SmartRetail - Multimedia Computing Project

Nádia Mendes 53175

Samuel Robalo 41936

Computação Multimédia 2022/2023

June 8, 2023

1 Introduction

In recent years, the retail industry has experienced a significant shift with the rise of online sales and e-commerce, despite this trend, many customers still prefer the in-store shopping experience. The SmartRetail project aims to achieve a system that enhances the in-store experience for customers and embraces the growing interest in innovative interaction technologies regarding the clients shopping experience.

The SmartRetail system will incorporate multimedia content, such as images and videos, will display product-related metadata on the screen in real-time, from a camera that monitor simulated in-store activity capable of detecting products in from of the camera or track various products on the shelves.

The system will be interacted with using a video camera and gestures, which can be simulated using a mouse. The camera will be positioned towards a shelf area and will identify different products using object detection. It will also detect activity, such as customers browsing or picking up products while showing in the screen relevant content. When no activity is detected the display will change its behaviour to show case other products and welcome messages or tips.

To track the products on the shelves, the system will use a combination of color, texture, edges, and object detection. The system will also display information about a specific product when a customer shows it, and tags will be used to decide what content to show.

The SmartRetail system aims to create an improved in-store shopping experience for customers, there are many challenges regarding the project that are explored in a higher detail on the next sections.

2 Related Work

Our project incorporates several key features that are already being utilized in ongoing projects and real-life applications. Some systems can be more complex than others and may require multiple cameras in a closed environment. During our research, we found that many shops in Japan are transitioning to fully automated stores that use advanced detection systems with multiple cameras for product image detection, monitoring customer actions, sleeves that can measure product weight, checkout monitoring, and redundancy to validate products in the shop.

However, there are instances where bugs and errors with detection can occur, leading to unexpected states within the system that may require manual intervention or observation within the shop and the system itself for added security, e.g misplaced item on a shelf.

One of the systems with similar use is the Amazon Go, which is a chain of convenience stores by Amazon, such system use multiple technologies like computer vision, deep learning algorithms and sensor fusion to enable a checkout-free shopping experience for customers.

Customers use the Amazon Go app when they enter the store, inside they can browse and store items on their bags. As customers pick up items, sensors and cameras track the items and add them to the customer's virtual cart, if a customer puts an item back it will be removed from the virtual cart.

After a client finished shopping, they simply leave the store without having to wait in line to pay, the App charges the customer's Amazon account for the items they have taken from the store and a digital receipt is sent to the customer's phone.

Another work we found interesting was about Recognition of Retail Products, it focus on an automatic checkout of products using OCR and uses features taken from an RGB camera and Line Scan Camera images, RGB and Gray colors respectively, the research is mostly about AI models of CNN and BERT and retaining knowledge acquired from different environments, it is interesting because it reads words and characters out of the product image, which is also a valid approach to match the same product with metadata, as distinct products should have different names and descriptions, making this good for product comparison, so we could conclude if it within our own images in back office and match accordingly, using OCR is currently outside our subject scope but it is still food for taught.

Can be hard for OCR to read from far away, but when a product is close to the camera it becomes viable option for analyzing, a single, close by product being shown by a client.

3 Specification

Our system requires 1 camera and a monitor, we will use OpenFramework and OpenCV along side with the Cpp programming language for development.

There will be various system states, when there is proximity of products or client interacting with the system, when it is in stand by an there is nothing to detect and when there are products in shelves being detected.

As shown in Figure 2, the main interface is simple and is comprised of:

1. Local system time.

2. Text Scrolling Information this can be information based on objects metadata their tags or related products or promotions could also be other information such as daily news.

3. Camera Feed Detection Screen, to detect existence of people, products and other user interactions such as hand movements and proximity of a product.

4. A contextual display, will show related images and video related to metadata provided from the camera detection.

As shown in Figure 3, the system will enter the shelf detection when was been a few seconds without no interaction with any client and there are items in the background.

This mode consists in the detection of products that are in the area covered by the field of vision of the camera, the system random pick a product in this area, and then will show some related images and videos in the screen, after a brief time the system will scan another random product in the area, if there is no item it will enter into standby mode, standby mode can be also triggered after certain elapsed time.

The interface is practically the same as shown in Figure 1, with the small difference that is the system random picking products in the area instead of being the customers scanning a product.

In Figure 2 we have an item that is found closer to the camera so brings us to UI 3 where there is an item detected from its features, item detection should be based of edge detection, average color of the item, edge histograms, comparison made with metadata from image version of the item stored in the back office, person gestures may be recognized to scroll to related images from the detected product.

In Figure 4 we have no special detection going on, no shelf items being detected no item or clients nearby, so it consists of displaying the products existing in the store and displaying the promotions of the store besides displaying consumer-friendly messages, this mode comes from the necessity of preventing unnecessary processing, making it more sustainable.

The contextual display is thought to be a good way to improve the relationship between the customer and the seller, since it shows the user related items information based of the metadata obtained from the video feed.

4 Application Implementation

To make our Camera Feed Detection Screen, to detect existence products and to compare with the data images and videos that we contain in our back office we compare the follow metadata such as tags, luminance, color, edge distribution e texture characteristics, this metadata we decided to store in a xml file in our back office contained in the paste images in our bin and then we use a series of algorithms such as ORB, Normalize Hamming to relate the extract metadata and make the correspondence.

To obtain the edges distribution we applied a set of filters in our class utils, namely the Gaussian blur that allowed us to reduce the irregularities of the images as well as smooth the pixel value of the images, we used Sobel in order to detect the horizontal and vertical edges of the images by calculating the partial derivatives in relation to the x and y directions. We also use Gabor Filter for edge detection in images since it allows capturing edges in different orientations and scales, thus allowing the detection of complex features. Finally, we used the Canny filter again to perform the edge detection since this proves to be quite effective as it allows us to smooth the image, calculate its gradient, perform non-maximum suppression and thresholding to identify edges with precision.

For detecting texture characteristics, we used Gabor Filter in our class utils, since it allows us to extract features in textures, in order to identify specific textual patters such as lines, dots and spots, thus allowing us to classify and recognize the textures contained in our images.

We use the ORB in the utils class as an alternative to the SIFT algorithm to detect key points and descriptors, thus interacting the resources of the detected image with the metadata contained in our back office in order to obtain a correspondence with the images contained therein. Finally, we use Normalize Hamming in the utils class, since this algorithm is particularly useful in scenarios where features may vary in terms of scale, intensity and other factors that are not relevant in the comparison we wanted. This algorithm by normalizing features using a Hamming window compensates for theses variations and ensures that image features are reliably and consistently comparable.

5 References

[1] Advanced shopping technology. Retrieved on March 16, 2023, from https://youtu.be/DKwa-ZFbYnk

[2] Amazon Go. Retrieved on March 16, 2023, from https://youtu.be/NrmMk1Myrxc

[3] Li, Z. et al. (2020). Augmented Reality Shopping System Through Image Search and Virtual Shop Generation. In: Yamamoto, S., Mori, H. (eds) Human Interface and the Management of Information. Designing Information. HCII 2020. Lecture Notes in Computer Science(), vol 12184. Springer, Cham. https://doi.org/10.1007/978-3-030-50020-7_26

[4] Shelfie (Youtube). Retrieved March 18, 2023, from https://youtu.be/-3VWDKonjkw

[5] Shelfie (Webpage). Retrieved March 18, 2023, from http://shelfie.labcd.unipi.it/

[6] Tobias Pettersson, Rachid Oucheikh, and Tuwe Lofstrom. 2022. NLP Cross-Domain Recognition of Retail Products. In 2022 7th International Conference on Machine Learning Technologies (ICMLT) (ICMLT 2022). Association for Computing Machinery, New York, NY, USA, 237–243. https://-doi.org/10.1145/3529399.3529436

6 Annex

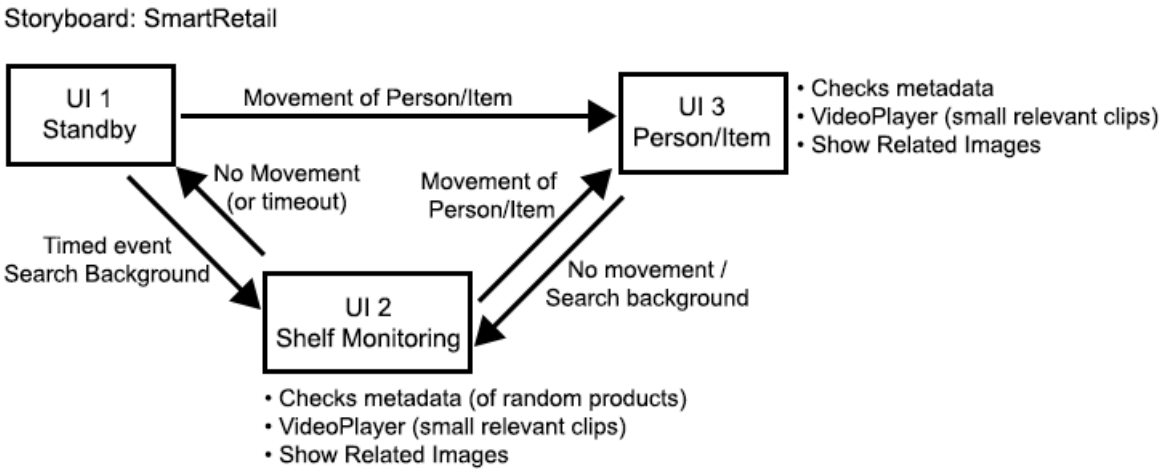

Figure 1: Storyboard for SmartRetail

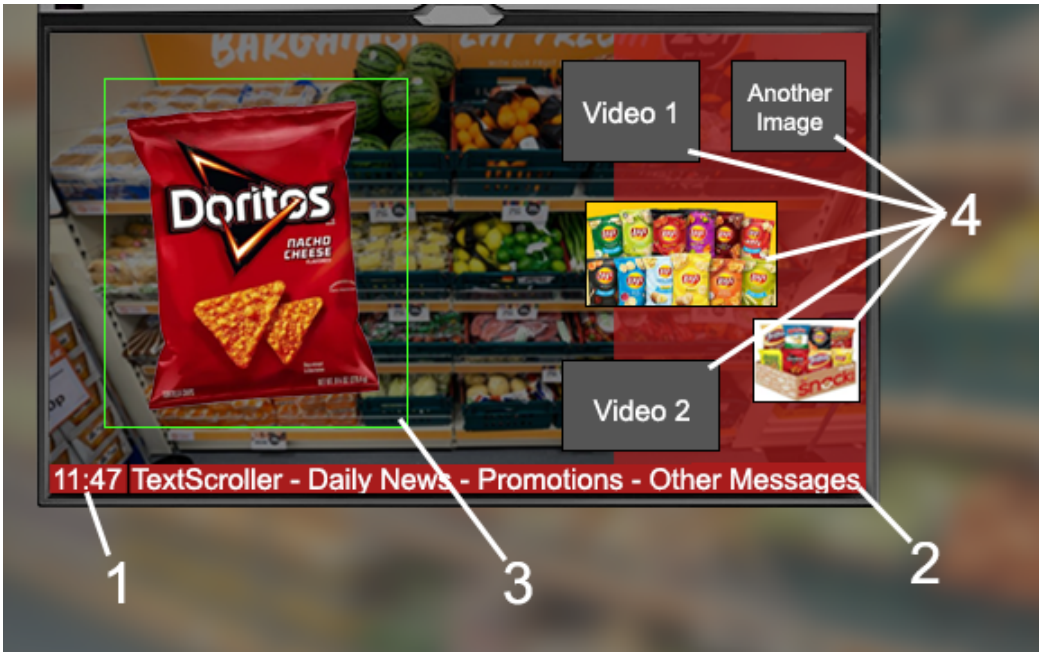

Figure 2: Example of the planned Prototype user interface

Figure 3: Items in shelves detection

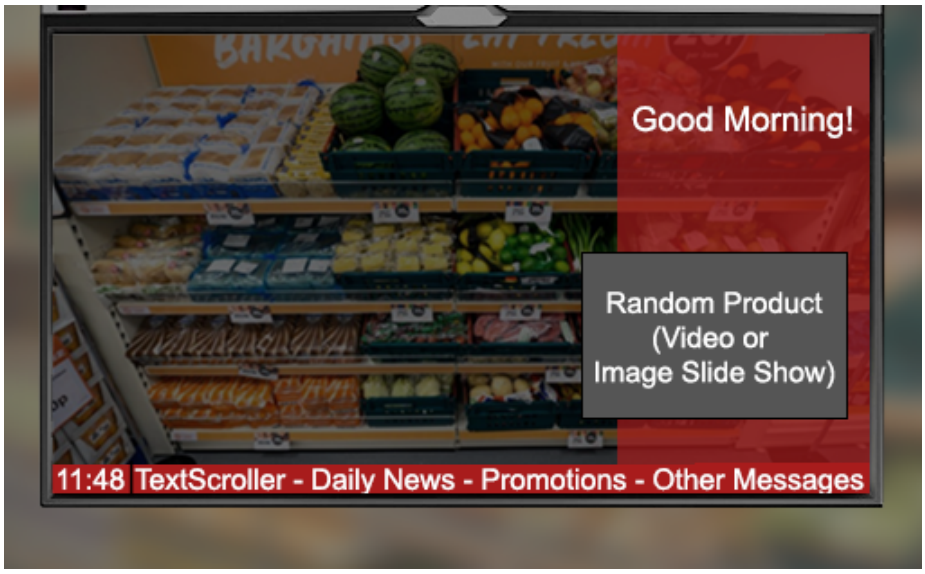

Figure 4: Example prototype in standby with no object detection

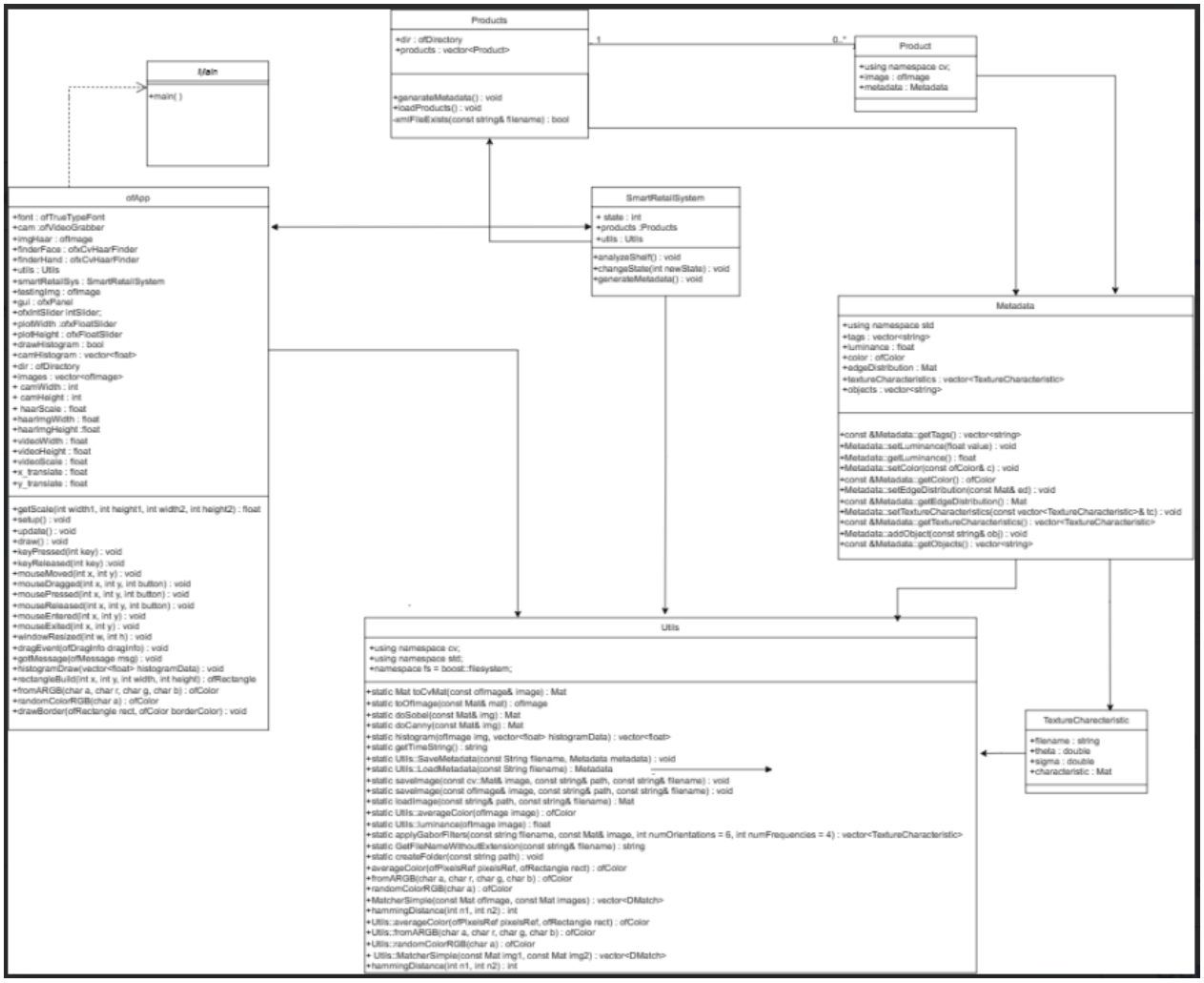

Figure 5: SmartRetail Class Diagram

7 User manual

The user must approach the camera, then grab the object that he wants to be detected and bring it closer to the camera and press the k key. As soon as it is detected, the image data begins to be generated and analyzed in order to obtain a correspondence and it will be displayed to the user. Users also have the following features:

If you wish, you can select the h key to view the histogram of the image to be detected, which can be moved along the display window and resized according to the desired dimensions by simply clicking on the histogram with the mouse by pressing the button on the left side.

The user can also increase the application's back ground visualization image according to the desired dimensions by simply clicking on one of its lower corners with the mouse, pressing the left button and dragging it until reaching the desired measurement.

The user can define the background color of the program and its opacity by pressing the i key, as well as the dimensions of the window by pressing this key, an interface of bars will be displayed, which the user will be able to adjust the values according to his preference. To remove the displayed interface, just press the i key again.

The user, if he wants to view the KeyPoints, just press the k key, and they will be arranged, if he wants to remove them, just press the k key again.

To view the FPS the user will only have to press the f key, to remove them just press the f key again.

To view the detection of the palm of the hand, the user must press the p key, if you want to stop viewing the palm detection rectangle, just press the p key again.

Finally, if the user wants to view the face detection, just press the key to stop viewing it, just press the o key again.