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.**

I. Introduction

For our second presentation, we summarized what we covered in the first presentation. We introduced new content revolving around the overall DevOps approach to software development versus the Computer Vision pipeline cycle, along with a real-life product that utilizes computer vision, the Amazon Go store. As stated in our first presentation, Computer Vision is a branch of Artificial Intelligence (AI) and Machine Learning (ML) technologies that process data through observation. Similar to how humans process data through perception, Computer Vision (CV) aims to mimic these traits to help perceive, identify, and understand objects through patterns and other forms of motion. It is worth mentioning, although CV is still entering its infant 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 advantages include, but are not limited to, Automation, Improved Insights, Enhanced User Experience, Safety, and Innovation, which make the resource itself stand out for any AI and ML traits. However, some disadvantages include how it relies on data, overall complexity, privacy concerns, and security challenges. Despite these differences, CV will continue to be the optimal solution for many years to come, as it will become the industry standard to integrate through their already existing technology.

II. DevOps Vs. Computer Vision Pipeline

During the presentation, we emphasized the DevOps approach to software development and compared and contrasted it to the standard modern CV pipeline. Both provide a generic sequence skeleton of development and many similar techniques including automation. In DevOps approach to software development, it consists of a sequence of Plan, Build, Test, Release, Deployment, Operation, and Monitoring. And in the modern CV pipeline sequence, it consists of Requirement Analysis, Data Collection and Preprocessing, Feature Extraction, Model Selection Training, Optimization Fine-Tuning, Deployment, Monitoring Maintenance. Although these approaches are different in the development world, they do help develop products at a higher speed and apply adaptation, inspection, and transparency. Unlike the traditional Waterfall or Vee model approach to software development, DevOps primarily focuses on the Agile Scrum pattern that applies faster deployment to already scalable application development. This allows many scopes of the application to be recorded and fully engaged within the agile team and allows automation for quality assurance. Similar to CV, a DevOps product is never completed, and will need to undergo proper continuous integration, heavy monitoring, and frequent updates to ensure the platform and other components are up to industry standards. Of course, some of the obvious differences between the two approaches are the tooling and technologies that are used. DevOps harvested tools in relation to Docker, Jenkins, and Jira, while CV utilizes tools within the Open Source Libraries such as PyTorch and OpenCV for their settings.

III. Data Collection and Feature Extractions

When indicating how the sequence pipeline is covered, it is important to realize how all the steps play a pivotal role in development. For our first presentation, we covered the first step of the pipeline: Requirement Analysis. This step involves gathering all the necessary product backlog and business requirements to build the application and define the main objective of the project itself, providing transparency and adaptability while the product is in development. For the second presentation, we covered the next two phases of the diagrams: Data Collection and Feature Extraction. This is a 5-step process consisting of 5 stages: Image Acquisition, Image Processing, Feature Extraction, Pattern Recognition, and Decision Making. In the Image Acquisition phase, the image or video is captured and stored. The captured image is then ready to be processed in the processing phase, where the images are refined, and important features are extracted. After this process, the system evaluates the most important parts of the captured images, including size, shape, color, and texture. Subsequently, the images undergo Pattern Recognition, where they undergo either statistical analysis or Machine Learning algorithms such as Clustering, Bayes, and many other Neural Network ML algorithms. After completing all these steps, the computer vision system proceeds to a proper decision-making process, utilizing an IF-then solution. Throughout this segment of the presentation, we compared how we, humans, process images to understand and make rational decisions based on what we witnessed. We also introduced a real-world product and how it processes the images as well, the Amazon Go Store's "Just Walk Out Policy" functions.

IV. Amazon Go Store "Walk Out Procedure"

The Amazon Go Store is revolutionizing the retail industry with its "Just Walk Out" technology. It is where customers can enter a store and walk out with the merchandise of their choice without any human interaction. We explored its essential components and their roles in creating a seamless shopping experience. It is a high-level architecture platform that is powered by pure AI, ML, and CV technologies so it develops its decision-making process and allows an amazing user experience while in stores. What makes the "Just-Walk-Out technology" so advantageous is its high-level architecture, which applies 3 major components. The first component is Person Detection, where the system continuously identifies and tracks the presence and movement of individuals who walk into the store. The second is Object Recognition, where the technology is capable of differentiating between the items available for purchase, allowing for accurate tracking of inventory and potential purchases. The third component is Pose Estimation, whereby the system analyzes the first two components to make rational decisions and conclude the outcome of the entire experience. These three components collaborate to make the Amazon Go Store the leading-edge technology that will one day revolutionize what we see today. It is apparent that the store itself harvests precisely calibrated camera positions to allow accurate movements and actions within the store to be measured. The system is trained to create a 3D map of the entire interior and exterior of the store and to provide dynamic corresponding views within the area, which is powerful enough to accurately determine actions.

V. Sensor Fusion and Activity Analysis of Amazon Go Store

Despite the camera's sophistication, it needs to aggregate significantly across different sensors of the system. As stated before, the multi-cameras orchestration is trained and equipped to have a complete view of the store environment, seamlessly tracking customers across different camera views and then ensuring it never loses sight of them or their shopping interactions. To further illustrate this, to communicate with the overall data process, it is necessary to undergo Data Triangulation to decide on the overall actions. When analyzing the activity actions of the CV, deep learning algorithm is then used to analyze the video to identify and track individual items and recognize actions.

VI. Conclusion

Overall, our presentation was highly informative and technical. We illustrated a real-world product that is poised to revolutionize the current standpoint for the retail industry. A CV system is complex and requires thorough training to become a reliable source of information. This training process may take years or even decades, depending on the companies and industry focus. Despite going overtime in our presentation, we effectively conveyed the concept and real-world applications of the pipeline approach. However, we acknowledge the need for a deeper understanding of deep learning algorithms, such as Artificial Neural Networks, Data Triangulation, and others. It is crucial to have a great understanding on the topic, but also to have a grasp of the algorithms that empower these cutting-edge technologies. In our next presentation, we will dive deeper into the algorithms used and explore data-processing algorithms that contribute to transforming these concepts into reality.

References

Performance Characterization in Computer Vision: Reference: https://link.springer.com/chapter/10.1007/978-1-4471-3201-11

Computer Vision By E.R Davies: E.R. Davies, "Computer Vision", Publisher: Academic Press, Year: 2005.

How the Amazon Go Store's AI Works By Ryan Gross: Reference: https://towardsdatascience.com/how-the-amazon-go-store-works-a-deep-dive-3fde9d9939e9