Some of the previously developed algorithms for 3D object reconstruction [1] will be used to streamline the development of the proposed solution.

Other risks include the ability to deploy a computer vision algorithm on an embedded system. Fortunately, the core computer visions algorithms and robot hardware have been developed and tested as part of the previous project.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Quarter 1</th>
<th>Quarter 2</th>
<th>Quarter 3</th>
<th>Quarter 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Computer Vision Algorithm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Framework Design and Implementation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Validation and Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Report Preparation and Demonstration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Timeline of research milestones. A large amount of time is dedicated to validation and evaluation, as validation work will be conducted using a physical UGV in a relatively large facility.
Next Steps
The proposed work will contribute to productivity gains during the construction and operation phase of the building lifecycle, and hence it will also be of interest to stakeholders in construction, facility management, operations, and maintenance. We expect that the framework will be particularly applicable for stakeholders who are interested in using robots to automate construction, operations, or maintenance. Furthermore, we will pursue government funding opportunities such as NSF, NIST, FEMA, Office of Homeland Security and others. To test our result with industry, we will seek feedback from our industry partner through continuous collaboration. Future work will explore how the dynamic building information model can be used with artificial intelligence techniques to automate the management of mobile robots in the building.

References