Project Title:

BRIEF Visualization

Authors:

A. Tarek Hatata and Thomas Magnan

Synopsis of Program:

This proposed project is an extension of a pre-existing “human-in-the-loop” application utilizing two robots and a turntable to image and grasp objects within a predetermined region.

This project focuses on advancement of the applications of the optical robotic arm. The end goal is to add functionality of the controlling GUI to easily move the arm and obtain images from other perspectives to build a refined 3D model of the object being imaged. Ideally this process can be automated through machine learning and analyzing an image stream to determine where another image is required.

The next stage of the project includes taking a series of images of this object over time and connecting the 3D models to create a 4D model where one can see the changes over time (the 4th dimension). An ideal object to model in three dimensional space over time is a plant, specifically a flower.

Keywords: Kinect, Computer Vision, Modeling, Reverse-Kinematics, Robotics, OpenCV

Intellectual Merit:

This Small Business Innovation Research Phase I project has a variety of technical challenges that can be broken down into two main areas: (1) the movement and location of the imaging arm, and (2) determining where new images are needed to make an accurate model and meshing models together to create a ‘4D movie’.

To achieve these goals, the optical arm must have the ability to orient itself, so that what it sees can be accurately modeled. Once that is achieved reverse kinematics can be used to easily move the arm from one point in space to another to get a new image. Ideally this process intelligently avoids all other obstacles in the space and finds alternative paths where necessary.

Broader Impact:

The first significant impact that this project relates to the 3D modeling technology alone. Through this project one will be able to obtain an accurate 3D model of an object and model over time. One potential application of this technology includes monitoring plant growth in a lab setting. This technology will minimize the monitoring and meticulous notes required by lab personnel when observing growth or changes on a daily, or even hourly, basis.

The second significant impact that this project will provide incorporates the grasping arm technology in a few ways. The 3D models created by this project can be further analyzed to grasp the object and even move it (e.g. helping a disabled person drink a glass of water). Additionally the algorithms created to control the movements of the camera arm can be modified to enhance the functionality of the grasping arm.

Elevator Pitch:

Good Morning, everyone! Who here made themselves breakfast or a cup of coffee this morning? Probably most of us, right? I know I might not be up here talking to you guys without my morning cup of Joe. Imagine not being able to simply pick up the cup and bring it to your mouth. There are over a quarter million people in the United States that can't perform this simple task due to spinal cord injuries.

My is Tom Magnan. And I am Tarek Hatata. And we are here to talk to you about a technology that would make these simple tasks possible - tasks that able-bodied people take for granted everyday. This technology focuses on creating efficient and accurate 3D models of real world objects.

To conduct our research we are working with a two-armed robotic setup. One centrally located grasping arm and a second optical arm on the perimeter of the turntable setup. This is a laboratory setup that can be easily converted to a breakfast table for the aforementioned example.

We know a lot of you are probably thinking that technology already exists to create models of real world objects. That is true, but none of those technologies are currently efficient enough to use our intended applications. The burden associated with a spinal cord injury is bad enough, they should not have to wait two hours to accurately model their coffee mug before taking a sip. Our technology uses a "human-in-the-loop" methodology to significantly reduce the time needed to model.

Through our GUI, a user will be able to select a pre-existing shape (i.e. coffee mug, bowl, spoon) and apply that mesh to the image stream and have it adjust to the real world object. We are focusing on an efficient algorithm so that given the first image, we only need to take pictures from obstructed angles rather than continually looping around the object like current technologies.

Because of the nature of the setup, the optical arm will understand its location relative to other objects in the "world" through reverse kinematics, so it can readjust its path and avoid obstacles. It will also have the ability to take its image stream and reorient itself with landmarks or tags in its view to ensure accurate data in the event that the arm went too far or too short.

In addition to assistance for the disabled that we already talked to you about, the modular technology can also be extended to a variety of laboratory settings. One that we are simulating in our current research is using the technology to replace the need for additional laboratory assistants in doing daily measurements of plant growth. This could all be done by our software.

We hope that you, too, see the desperate need for technology like ours. We can only make it possible with your help.

The Commercial Opportunity:

The market for 3D modeling is one that spans across multiple industries. Architecture, for example, utilizes 3D modeling to accurately model a potential home and allow a future homeowner to see what his or her home will actually look like. Another big industry that relies on three-dimensional objects is the film industry. As movies like *The Avengers* or *Avatar* that utilize computer-generated imagery (CGI) continue to dominate the box office, the need for more accurate CGI will grow as the viewers start to expect more realistic visuals. The landscape of a digital artist is also quite competitive, and the accuracy of their 3D models make these artists more appealing to films that require CGI.

The industries that we're addressing and people that we're targeting with our research do not normally utilize this type of technology. It's rare to find a laboratory that automates their botanic research or someone that has suffered a spinal-cord industry to take advantage of 3D modeling. It'll be important to convey to these customers the benefits of our technology, specifically saving them time and money and ultimately improving the efficiency of things that they do every day. As a biologist, our technology aims to save them the money to hire a lab assistant to monitor a plant, and may even make their measurements or other data more accurate through more accurate 3D modeling. As a victim of a spinal cord injury, our technology will save that individual the time to perform rudimentary tasks, such as the coffee example mentioned previously, and can ultimately improve any physical task that the individual has to perform.

There are many companies that have software to create three-dimensional objects, including SketchUp and even Autodesk Maya. These companies developed software to allow the average consumer to easily create objects in 3D, from simple cubes to very realistic video game characters. Our product will introduce a new way to model 3D objects, a way that will change the way the consumer even thinks about 3D modeling. Rather than building an object from scratch, the user will have the ability to take an image of an object, identify that object, and allow our software to accurately model the object for them. Saving the user the long and meticulous hours spent getting every detail right modeling an object will be a huge advantage of our technology over the competition.

The key risk in bringing our technology to market is the ability to convince the consumer that previous three-dimensional modeling technology is obsolete and that our product is unique. From a consumer's perspective, it's difficult to physically look at two different 3D models of the same object and choose one over the other. However, conveying the power and efficiency of our modeling algorithm and software will influence our consumer to choose our technology because of the importance of time. If someone with a spinal cord injury simply wants to model a coffee mug to pick it up, they'd rather wait the least amount of time to accurately model that object, and hence choose our product. The same holds true for biologists, who would rather spend time working on more advanced tasks than simply monitoring and measuring the growth of a plant.

The way in which our project would be commercialized would involve a few different aspects. While the timing of our product launch or the location of the launch wouldn't be essential, it is important to consider exactly *who* we want to target, because 3D modeling isn't exactly an innovation that everyone needs in their lives. Ideally, we'd start by targeting botanists through websites like ScienceDaily or even television networks like the Discovery Channel. But as with any product, our project will need to be summarized concisely and in a way that will convince a typical botanist to try our technology. For those that suffered from spinal-cord injuries, it'll be important to fully outline exactly how our technology can improve their quality of life from the perspective of an average person.

As we've stated, this project specifically targets biologists and other scientists that have a need for plant examination and manipulation. Eliminating the need for a lab assistant to measure and manipulate a plant will save that biologist a lot of money and allow him or her to focus on other tasks. Glassdoor estimates that the national average salary of a lab assistant is $29,203/year. While many science labs can run on a budget of upwards of $1,000,000, many labs including those of a university do not have the same resources that a lab at Google may have. And saving tens of thousands of dollars can often allow researchers to invest in other areas and explore new innovations that they otherwise could not because they had to pay the salary of a lab assistant.

Societal and Global Impact:

The goal of our research is to create a modular enough algorithm and product that can be extended far beyond the initial issues that we are trying to address when conducting the research. At the current stage of our research our focus is two-fold: (1) addressing inefficiencies in conducting biological research, and (2) enhancing the quality of life for those affected by neurological or spinal-cord injuries.

Everyday some of the world's most brilliant minds spend their days taking care of plants. They make meticulous growth measurements, water them, and rotate them to make sure that they get enough sunlight. Our technology would be able to do all of this for them. It would even provide more accurate measurements that are digitally stored and easily analyzed.

By maximizing the use of a scientist's mind, through the use of our technology, they can spend more of their time working on new theories and projects. Not only would it benefit the individuals, but society as a whole. Each of those potential projects that they are now able to spend more time on would advance many of the fields studied by these researchers such as the environment and food. These projects could include studying plants to better protect our watersheds, more efficient green roofs, or preserving a dying plant necessary for environmental equilibrium. Even if one does not see the importance of the environment, each of these steps taken result in dollars saved on energy, which everyone values.

Many scientists who would utilize our technology also do research focused on food such as maximizing the nutrients we can get out of single plant or single square meter of land. By saving the labor costs associated with the projects and allowing biologists to further the field, greater society would gain from better quality food at a lower price. In turn would assist in the battle against food insecurity in america, access to healthy fresh food, and our nation's weight problem.

As we mentioned, the second focus of our research focuses on restoring a better quality of life to those who are affected by spinal-cord injuries. Eating breakfast or having a morning cup of coffee are just a few things that able-bodied persons take for granted. Our technology would allow everyone, regardless of ability, to enjoy these things with minimal assistance. This is where the efficiency provided by our modeling algorithm is particularly useful over other modeling software. It allows people to drink their coffee while still hot, and cereal not mushy.

As mentioned previously, our work would help almost every citizen who purchased food using our technology. It would most directly help the biologists working on these advancements the most by saving them the cost of doing these monotonous tasks. Our technology would be the most beneficial to smaller labs who may have equally brilliant scientists working in it but not the resources to pursue larger projects due to the cost of managing plants. Our technology would allow them to compete on this stage and potentially bring forth scientific advancements that would otherwise never occur without our algorithms.

In the context of the spinal-cord injuries, obviously would assist the disabled the most by increase quality of life. It would instill in them a newfound sense of confidence, and ability to be a larger contributing member of society. And by extension, our technology would also the their loved ones - the ones spending so much of their time currently being a caretaker. Many of the tasks that they spend hours of their days doing out of love and support could be done by our technology. This would allow them to work, have time to themselves, and provide the emotional support that our technology cannot provide. Potentially allowing them to change the world for the better in ways we can only imagine.

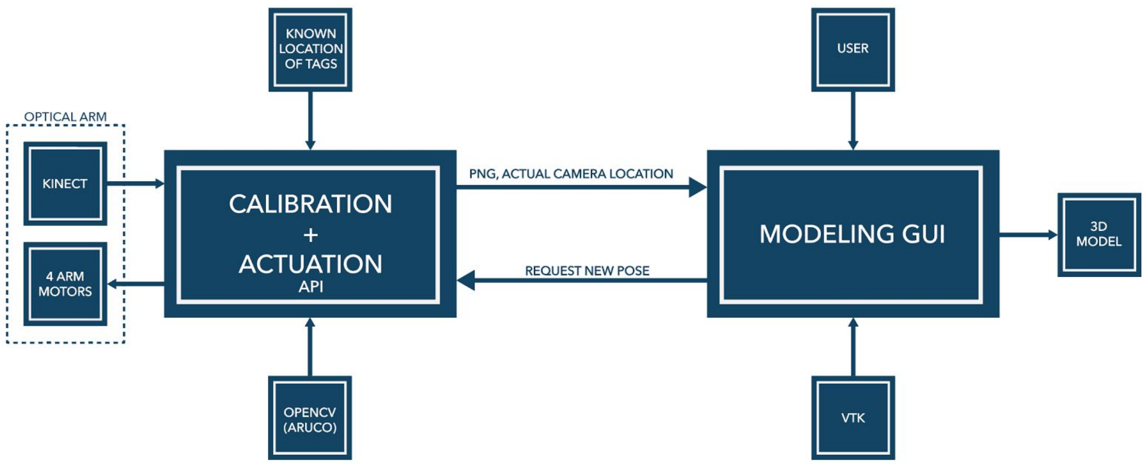
Like the introduction to any technology, there are a variety of potential issues associated with our technology. Although our technology could allow for advancements in environmental issues, currently the system must be powered up and operational at all times for maximum efficiency. No environmental studies have been done to know what the footprint of doing so has been done at this point. That being said, the high cost to produce and setup we hope would deter the manufacturing of unused setups, limiting this potential environmental issue related to physical resource waste.

Cleanliness is crucial with the use of our product. Before release we plan to encase all exposed parts to prevent things such as dust or grease to compromise the plant growth or enter one's food. Assuming these standards are met we do not foresee any health related risks with our product. Yet, our product is not appropriate for children. It still requires a certain level of intellect to operate and could result in one harming themselves if misdirected. Regardless of any potential risks associated with our product, our product is equally functional and accessible across race, gender, sexuality, ability and/or disability.

At this time we do not see any potential unethical uses for the technology, but understand that we cannot possibly think through every possible use for a technology. For that reason if we were to be so lucky to take this product to the public we would be sure to emphasise the importance of only using our product for good. Everyday we all make a conscious choice to be ethical or unethical and we challenge everyone who uses this to make the ethical choice. For these reasons we do not feel that governmental regulation is required for our product at this time. We remain open to the idea and willing to negotiate this point at any point.

Many of these benefits and issues can easily be extended on a global scale. Our technology can easily be modified due to its modularity to assist disabled people around the world speaking different languages. The time saved by scientists could also allow them to connect on a global level and use that time to collaborate on projects that would otherwise be impossible.

Technical Discussion:



The two main technical challenges we foresee in bringing our application to market surround efficiency and automating our modeling process to eliminate user interaction through machine learning. The biggest selling point of our application is its efficiency in producing accurate 3D models, and to achieve that we want to ensure communication between the calibration API and modeling GUI is smooth so that the user is quickly presented with their requested poses. The optical arm should have efficient communication between the Kinect camera and the arm motors, so movements are quick and accurate. Our algorithm to build 3D models must efficiently reconstruct objects based on the user's adjustments, and while we believe our underlying algorithm is already more efficient than the market, we want to improve it even more through machine learning.

Ultimately our algorithm relies on a "human-in-the-loop" methodology, where the user identifies a common object that is then reconstructed from predetermined characteristics of the selected object. We believe that the image recognition and reconstruction can be automated to simply eliminate the user interaction all together. The user simply identifies the object to the application, and everything from selecting different poses for complete modelling to reconstructing the object will occur behind the scenes. However, we see this as the biggest challenge to complete, and consider it more of a future enhancement than our immediate goal. Our focus is to build a user-facing application to build 3D models easily and more efficiently than others on the market.

As previously explained in the Broader Impact section, we believe that 3D modeling is becoming more important in society today, from making impacts in the medical industry to the entertainment industry. The modeling process is considered 3D reconstruction, and the early method to construct a 3D object involves photogrammetry, the science of making measurements from photographs. The 3D meshes produced by these applications are dense and while they can be highly accurate, they often take a very long time to produce. Our algorithm is a process of image recognition, when the user identifies a common object, followed by image reconstruction, where the algorithm takes predefined characteristics of that object and adapts them to the images from the optical arm to build a 3D model. While our algorithm is limited in the number of objects to model, our approach uses known geometry of the object and a small number of poses to model the entire object accurately, and do so much quicker than existing applications.

As an example, Microsoft created a *3D Scan* application that allows users to make 3D scans in real-time using a Kinect camera. To scan an object, the user must hold the Kinect camera in one position and slowly rotate the camera around the object for a full 360 degrees. Our optical arm holds the Kinect camera in position, and the modeling GUI communicates with the arm to capture requested camera poses until the object is modeled sufficiently in 3D, reducing the time needed to rotate completely around the object. For the plant monitoring example mentioned previously, our optical arm will capture the data necessary to model the plant in 3D and determine any physical changes from previous measurements or models.

During this phase of our research we hope to complete a series of core objectives: (1) successfully locate and calibrate the robotic arm, (2) efficiently move the arm, (3) create a 3D model with as few images as possible, make this process possible through (4) the creation of a modeling GUI that the user can interact with and build their model off of a (5) library of pre-built meshes. Before we can accomplish these objectives we must answer a variety of questions to determine the technical and commercial feasibility previously discussed. Many questions have been answered throughout the document, but some of these questions include:

- Where will we call the center of the world?
- What type of coordinate system should we use?
- What models do we to have meshes premade for?
- What angle is a good angle have between the two poses presented to the user?
- How accurate do our locations need to be?
- What is a good enough margin of error for the robotic movement?
- How many VR tags are required for the desired level of accuracy?
- How efficient do we need to the motors to move?
- What options for translation/rotation will be available in the GUI?
- Should the functionality be modular enough to extend to work with different cameras?
- How easy will it be for users to set up on a new machine?
- What competitors are in the market that we need to be concerned with?
- How much will it cost to market the product?

To accomplish these objectives and provide the answers to these questions, we have set forth a series of technical milestones:

- Determine a coordinate system and world center
- Determine a list of meshes that we would like to give the user, and build them out one by one
- Successfully integrate with Aruco to interact with VR tags for calibration
- Successfully connect Kinect camera with opencv as a video source
- Perform sufficient tests with the Kinect and 1-4 VR tags on the turntable from various angles
    - For each of these angles manually measure the location to compare against the readings
- Socket communication between our code base and the python-based API for controlling the motors
- Control actuation at varying degrees of freedom: (1) Rotation - stepper motor, (2) Height, (3) Radial distance.
- Allow user to stretch, rotate, scale, or otherwise manipulate the model to more accurately represent the model in 3D space
- Allow user to choose a desired output file format and quickly output a created model
- Create and apply a machine learning algorithm to further improve our modeling process

Some of these milestones focus on technical decisions that must be made before coding out the program can continue, others involve building out layers of functionality, and many focus on testing the functionality of what we built along the way. Most milestones were designed such that they can be done in parallel as much as possible. We have gone into more detail in the follow section breaking out these milestones by core objective.

Research & Design Plan:

To achieve our objectives we have broken our project down into three main modular components with various experiments and demonstrations to support their objective to support the overall objective.

The first objective of the project is to create a GUI that the user can use to create their models while also controlling the amount of automation they choose to give the algorithm when making the model. We have a series of experiments and demos planned along the way to test the functionality. All have been listed previously in this document, but a few relevant to this objective include creating simple presets of common shapes and objects, such a cube or mug, and outputting the model to the desired format portably and efficiently.

Our second objective is creating a modeling API to build out the GUI. The modeling API must present the user with images taken from the optical arm, and have options to select the object to model and actually manipulate the object over the images. We have a series of experiments and demos planned along the way to test the functionality, including testing with cubical or spherical objects that are simple and easy to model.

Our last objective required to obtain photos of the object we are modeling is our Calibration and Actuation API. This API must answer three core functionalities: (1) Where the camera is with respect to the object being modeled, (2) manipulating the four arm motors to move the optical arm to a desired position, and (3) calibrating the camera to account for motor mechanical failures. We have a series of experiments and demos planned along the way to test the functionality.

To accomplish all core functionalities we have broken it down into four levels of implementation: (1) technical design choices and setup, (2) understanding and connecting opencv, aruco, and the kinect on a test machine, (3) final implementation of the code base and aruco tags on final machine, and (4) manual testing and calibration.

These core functionalities are developed in such a way such that they can be worked on in parallel for levels 1 and 2 of implementation. All core functionalities must be at level 3 before moving onto level 3. Elements of level 4 can be tested independently by each core functionality at level 2 but will have to be retested at level 4.

These three core objectives are designs so that they can all be built out independently of one another Once we have built out these three objectives we will perform some integration testing. Even before we can test out full function calls of the API we plan to test the calls with simple print statements within the functions once they have all been built out.
Once integration has been complete we will need to make the code production ready by cleaning up unnecessary code, ensuring proper debugging runtime statements are in place, and packaging it into a zip file with proper installation scripts and documentation.