e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

AUTOMATED CASHIER STORES: AMAZON GO TECHNOLOGY

Rohit Kumar*1, Salwa Shaikh*2

*1M.Tech Student, SMST Dept., Indian Institute of Technology,
Kharagpur, India.

*2B.Tech Student, ECE Dept., MJP Rohilkhand University, Bareilly, India.

ABSTRACT

Amazon Go is a chain of convenience stores in the United States and the United Kingdom, operated by the online retailer Amazon. Now a day no one likes to stand in a checkout queue while shopping. So to automate this process Amazon came with automated Go stores, incorporating computer vision, deep learning algorithms, and sensor fusion for the purchase, checkout, and payment steps associated with the retail transaction. As of March 2021, there are 32 stores located (both established and announced) in the United States and 15 in the United Kingdom. In this paper, we will go through each aspect of this technology.

Keywords: Computer Vision, Deep Learning, Amazon Go, Sensor, AWS, RL.

I. INTRODUCTION

Amazon Go stores were conceptualized and tested by a team of Amazon executives, who constructed a 15,000-square-foot mock supermarket in a rented warehouse in Seattle. A cashier-less store (also called a till-less store, checkout-free store, or Just walk out store) is a store that allows customers to shop for their products and leave without having to wait in line and pay at checkout. Stores use a combination of technologies such as cameras, sensors, computer vision techniques, and deep learning to register which products someone picks up or puts back. In April 2022, Minute Maid Park in Houston, Texas became the first Major League Baseball stadium to incorporate cashier-less stores, installing Amazon's technology at two of it's concession stands.

II. AMAZON GO STORES NOW &THEN

In 2016, Amazon announced Amazon Go, a brick-and-mortar store featuring various technologies to eliminate the checkout. People require the Amazon app (formerly a specific Amazon Go app) to enter the store, grab whatever they need, and exit to be billed afterward through their account.

The first Amazon Go store opened in 2018. In 2019, Sainsbury's opened the first cashier-less store in the United Kingdom. However, it was closed a few months later. In July 2020, a Finnish company called Korttelikauppa opened a cashier-less store in Kuninkaantammi, Helsinki. In the Spring and Summer of 2021, Korttelikauppa opened 6 stores in the following cities:

- Helsinki
- Espoo
- Vantaa

In April 2022, Minute Maid Park in Houston, Texas, became the first Major League Baseball stadium to incorporate cashier-less stores. In 2022, Aldi Nord will open the first cashier-less store in the Netherlands, in the city of Utrecht. Amazon plans to introduce its cashier-less technology in two Whole Foods stores (in Washington, DC, and Sherman Oaks, California, in 2022).

III. ARCHITECTURE OF PLATFORM

At Amazon's RE: MARS conference, they did the first tech deep dive on the vast array of innovations required to provide the "Just Walk Out" experience. I will walk you through what they shared so we can appreciate their astounding achievement.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

1. Blueprint

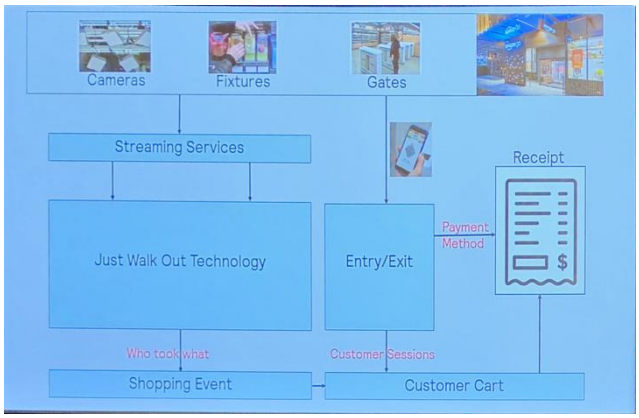

Fig.1. The high-level architecture of the platform

2. Computer Vision Core: "Just Walk Out" Technology

The heart of the Go Store is the Computer Vision-based Machine Learning which is used to seamlessly track and estimate the intention of everyone in the store. Amazon went into a surprising level of detail on their implementation of this technology. While they didn't go so far as to show the exact Neural Architectures for their models, they did show the specific problems those individual models solve and how they are combined to build the further solution.

Dr. Gerard Medioni — chief scientist and former USC Computer Vision professor, presented this session. Not only is he an excellent speaker but also a brilliant researcher. He managed to cover intense topics in just a few seconds to answer the seminal question behind the go store: **Who took what?**

Fig. 2. Dr. Gerard Medioni — chief scientist and former USC Computer Vision professor.

Fig.3. Amazon's re: MARS conference June 4, 2019 - June 7, 2019

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

3. Core Problems

There were 6 core problems that needed to be solved to provide the experience.

3.1 Sensor Fusion: Aggregate signals across different sensors (or cameras because this was solved using nothing but computer vision).

3.2 Calibration: Have each camera know its location in the store very accurately.

3.3 Person detection: Continuously identify and track each person in the store.

3.4 Object Recognition: To distinguish the different items being sold.

3.5 Pose estimation: To detect what exactly each person near a shelf is doing with their arms.

3.6 Activity Analysis: To determine whether a person has picked up vs. returned an item.

IV. COMPONENTS REQUIRED

The detailed architecture for the Just Walk Out technology includes the following components:

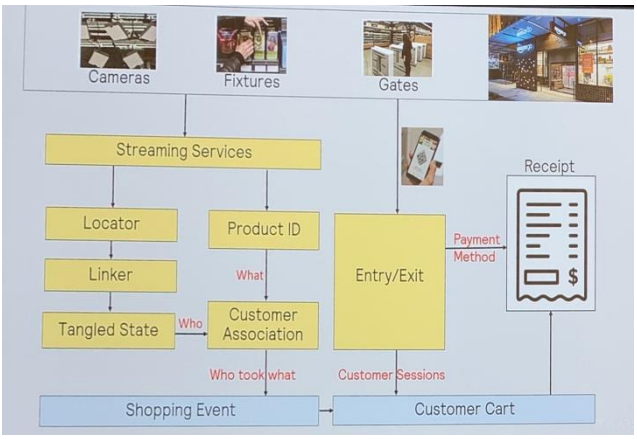

Fig.4. Just Walk Out Architecture

1. Person Identification

Following components are important for it:

1.1 Locator: The question "Who took What?" could not be solved as a series of independent picks. Amazon had to track each person the whole time they were in the store, from the moment they walked in until they had left. Some of the problems that had to be solved by the Locator component were:

a. The Occlusion, where a person is blocked from view by something in the store.

b. The Tangled State, where people are very close to each other.

1.2 Linker: The next task was to ensure the labels were preserved across frames in the video, moving from locating to tracking the customers in the store. The problems experienced in this phase were:

Disambiguating Tangled States: When people get very close together, this lowers the confidence of who's who. The go store technology handles this by marking these customers as low confidence, so they get scheduled to be re-identified over time.

There is a follow-up phase for distinguishing Amazon Associates, who likely perform different behavior than customers (for example, they would likely put items on shelves rather than taking them off).

2. Item identification

2.1 Product ID detection: The prominent question to answer here is: which specific items are off the shelf and in someone's hand. Some of the problems faced and their solutions in phase were:

a. Items that are very similar, like 2 different flavors of the same brand of drink, were distinguished using residual neural networks that do refined product recognition (I'm guessing across multiple frames) after the CNN layer identifies the item class.

b. Lighting and deformation changes the items, which was solved using a lot of training set data generation for these specific challenges.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

3. Customer association

Probably the most challenging problem is combining all of the information from the above steps to finally answer the "Who took what?" question.

4. Pose Estimation

The Location tracking Go store cameras look top-down, not from an isometric view, so they need to trace a path through the pixels representing the arm between the items and a customer. A simple top-down model did not work well enough to solve this problem, so the team set out to build a stick-figure-like model of the customer from the video.

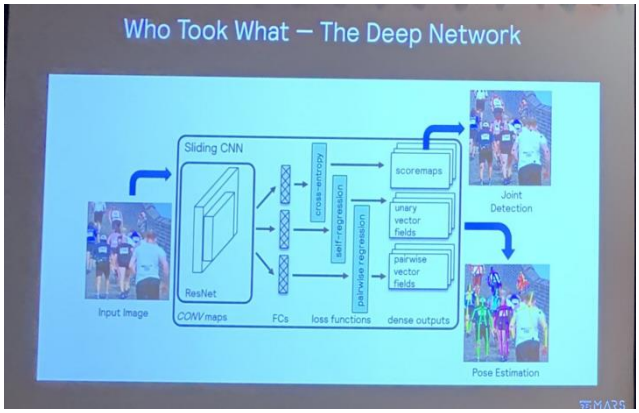

Fig.5. Who Took What

A novel new Deep Learning model was needed to build an articulated model of each customer from the video. It uses a CNN with a cross entropy loss function to build the joint detection point cloud, self regression for vector generation, and pair-wise regression to group the vectors together. This model is extremely interesting in and of itself. They showed that it can be used on any video clips to aids solving many other problems that rely on pose estimation.

5. Action determination

To avoid charging customers for items they didn't take, the system must accurately account for a world where the customer can put items back on the shelf.

One of the problems in this area can be seen in the picture above. The obvious answer to the question is that an item was taken, but this is incorrect. Instead, a customer put an item back and pushed the remaining ones further back on the shelf. To solve for this, the system needs to count all the items on the shelf rather than using a simple assumption based on space.

6. The long tail

There are massive number of poses people can be in when picking an object off the shelf, especially when you consider multiple customers nearby. There isn't enough labeled data to train a model for each of these. Even with human labeling, it wouldn't be possible to scale the training dataset (in both money and time).

To solve this, the team took on the ambitious project to generate synthetic activity data using simulators. Within these simulators, they needed to create virtual customers (including variations in clothing, hair, build, height, etc.)

(Including variations in clothing, hair, build, height etc.), cameras, lighting & shadows, and simulate the same camera hardware limitations.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

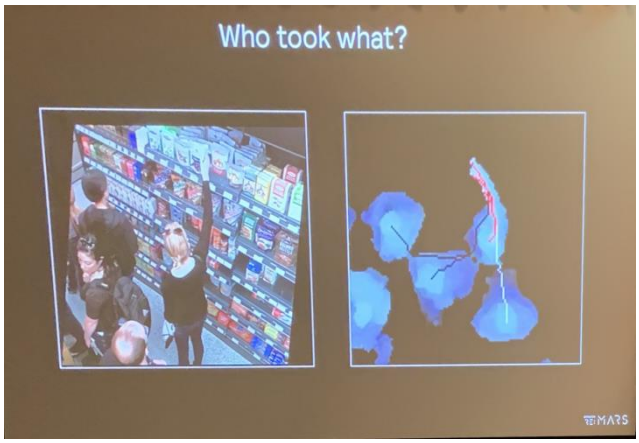

Fig.6. Simulated data is used to solve for difficult situations

V. USED HARDWARE & SOFTWARE

Table 1: Font Hardware and Software requirement for implementation

	Appearance (in Time New Roman or Times)		
	Person	**Shelf**	**Entrance/Exit**
Software	Amazon Go app	Cameras, pressure sensors, infrared sensors, scales, volume displacement sensors, light curtains, etc.	Customize system
Hardware	Smart Device (Smart Phone)	Management System(WMS)	2D Barcode

VI. ENTRY & EXIT DETECTION

The next challenge is detecting when people enter and exit the store to create the shopping session. This system has the following components:

a. Mobile App to scan QR when you show up at the store. They spent a lot of time doing UX testing on this (scan with phone up or down, how to handle groups, etc.)

b. Association System associates your likeness in the video to your account based on position in the store entrance when you scan the QR code.

While implementing the system, the team had to solve a few additional scenarios. First, people might scan multiple times, so they had to delete any session with no items on a second scan. A more difficult problem is customers (especially families) want to shop as a group but only have a person to pay. To enable this, the head/payer scans the same code for each person as they enter the store. This creates a session that links all of the people in the group to the same account. From there, the people in the group can leave the store whenever they choose. By moving the session up to the group level, and treating individual shoppers as a "group of one", the team was able to overcome this challenge and let individuals https://meet.google.com/fkw-uuqt-yjy https://meet.google.com/fkw-uuqt-yjyenter or exit the group at any time.

VII. CART, PAYMENT, AND RECEIPTS

These all are basically the same as what you have on Amazon.com, so there wasn't much innovation to discuss here. They covered these sections for all of 10 seconds. Of course they ended the talk in classic Amazon style by

saying "It's Day 1 at Amazon Go" and hinting that there are major improvements on the horizon. I left the talk feeling both humbled at how far behind I am compared to these types of achievements and extremely excited for the golden age of AI to come.

VIII. CONCLUSION

We can say that this type of technology explores the evolving relationship between human and machines, examining the ways that robots, artificial intelligence and automation impacting our lives.

We could see Amazon Go storefront play an instrumental role in Just Walk Out technology. This technology could solve the perennial problem of retail theft, because even if you walk out stealing an item out of a store, you will still be charged.

ACKNOWLEDGMENT

This paper is detailed version of RE: MARS conference of Amazon held in June, 2019. Also we want to thank "Emerging Tech & Data Leader" Dr. Ryan Gross for publishing such tech articles.

IX. REFERENCES

[1] Amazon Go Editorial Staff (2017). Amazon Go, Retrieved from: https://www.amazon.com/b?node=16008589011

[2] Vinod Suryawanshi, Aditya Gosavi, Unmani Joshi, Sagar Suri," Automatic Toll Collection System Using QR Code", InternationalJournal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 3 March 2017, Page No. 20614-20616

[3] Uke Nilesh J., and Ravindra C. Thool. "Moving Vehicle Detection for Measuring Traffic Count Using OpenCV." Journal of Automation and Control Engineering Vol 1.4 (2013).

[4] Amazon.com. (Accessed on 2017, July 12) Retrieved from: https://www.amazon.com/b?node=16008589011

[5] Brian Roemmele. (2016, December 6). "Amazon Go Store: ABL- AAlways B- Be L- Learning" [Blog post]. Retrieved from https://www.quora.com/How-is-deep-learning-implemented-in-AmazonGo

[6] C.M. Roberts, "Radio Frequency Identification (RFID)," Computers & Security, Elsevier, 2006. James J. Barbelllo. Handbook for parallel port design. Prompt Publications, 1999.

[7] Nick Wingfield, (2016, December 5). "Amazon Moves to Cut Checkout Line, Promoting a Grab-and-Go Experience" Retrieved from The New York Times, https://www.nytimes.com/2016/12/05/technology/amazonmoves-to-cut-checkout-line-promoting-a-grab-and-go-experience.html

[8] Rabia Jafri, Syed Abid Ali, and Hamid R. Arabnia," Computer Visionbased Object Recognition for the Visually Impaired Using Visual Tags", Retrived from: http://weblidi.info.unlp.edu.ar/WorldComp2013-Mirror/p2013/IPC2637.pdf

[9] Anderson, G. (2014, August 14). WalMart's Scan & Go is a no-go. RetailWire. Retrieved from http://www.retailwire.com/discussion/walmarts-scan-and-go-is-a-no-go/

[10] Barnard, S., Calderara, S., Pistocchi, S., Cucchiara, R., Podaliri-Vulpiani, M., Messori, S., and Ferri, N. (2017). Amazon Go. Retrieved from https://www.amazon.com/b?node=16008589011.

[11] Bowman, J. (2017, November 3). Whatever happened to Amazon Go? Retrieved from https://www.fool.com/investing/2017/11/03/whatever-happened-to-amazongo.aspx

[12] Del Rey, J. (2017, December 20). Walmart is developing a shopping service with no cashiers. Recode. Retrieved from: https://www.recode.net/2017/12/20/16693406/walmart-personal-styling-jet-black-amazon-go-prime-no-checkout-store

[13] Dipert, B. (2017, January 23). Visual depth sensors: Multiple applications and options. Vision Systems Design. Retrieved February 25, 2017, from:

http://www.visionsystems.com/articles/2017/01/visual-depth-sensors-multiple-applications-andoptions.html

[14] Hofbauer, R. (2017, November 22). Is Amazon Go nearing public debut? Progressive Grocer. Retrieved from https://progressivegrocer.com/amazon-go-nearingpublic-debut

[15] Lamm, G. (2017, March 27). Amazon Go store could be shut down. Digital Editor. Puget Sound Business Journal. Retrieved from:
https://www.bizjournals.com/seattle/news/2017/03/27/amazon-go-difficulties-seattle-betatest.html

[16] Rash, W. (2016, December 8). Amazon Go won't kill many jobs, but it may prove a boon to crooks. eWeek. Retrieved February 25, 2017, from http://eds.a.ebscohost.com.