Computer Vision for Pattern Recognition

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

***Abstract*—This paper outlines a systematic design approach for contemporary Computer Vision technology. It will indicate potential trade-offs and potential remedies for product failures.**

I. Introduction

For our project, we aim to develop a more refined Computer Vision (CV) technology by identifying, analyzing, and proposing new ideas. CV is a branch of machine learning that allows machines to understand and process visual data. Unlike standard Artificial Intelligence (AI) products, CV handles and processes any complex data through observation. It has the potential to improve existing AI products by providing a more accurate and scalable form of visual representation, leading to improved image processing capabilities. Similar to how humans learn from observing the world around them, CV mimics this process, enabling AI to make more rational decisions outcomes. The benefits behind a successfully trained CV system includes cost efficiency, decrease in staff, and overall reliability and accuracy when analyzing through large amounts of data. In other words, it helps identify specific patterns and immediate proper decision-making responses. However, a common challenge with Machine Vision (MV) products is that they can become overly complex, making them difficult to maintain and develop further. As AI becomes increasingly available, it's crucial to establish sophisticated CV requirements to ensure long-term success. For our project, our team will focus on creating a sophisticated yet manageable CV technology system design by proposing and implementing innovative approaches.

II. History Of Computer Vision Technology

Computer vision technology has been around for many decades. Dating back to the 1950s and 1960s, engineers were fascinated by the challenge of enabling early computers to process information from the physical world through means of observation. Of course, as computational power and technological advancements continue to grow, so do the capabilities of computer vision technology. It has proportionately become increasingly refined and prevalent as time progresses; making it a high-powered impact technology. As time continues to proceed, so do the visionary technologies integrating optical factors. An example of an industry that harvests its capability is in security monitoring; where cameras and other sensors leverage computer vision to perform tasks like object detection and anomaly recognition. To further extend this point, technologies ranging from autonomous vehicles to Augmented Reality gravitates to integrate CV technology to their already dynamic System Design. Lets understand the types of domains that utilizes CV technologies

III. Understanding the Domains of a CV technology

The initial phase before refining a technology is identifying its own domains. A CV technology has the potential to perform complex-data processing and decision making behaviors through what is being captured. As stated earlier, almost any industry can adapt to its generalized capabilities. Almost every industry has at least one AI product integrated into its system, but it's almost incomplete. What we mean by this is most of these AI products have very little existing vision technology integrated into the system. Let's explore the various industries that can utilize CV technology.

A. Here are some industries that would benefit from having a defined CV technology

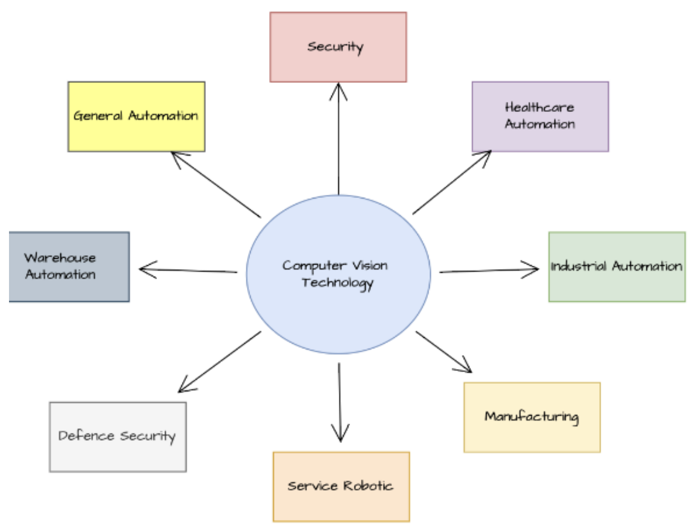


Fig. 1. Various Domains of where Computer vision can be applied

- **Security**: A MV should be trained to ensure any threats to the public are captured in real-time and reported. E.G. Airports, Schools, Buildings, and more.
- **Warehouse Automation**: A CV has the potential to be used to detect and help in inventory management within a warehouse.
- **Service Robotics**: A MV machine should be trained to help anyone who has a vision disability. To help guide and assist with its pre-trained observation field.

It is apparent that Computer Vision technologies can have an impact. As AI becomes more readily available and integrated to an existing industry; CV will become a factor that will be proportionality significant. Implementing a proper MV mechanism takes time and training, but if executed correctly; the outcome will be a game-changer in any industry.

B. *Modern CV Pipeline Sequence*

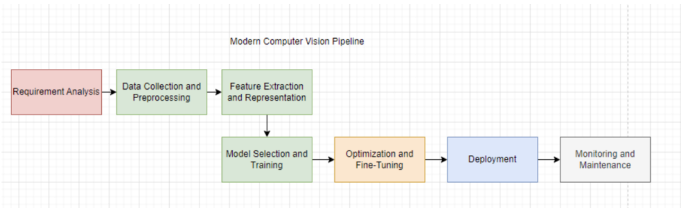


Fig. 2. An example of a Sequential CV Pipeline

- **Requirement Analysis**: Define the problem a CV system has to solve.
- **Data Collection and Prepossessing**: Gathering a dataset circulating around the relevant topic. Used for training and testing CV models using certain key metrics and physical representation to detect the target domains.
- **Feature Extraction and Representation**: Identify meaningful patterns, textures, shapes, and other physical attributes from the input data. More data analysis because it includes techniques such as edge detection and feature decipher extraction of the data.
- **Model Selection and Training**: Choosing the appropriate CV model architecture that will help assist an optimal performance on the targeted task. Convolutions Neural Network (CNNs), You Only Look Once (YOLO), Recurrent Neural Networks.
- **Optimization and Fine-Tuning**: Evaluate, validate, and make adjustments based on the general performance.
- **Deployment**: Release the entire application.
- **Monitoring and Maintenance**: Continue to monitor and maintain the system to ensure its optimal performance stays consistent overtime.

In the modern CV pipeline sequence, it consists of Requirement Analysis, Data Collection and Prepossessing, Feature Extraction, Model Selection Training, Optimization Fine-Tuning, Deployment, Monitoring Maintenance. Each of these phases are built from the initial phase of requirement analysis. In the requirement analysis phase, we must collect all objectives and business perspectives so we can draw out the root outcome of the entire system. For example, in the healthcare industry, if we want a CV product to detect any abnormal psychological behavior an individual is performing, we must highlight that within the business requirement. This is to help ensure we build the entire system requirements with respect to the original intentions. To properly address many of the components of a MV system, we must understand what makes a concrete Requirement Analysis, and will draw an optimal solution later on this page. Despite all the sequence providing insights on how to properly train the entire CV, it always has the potential to be wrongfully executed. As stated before, every sequence is traced back to the initial requirement stage, where many of the functional analysis and non-functional requirements are defined. It is apparent understanding the requirements of the System Requirements is the essential key to any system design process. Let's identify some of the issues with a Modern CV Product if the requirements are not fully understood.

C. *Identifying Issues with Modern CV Pipeline*

Not all existing CV products are fully trained, and many of which still need human interventions to ensure proper accuracy. Many other problems that arises, but are not limited to:

- **Issue #1**: Lack of inspection, transparency, adaptation of what CV technology.
- **Issue #2**: Misinterpretations of the Requirements highlighted for the MV technology.
- **Issue #3**: Over-Design documents which leads to high-level bottlenecks of the system design.

Some of these issues are addressed, but become a long-term problem when being able to maintain the entire system. Despite certain photographic sensors already incorporated CV technology, most of which are incomplete and incompatible. The goal of a CV product is to mimic and process any captured information similar to human traits.

IV. Example of a CV's Operations - The Amazon Go Store's "Just Walk Out" policy:

1) **Entry and Identification**: Initially, Amazon Go stores employed a QR code system for customer entry. Customers would access the Amazon Go app on their smartphones and scan a unique QR code generated within the app. This code linked their Amazon account to the shopping experience, ensuring accurate billing upon exiting.

   In 2020, Amazon introduced a more streamlined entry process utilizing biometric palm scanning technology. Customers simply wave their palms over a designated palm scanner. This scanner captures an image of their palm, focusing on the unique vein patterns. The captured image is then compared against stored user data in a secure database using neural networks. Upon successful verification, access to the store is granted.

2) **Tracking and Monitoring Customers' Movements.**: Upon entering the store, a network of strategically placed cameras tracks customer movement and items they picked up or held. These cameras utilize advanced computer vision software to identify products and monitor customer behavior in real-time. Additionally, each item in the store is equipped with RFID tags,

streamlining the tracking process.

Before 2022, the Amazon Go store combined a network of depth cameras with load sensors to identify customers' movements. Depth cameras don't capture color like RGB cameras but build a 3D image of the scene. The images captured from these cameras are sent to a central processing unit and quickly and accurately identify different customers in the store and objects being picked up or held. Depth camera identifies the object's shape and location. Load Sensors embedded in the shelves confirms a weight change, indicating an item being taken.By combining depth cameras and load sensors, the system gains a more complete picture.

In 2022, the Amazon Go store updated its camera system to RGB cameras. It operates more accurately and faster than common depth camera systems. Once the customers enter the store, the RGB cameras apply the CNN layer to identify the item they picked up. For very similar items, like 2 different flavors of the same brand of drink, RGB cameras distinguish them by using residual neural networks that do refined product recognition. As customers pick up items from the shelves, the system identifies who take what as we described above and automatically adds them to their virtual shopping cart. The network of RGB cameras ensures accurate tracking of the items selected by the customer and determine which items are being taken or put back on the shelves.

3) **Completion of Purchase and Payment**. Once customers have finished purchasing, they simply exit the store. The "Just Walk Out" technology calculates the total expense of the items in their virtual shopping cart using the data collected during their shopping process. The payment is then automatically processed through their Amazon account, and a digital receipt is sent to the app.

A. Recent News

According to recent news, the Amazon Go store's computer vision (CV) technology fell short of expectations. The system relied on over 1,000 people overseas manually reviewing video footage of customers shopping, highlighting the limitations of current AI and machine learning for achieving complete autonomy. This incident suggests that a deeper understanding of user needs and system requirements is crucial for successful AI implementation. While over-engineering a product can be a concern, proper upfront planning regarding how the system should process data can eliminate many potential bottlenecks. In the case of Amazon Go, focusing on perfecting the technology overshadowed defining clear expectations for image processing. Learning from this, our team has identified three key requirements to improve such technologies.

B. Visual Diagram of Amazon Go Store

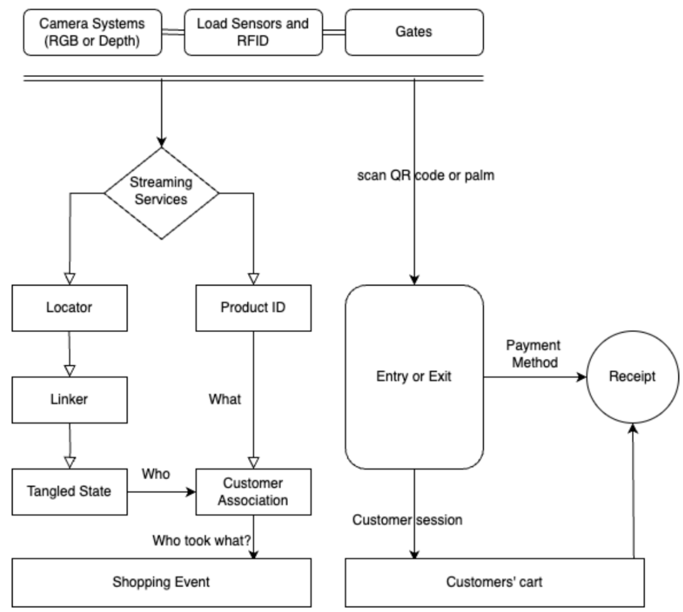


Fig. 3. The following diagram illustrates how the entire process of Amazon Go store operates

Business Owner and Stakeholder's Requirement #1: "I want a customer to enter a building, scan their Prime membership, pick up any fixtures, and Just Walk Out with no human interaction." The standard objective is to eliminate customer to employee interactions and allow camera sensors to make any decision-making results. The system will be heavily monitored using MV.

Business Owner and Stakeholder's Requirement #2: "The system must accurately track inventory in real-time. This includes detecting when items are removed from shelves (to add to customer baskets) and returned (to remove from customer baskets)."

Interpretation: Taking into the standard consideration Inventory Management, the system needs to indicate what items are taken from the shelves in real-time.

Business Owner and Stakeholder's Requirement #3: "The system has to define a 3D virtual zone around the product shelves area, so it can identify when stock is running low and trigger automatic alerts for staff to reload shelves."

Interpretation: "The system can identify 'n' number of items and its quantity within a particular zone, and has to trigger an automatic alert indicating Low-Stocks for the staff."

Note there will always be a finite number of requirements given to by the stakeholders and Business owner. With that being said, most of the requirements are the blueprints in any modern system definition. As indicated from the 3 requirements, the key factors to consider are Decision-Making Results, Inventory Management, and Low-Stack Management. With all of these illustrated, we can construct an optimal approach of the overall configuration and orchestration of the design. These requirements set the precedence and expectations that the product should deliver, eliminating any forms of ambiguity. It is up to the design team to decipher the requirements into a more technical perspective.

V. User Requirements

The user requirement builds upon the foundation laid by the business requirement. It segments and translates the business requirement into user-specific needs, prioritizing them based on their importance to the user. Understanding certain key requirements, constraints, and the application of the system is essential for designing efficient user requirements. Based on the requirements mentioned as a starting position, here are some user requirements that could be constructed:

- **User Requirement #1**:
  The CV system must accurately distinguish between a customer picking up an item and simply browsing it. This is crucial because the entire system relies on CV to automatically track items a customer obtains without having the traditional checkout process. The system should have access to every Stock-Keeping-Units, and the quantity present given that particular day. The entire system must be trained to optimize what is potentially being purchased versus what is being picked up. If the CV system cannot differentiate between picking up vs. browsing, then it can lead to:
  - Incorrect Billing: A customer will be charged more, and it will lower the customer's experience to shop at the store.
  - Inventory Inaccuracies: Items will show as missing, and it will greatly impact the Inventory Management requirements.

  Technologies that can be used include RFID, Sensor Fusion, Object Recognition, and LiDAR scanners.

- **User Requirement #2**: The QR Scanner must identify between active members and non-active members. It should be simple, user-friendly, and be able to determine if the person is authorized to enter the store.
  - Denial of Access: If membership is expired or not valid.
  - Security: The QR code scanning process should be secure. Users should be confident that their personal information and account details are securely safe, and there will be no data leakage. Only active users can change information within the account, including payment methods, contact details, and personal information. The Retail shop can only store this information if the user enters the store at a given time frame of the customer entering the store, and only reports back analytics data to determine trending, inventory, and Frequently Bought Items.
- **User Requirement #3**: The system shall be capable of capturing and analyzing complex, unstructured data to identify emerging opportunities and threats and perform decision-making outcomes similar to humans (Make-Decision Results). Be able to make autonomous decisions based on the image, similar to how humans make decisions based on means of observation.

- Inventory Management: Allows a visual representation of when a certain item is running low and needs to be restocked.

- Alerts team over any unusual activity that is taking place within the venue during business hours.

- Customers should not be worried about any specific human interactions within the store venue, and should be able to walk out of the store. The CV system should contain all the tools and technologies that can substitute into the role of employees.

Taking into account all three of these user requirements can help illustrate how the system should behave. Like many retail industries, we must take into consideration the fundamentals a traditional store has. We are not changing the retail environment, but using the CV to assist in eliminating staff and human interventions.

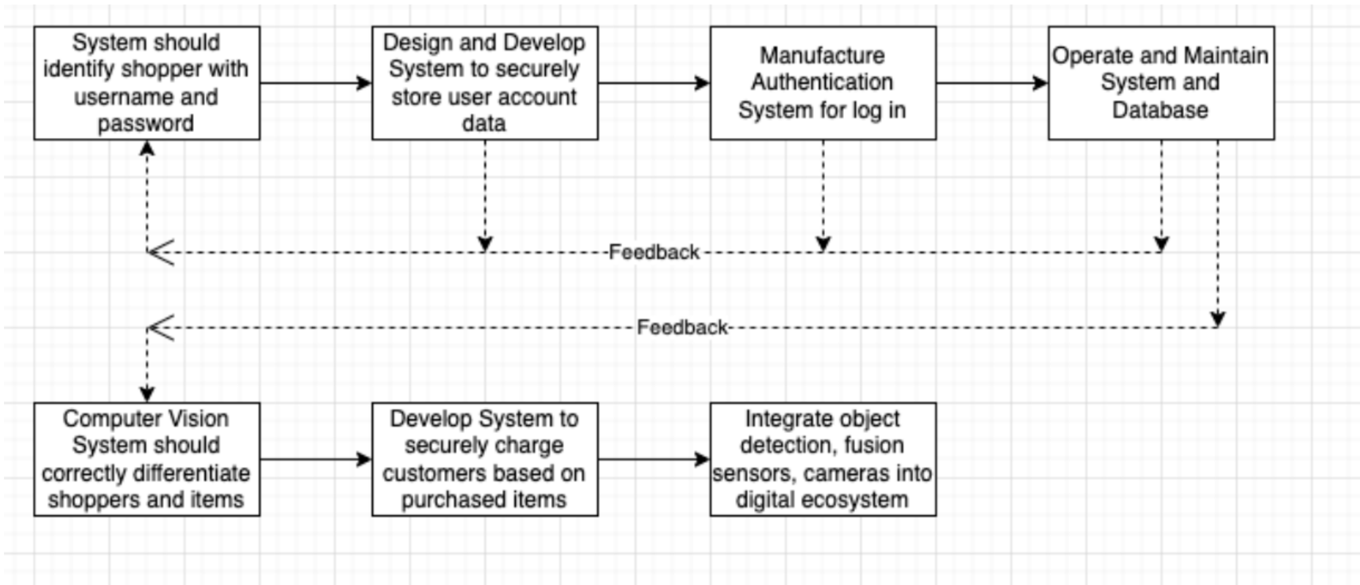
VI. System Requirements

- **Requirement #1**: The Computer Vision (CV) system should utilize advanced computer vision algorithms, e.g., object detection, pose estimation, and action recognition algorithms, to analyze customer movements and interactions with items on the shelves.
  - **Accuracy Goal**: The CV system must have an accuracy of at least 95
- **Requirement #2**: The QR scanner should be integrated with the Amazon Go app and capable of scanning QR codes generated for both active members and non-members.
  - **Acceptance Criteria**: The QR scanner must successfully authenticate users based on their membership status and authorization, allowing only authorized individuals to enter the store.
- **Requirement #3**: The system architecture should be designed to handle peak loads and accommodate future

growth through salable infrastructure and efficient resource allocation.

- **Availability Goal**: The system must maintain high availability (at least 99.9 percent) and responsiveness under varying loads and usage patterns.

A. Functional Requirement



The process of functionally analyzing a system entails disassembling the intended functionality into more manageable, smaller parts. The functional analysis may go like this in the event that a digital ecosystem requirement exists to distinguish purchased products and identify store shoppers.

Users ought to be able to register for accounts on the system by combining several forms of authentication, such two-factor authentication, biometrics, and passwords. After a person logs in, the system needs to recognize them and link their activities to the appropriate accounts. It should be possible for users to edit their profiles, adding and removing preferences, payment options, and personal data. Features like order tracking, subscription management, and purchase history viewing should all be available to users of the system.

The system will distinguish between products that have been bought and those that haven't. This might entail entering the acquired things in the system's database as "owned" or "purchased". One way to differentiate products may be visually, by using a distinct color or icon for purchased things, or by using filtering options. The system has to adjust the inventory when things are bought to account for variations in stock levels. It should be ensured via inventory management features that users are unable to purchase things that are out of supply. Item names, quantities, pricing, timestamps, and user data should all be included in transaction details. For a number of uses, including analytics, reporting, and customer service, this data is essential.

In addition, the system should integrate with payment gateways to facilitate secure transactions. Users should be able to choose from various payment methods (credit/debit cards, digital wallets, etc.). Integration should ensure compliance with relevant security standards (e.g., PCI DSS) and provide a seamless checkout experience.

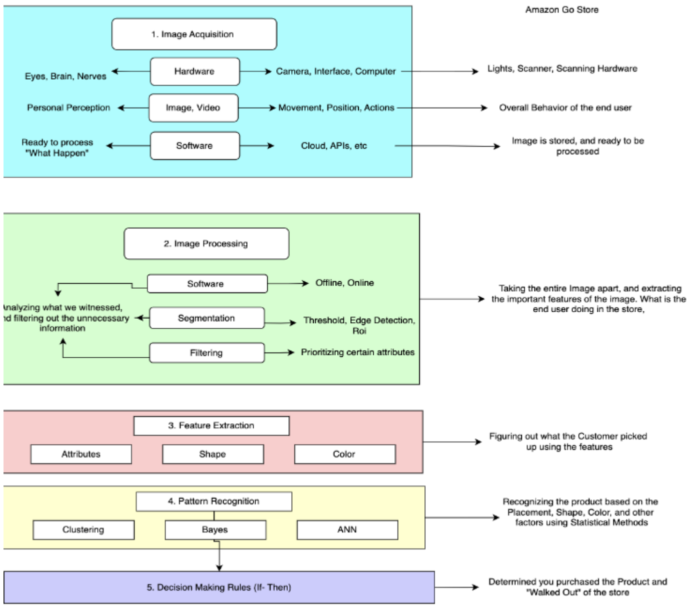


Fig. 4. Taking Characteristics form human and processing through

On the left indicates human attributes to image processing, on the right (internal from the diagram) represents technological substitution, and on the right (outside of the body) represents Amazon Go Store Image processing strategies.

VII. Test Performance Measurements

TPM Category	Description	Priority (%)	Weightage Example
High (70%)	Critical to customer experience and core functionalities		
Accuracy	Percentage of correctly identified customer purchases	40%	Ensuring trust and avoiding revenue loss
Speed of Checkout	Average time from product selection to checkout completion	20%	Frictionless and fast checkout experience
System Availability	Uptime percentage of the system	10%	Minimizing customer inconvenience and lost sales
Medium (20%)	Important for operational efficiency and security		
Inventory Management	Inventory shrinkage and stock outage rates	8%	Efficient stock management and customer satisfaction
Loss Prevention	Number of unidentified items leaving the store	6%	Maintaining profitability by preventing theft
Fraudulent Use	Monitoring suspicious purchase attempts or unauthorized access	6%	Protecting customer data and preventing misuse
Lower (10%)	Important for long-term performance and optimization		
Computer Vision Performance	Frame rates, resolution, and processing latency	4%	Smooth operation as long as accuracy isn't compromised
Shelf Weight Sensor Accuracy	Accuracy of weight sensors used to identify product selection	3%	Acceptable variations as long as overall accuracy remains high
System Scalability	Resource utilization and planning for future expansion	3%	Important for future growth but might not be immediate focus

Fig. 5. Taking Characteristics from human and processing through

To ensure the entire system performs accurately, we must take into consideration the traditional retail store's expectations. Similar to 5G's Key Performance Indicators, our system design approach will prioritize Accuracy at 70 percent to ensure trust, minimize inconvenience, and provide a fast checkout experience (which is highlighted in the business requirement). The reason for this 70 percent metric is that we are eliminating staff and adopting a more simplistic approach using CV technology. CV technology must be trained to ensure accuracy. 20 percent of the Technical Performance will align towards inventory management and general loss prevention, while 10 percent will measure the physical technology and smooth operations and versatility within the system itself.

VIII. Physical Analysis

Physical analysis of a computer vision automated grocery store like Amazon Go involves understanding the infrastructure, hardware, and technologies deployed to achieve its functionality.

Computer vision systems based on cameras are a major component of Amazon Go Store. Throughout the business, high-resolution cameras are positioned in strategic locations to record and examine client movements, interactions with merchandise, and ultimately, the things that they choose to buy. All of these cameras frequently include sophisticated features including the ability to recognize images and sense depth. It is apparent, Amazon Go store relies heavily on camera-based computer vision systems. High-resolution cameras are strategically placed throughout the store to capture and analyze customer movements, interactions with products, and ultimately what items they select for purchase. These cameras are often equipped with advanced features such as depth sensing and image recognition capabilities. The data recorded by the cameras is processed in real-time by complex computer vision algorithms operating in the background. These algorithms have been trained to correctly follow client behavior, detect objects, and recognize gestures. Convolutional neural networks (CNNs), one type of deep learning technology, are widely employed.

Behind the scenes, sophisticated computer vision algorithms process the data captured by the cameras in real-time. These algorithms are trained to detect objects, recognize gestures, and track customer behavior accurately. Deep learning techniques, particularly convolutional neural networks (CNNs), are commonly used for such tasks.

To process the massive quantity of visual data produced in real-time by many cameras, it requires an immense amount of processing and computing power. In order to reduce latency and guarantee quick decision-making, edge computing solutions—where data processing is done closer to the data source (i.e., within the store premises)—are frequently used.

Other sensors, such RFID tags on items or weight sensors on shelves, can be used in addition to cameras to supplement visual data and improve transaction processing accuracy. Sensor fusion approaches combine information from several sources to offer a more thorough picture of consumer behavior and product interactions.

Cloud infrastructure is essential for storing and analyzing massive amounts of historical data, building machine learning models, and handling urgent data processing requirements. Cloud services also facilitate remote monitoring, maintenance, and software updates for the entire network of automated stores.

Strong security measures, such as encryption protocols, access limits, and anonymization techniques, are put in place to safeguard consumer privacy and stop unwanted access to sensitive information since personal shopping data is sensitive data. For effective inventory management and invoicing, the computer vision system's output—which contains data on the things that customers have chosen—must be smoothly linked with the store's point-of-sale (POS) system. Custom software interfaces, or APIs, are created to make this integration possible.

The system is continuously optimized and refined by means of feedback loops powered by machine learning and data analytics. Customer behavior patterns, sales trends, and operational indicators provide insights that are utilized to optimize inventory stocking, boost system efficiency, and improve the entire shopping experience.

In summary, the physical analysis of Amazon Go and similar computer vision automated grocery stores encompasses a wide array of technologies including advanced sensor systems, computer vision algorithms, edge computing infrastructure, cloud services, and integration with existing retail systems. All of these components work together to create a seamless and efficient shopping experience while leveraging data-driven insights to drive business growth and customer satisfaction.

IX. Trade off

Amazon Go's "Just Walk Out" technology relies on a sophisticated blend of systems to support various customer scenarios. Here's how the technology addresses the use cases mentioned above:

- **Early QR Code System and Biometric Palm Scanning**: The initial QR code system linked a customer's Amazon account to their shopping experience. This allowed for easy identification and accurate billing upon exiting the store. The introduction of palm scanners in 2020 replaced QR codes. This system leverages unique vein patterns for secure identification and eliminates the need for physical scans (like QR codes) that could be less convenient. The palm scan connects directly to the customer's Amazon account for seamless billing.
- **Depth Cameras System and RGB Cameras System**: The pre-2022 system combined two technologies: depth cameras and load sensors. Depth Cameras identified individuals using 3D imagery and tracked their movements within the store. However, depth cameras don't capture color, making it challenging to distinguish between similar items, e.g. different flavors of the same product. The post-2022 updated system (RGB Cameras) utilizes RGB cameras that capture color information, improving accuracy and speed. These cameras integrated with Convolutional Neural Networks (CNNs) and Residual Neural Networks efficiently identify objects picked up by customers based on visual features and perform more precise product recognition, ensuring accurate tracking and monitoring.

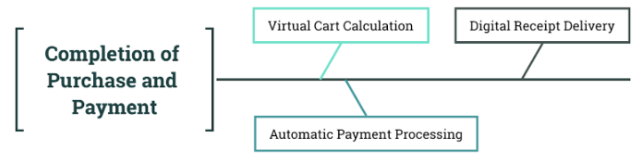


Fig. 6. Taking Characteristics from human and processing through

- **Completion of Purchase and Payment**: Throughout the shopping experience, the system builds a virtual shopping cart for each customer based on identified items. Once a customer exits the store, "Just Walk Out" technology uses the data collected to calculate the total cost of items in the virtual cart. The system automatically charges the customer's linked Amazon account for the identified purchases. Finally, a digital receipt will be sent directly to the customer's Amazon Go app for their easy reference.

Amazon's Just Walk Out system has revolutionized the retail experience by eliminating the need for a traditional checkout line. However, its implementation raises several questions regarding potential conflicts and the need for stringent quality assurance techniques.

1) **Entry and Identification**:
   Input: customers scan a QR code generated by the Amazon Go app or wave their palms above a palm scanner.
   Output: unique identification linked to their amazon account allows entry to the store.
2) **Tracking and Monitoring Customers' Movements**:
   Input: Cameras and RFID tags track customer movements and item interactions.
   Output: real-time monitoring of customers and items using computer vision algorithm and RFID technology.

Amazon's just Walk out system has simplified our shopping experience by eliminating the in-store and checking out process. However, the convenience is also require some trade-off selection between each categories:

- **Privacy vs. Convenience**: Trade-off: The implementation of the system that tracks customer movements and interactions raises privacy concerns. Consumers are often willing to sacrifice some privacy in exchange for greater convenience,but there's a delicate balance to strike.
- **Accuracy vs. Cost**: Trade-off: Achieving high accuracy in tracking customer movements and item interactions may require sophisticated technologies such as advanced computer vision algorithms and RFID systems, which can be costly to implement, the trade off of this is usually accuracy.
- **Scalability vs. Complexity**: Trade-off: Scaling the Just Walk Out system to accommodate larger stores or higher customer traffic levels may

introduce complexity in managing and maintaining the system.

- **Security vs. User Experience**:
  Trade-off: Implementing security measures such as fraud detection algorithms or access controls may introduce friction into the user experience.
- **Data Collection vs. Trust**:
  Trade-off: Collecting customer data to improve the shopping experience may erode trust if not handled transparently or if customers feel their privacy is compromised.
- **Speed vs. Reliability**:
  Trade-off: Maximizing the speed of transactions to reduce wait times at the store's entrance and exit may compromise the reliability of the system in accurately tracking items and customer interactions.

By carefully considering and managing these trade-offs, Amazon can design and implement a Just Walk Out system that effectively balances the needs of customers, the company, and other stakeholders while delivering a seamless and efficient shopping experience.

X. 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.

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

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.

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