Computer Vision Reflection #2: Amazon Go Store

Manjesh Prasad, Duy-Anh Dang, Nicolas Yang, and Shuzhu Chen
San Jose State University, San Jose, CA

***Abstract*—This paper presents a summary of our second presentation focusing on how computer vision is utilize , its feature extraction sequence, and a insight of a modern technology.**
***Abstract*—This paper presents a summary of our third presentation, focusing on a system design process, the functional analysis aspects, trade-off components, and an insight into how a flopped modern computer vision technology flopped.**

1. Introduction

Computer Vision is a branch of Artificial Intelligence (AI) and Machine Learning (ML) technologies that perform complex data understanding through means of observation. Similar to how humans understand data through spectating, Computer Vision aims to mimic these traits to help perceive, identify, and understand objects through pattern recognition and other forms of motion. Even though CV is still in its infancy stage of development, it is still considered a leading-edge technology that will consistently drive innovation and allow new products to fully adapt to its advantages. Some of these products include, but are not limited to, Automation, Improved Insights, Enhanced User Experience, Safety, and Security; which make it an outstanding advance technology. For our 3rd presentation, we covered more of a sequential system design approach analysis of modern CV technology and mentioned why the Amazon Go Store failed to properly incorporate CV technology successfully.

2. Communication Gap Between Technical and Non-Technical Individuals

When discussing proper system design and operation, every component traces back to the initial written requirements. These requirements can range from a standard use case presented in a product prologue to user stories and many other system engineering documentations. When discussing the philosophies of constructing optimal system requirements, we must acknowledge the language barrier that often exists between technical and non-technical individuals. It is important considering that most stakeholders tend to be non-technical, and many individuals in the design team are people of technical backgrounds. Our group proposes a sequential pattern to achieve optimal system design with respects to the non-technical stakeholder's requirements. Understanding, translating, and interpreting the requirements provided by non-technical individuals is a critical step in the software development process. It sets the foundation for the next phases of the development lifecycle, and is the difference between a successful product and potential bottlenecks.

In the past, we prematurely covered the requirements analysis phase in our presentation. However, for this presentation, we decided to re-illustrate that particular phase to capture all the key information so we can decipher the requirements. We decided to give the point of view of a verbal requirement of the "Amazon Go Store" a Stakeholder would communicate to an engineering team.

2.0.1. Example of a verbal communication between a business owner to a technical individual. Subsubsection text here. When discussing proper system design and operation, every component traces back to the initial written requirements. These requirements can range from a standard use case presented in a product prologue to user stories and many other system engineering documentations. When discussing the philosophies of constructing optimal system requirements, we must acknowledge the language barrier that often exists between technical and non-technical individuals. It is important considering that most stakeholders tend to be non-technical, and many individuals in the design team are people of technical backgrounds. Our group proposes a sequential pattern to achieve optimal system design with respects to the non-technical stakeholder's requirements. Understanding, translating, and interpreting the requirements provided by non-technical individuals is a critical step in the software development process. It sets the foundation for the next phases of the development lifecycle, and is the difference between a successful product and potential bottlenecks.

In the past, we prematurely covered the requirements analysis phase in our presentation. However, for this presentation, we decided to re-illustrate that particular phase to capture all the key information so we can decipher the requirements. We decided to give the point of view of a verbal requirement of the "Amazon Go Store" a Stakeholder would communicate to an engineering team.

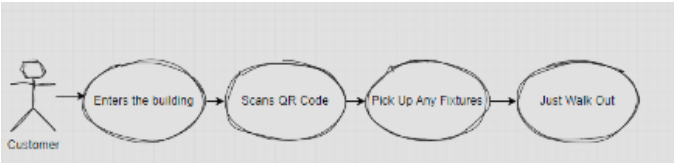

Figure 1. This is an example of Amazon Go Store "Just Walk Out" Procedure in a simple sketch format

Business owner would say one requirement

- **Requirement 1:** "I want a customer to enter a building, scan a membership QR code, pick up any fixtures, and simply walk out of the building with no human interaction".

Design Team Can Interpretation

- **User Requirement 1:** "As an Amazon Prime member, I want to enter an 'Amazon Go Store,' scan my QR code with my authentications, pick up any items, and simply walk out without any human interactions."
- **User Requirements #2:** "Surveillance cameras within the retail store must meet the following specifications:"
  - "The sensors must have high definition resolution to ensure clear and detailed image and video footage."
  - "The camera sensors should be properly calibrated to maintain accurate color representation and image quality. This can be a physical and software-based requirement."
  - Integration with knowledge within the 3D store model, fusion sensor, and object detection capabilities to enhance surveillance and security monitoring."

From the example above, all of the requirements that a technical individual has interpreted must match the requirements that were presented by the non-technical individual before officially being placed in the product backlog.

2.1. Example of the System Design

During the presentation, we emphasized the technologies that make up the entire Computer Vision from requirements 2.

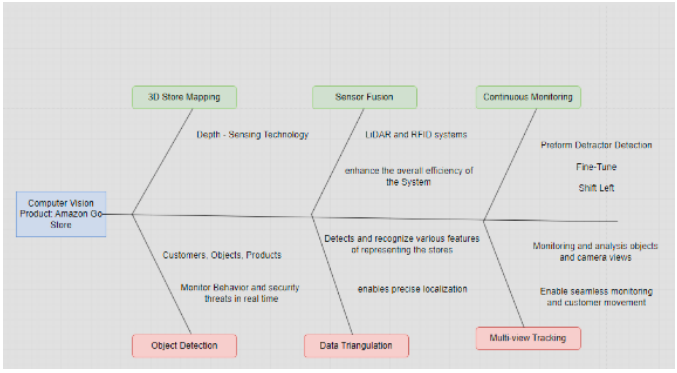

Figure 2. Based on requirements 2, there are 6 techniques to meet the requirement. We only covered 3D Store Mapping, Sensor Fusion, and Data Triangulation during the presentation. These, however, are the core components that collaborate systematically to ensure accurate procedures.

The key requirements of a successful interpretation is within the 3D Store Mapping. It is the process of obtaining a proper overview of the entire building's interior and exterior. This is treated as an environmental data variable presented by the cameras and/or LiDAR scanners. It creates a digital model of the store layout, including shelves, walls, and other objects pertinent to the retail industry. Sensor Fusion, as its name suggests, involves the integration of all various data. Similar to 5G, every sensor has its strengths and weaknesses, but fusing the information allows the system to compensate for any limitations within the network, resulting in a better comprehension of the surroundings. In Data Triangulation, utilizing information from multiple sources such as cameras can help determine the precise location of an object in a 3D space. It is equivalent to taking a photo of the same object from two different angles; it can help analyze the slight differences in perspective and pinpoint the exact location. Even though all three components rely on a combination of data, obtaining an image is still pretty complex. As a group, we did not have enough time to discuss the topic, but were planning on mentioning Convolutional Neural Network as a guidance to decision making.

3. Functional Analysis of Business Owner Requirements

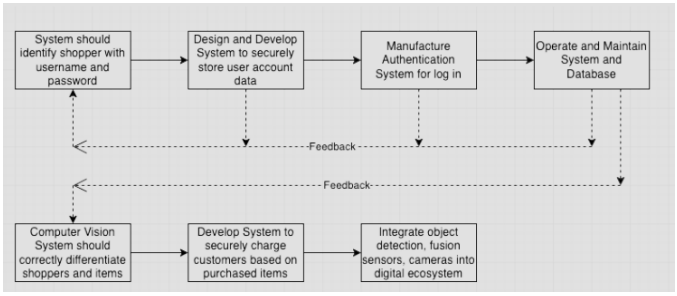

Figure 3. Ex. From Business Requirements

To ensure a requirement system works properly, we need to break it down into smaller features to ensure proper sequential functionality. The process of functionally analyzing a system entails disassembling the intended system into more of a manageable, and smaller group to ensure proper clarity. Taking into consideration the entire digital ecosystem of the Retail industry as well as our Requirements, proper analysis of functionality is crucial. In the end, it's about having proper documentation and user experience throughout the entire process.

Analyzing Requirement 1, users should be able to scan their QR Code to prove proper authentication. Once the system authorizes that individual, the customer can be granted access to enter the store and products. Once the user is granted access, the system securely stores the user's account data into the backend and links any activity performed by that particular user based on the information stored using CV. The user should have the ability to edit their profiles within the certified app, or link any default payment methods of their choice while they are in or away from the store. And the QR code will be frequently updated in real-time with the information the customer provides.

While in the store, as the user performs certain actions, the CV system simultaneously manages product inventory. It must determine the decision-making process if the acquired feature is potentially being purchased or owned. If the feature is no longer detected by the sensors, the entire Machine Learning (ML) technology indicates that the customer has "Just Walked Out" and will process payment. After proper payment has been made, a digital receipt is sent to the customer, and a report of the updated inventory will be sent to the retail's engineering team if any problems arise.

4. Trade Off in Computer Vision System

Towards the end of our presentation, we propose two alternatives in case the entire camera system's sensors are not functioning properly. The first alternative is using RGB cameras, and the second is using depth cameras with load sensors.

For the RGB camera, a customer could wave their palm above a receiver, which will take two photographic images. One image would be of the surface of the shopper's palm, and the other would be of the subsurface vein and artery pattern of the palm captured through infrared technology. These images would then be passed through a pre-trained neural network, which would compare them with existing feature vectors in the database. If there is a match within the store's database, the customer would be granted access. This proposal incorporates concepts from both CV technology and thermal camera sensors. The benefits behind this approach are twofold. First, RGB cameras could potentially achieve high accuracy in identifying shoppers and items. Additionally, palm scanning can provide unique user identification and allow for more personalized shopping recommendations. This can also help with inventory management, as it can provide insights into trending items that customers might buy within the store.

For the second proposal, we proposed depth cameras and load sensors. These cameras do not capture colors (unlike standard RGB cameras), but do construct a 3D image of the scene. Similar to 3D store mapping, the depth cameras identify the object's shape, location, and can help make decisions based on what is being taken. The load sensors are the weighted scales that are embedded in the floor or shelves. This can help determine the weight changes, and can help indicate an item being taken from the shelf. Both of these technologies work together to entire proper accuracy. The benefits to this is more cost-effective than standard RGB cameras, and it ensures privacy since it does not capture detailed shopper images. Another advantage behind this is it can help in a CV related neural network process, CNN. On the other hand, the overall accuracy aims to challenge similar items or handle situations with shopper's actions. They do not provide any additional data beyond the item they take, and are unable to provide any form of transparency in the promotion and personalized shopper's experience.

5. Why Did The Amazon Go Store Flopped?

Even though our previous presentations revolved around a modern system design approach, we have drawn significant inspiration from the Amazon Go Store documentation. As mentioned earlier, Amazon introduced their own version of incorporating CV technology into the retail industry, introducing the concept of "Just Walk Out." Announced in 2016, it aimed to promise customers a shopping experience without human interactions. Despite revolutionizing the retail industry significantly, in April 2024, they announced it as their biggest flop. The CV technology that was introduced never really worked as intended, relying instead on 1000 overseas workers to monitor customers as they shopped for merchandise. This suggests that CV technology is not yet mature and still has a long way to go before becoming a reliable source of information. Human intervention is still necessary for proper decision-making, and it will likely be decades before we see widespread adoption.

5.1. As a group, we indicated 2 proposal theories of why the Amazon Go Store failed:

- The concept of "Over-Design and Over-Engineering":
  - The design team has taken the requirements from the stakeholders and wanted to implement the sensors with all powerful tools that make it powerful. This is a bottleneck approach that costs more time, money, and staff, which leads to the entire system to be over-engineered and under-performed.
- The "Being First Approach":
  - Amazon wanted to be the first retail company to integrate this powerful ML technology. Implementation over the system requirements takes time to fully capture, and they wanted to eliminate human intervention prematurely. As of 2016, no company

has successfully integrated CV technology into their services, and Amazon wanted to be the first one. This leads to higher expectations and under-performing to meet certain requirements.

6. Conclusion

Overall, our presentation was very informative, taking into consideration the system requirements process, introducing potential trade-offs, and addressing current issues. However, what we need to improve on is properly mentioning other key factors that can work with a sensor. Rather than relying solely on RGB, LiDar, and Sensor Fusion, we did not mention RFID technology, which could make the entire system more simple and precise. Additionally, we overlooked the core aspects of machine vision, such as Convolutional Neural Networks, which aim to process captured images, simplify them, and feed them to a prediction model. For our next presentation, we plan to delve deeper into the more sophisticated aspects of the technology, mention other simpler key indicators of CV technology, and work more closely on potential fixes to certain issues that may arise.

References

[1] Haralick, R. M. (1992). Performance characterization in computer vision. In BMVC92: Proceedings of the British Machine Vision Conference, organised by the British Machine Vision Association 22–24 September 1992 Leeds (pp. 1-8). Springer London.

[2] Davies, E. R. (2017). Computer vision: principles, algorithms, applications, learning. Academic Press.

[3] Wankhede, K., Wukkadada, B., Nadar, V. (2018, July). Just walk-out technology and its challenges: A case of Amazon Go. In 2018 International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 254-257). IEEE.

[4] Polacco, A., Backes, K. (2018). The amazon go concept: Implications, applications, and sustainability. Journal of Business and Management, 24(1), 79-92.