

Employing Video-based Motion Data with Emotion Expression for Retail Product Recognition

Ahmad B. Alkhodre, Abdullah M. Alshanjiti
Faculty of Computer and Information Systems
Islamic University of Madinah (IU)
Madinah 42351, Saudi Arabia

Abstract—Mining approaches based on video data can serve in identifying stores' performance by gaining insight into what needs to be proceeded to further enhance customers' experience, leading to increased business profits. To this end, this paper proposes an association rule mining approach, depending on video analytic techniques, for detecting store-items that are likely to be out of demand. Our approach is developed upon motion-tracking and facial emotion expression methods. We used a motion-tracking technique to record information related to customers' regions of interest inside the store and customers' interactions with the on-shelf products. Besides, we have implemented an emotion classification model, trained on recorded video data, to identify customers' emotions towards items. Results of our conducted experiments yielded several scenarios representing customer behavior towards out-of-demand stores' items.

Keywords—Shopper Behavior; motion tracking; emotion classification; machine learning; association rule learning

I. INTRODUCTION

The advancement of Internet retailers and online stores for over two decades has empowered and facilitated consumer experiences. Unlike visiting physical stores, online shoppers can comfortably search for their needs, place them at competitive prices, and then request a home delivery service with ease. To a considerable extent, Internet retailers are highly lucrative as they offer customers the facility to order a universal variety of products from all over the world. However, traditional retail stores necessitate being more competitive with offering services that guarantee customer satisfaction and loyalty. Consequently, marketing specialists and store owners are perpetually attempting to find any intelligent solution that could enhance the customer's shopping experience, using, e.g. sensors equipped with computer vision technologies [1]. Such technologies can aid retail stores in staying competitive and offer the best desirable services [2].

Understanding customer behaviors on one hand and business requirements to success, on the other hand, is essential for the sustainability of e-shopping. There are many traditional methods used by businesses to collect information about the needs of their customers (such as feedback questionnaires). These methods can provide good insight, but they have some significant drawbacks. Therefore, with the advancement of data collection and processing technologies as well as the use of highly accurate sensors, one can collect vast amounts of data about the customers and get many valuable insights [1].

Depending on video analytics, data mining approaches can serve in identifying a store's performance by gaining insight

into what needs to be carried out to further enhance customers' experience, leading to increased business profits [2]. Some helpful information that is obtainable through the use of Big Data analytics such as customers' regions of interest, customer count during a particular time, customer's emotion recognition, general information about the customer like their age and gender, interactions with on-shelf products, purchasing patterns and products that are likely to be bought together. Business owners can leverage this information to achieve great results that help in keeping customers returning to their stores and increase their financial returns.

This paper studies a learning-based solution for predicting the sales of items in retail chains and/or physical stores. In particular, we propose an association rule mining approach, depending on video analytic techniques, for detecting items that are likely to be out of demand. The association rule is a data mining technique that we apply to a store's transactional database to extract information about items (mainly about explicit features). Relying on available video analytic tools, we use a motion-tracking technique to record information related to customers' regions of interest inside the store and customers' interactions with the on-shelf products. Besides, we implement an emotion classification model, trained on recorded video data, to identify customers' emotions towards items. In a nutshell, this paper makes three contributions as follows:

- We propose an association rule-mining approach compounded with video analytic techniques for predicting the sales of items in physical stores, including out-of-demand items.
- We report on seven different scenarios representing customer behavior towards items that are out-of-demand.
- We conclude the paper by giving broad recommendations to tackle the symptoms of out-of-demand items.

The remainder of this paper is organized as follows: Section II represents the related works, shedding light on data analysis, motion-tracking, and emotion detection techniques. In Section III, we introduce a high-level design of our proposed predictor, and then we give a brief description concerning the technical tools used for implementation. Section IV presents and discusses the conducted experiments for assessing our proposal. Last but not least, Section V concludes the paper with suggestions for future research to consider.

II. RELATED WORK

This section reviews different sorts of related work, described into three aspects: data analysis, motion tracking, and

emotion detection aspects. The original contribution of this paper lies in mixing these three aspects in our proposed solution.

A. Literature Related to Data Analysis

The system suggested in [3] mines data from a transaction database of a retail market. It is supposed to be similar to popular mining tools such as Weka and RapidMiner. The data is stored in a MySQL database which is later accessed by simple Java and Python programs to apply a classification algorithm (C4.5) and an association rule mining algorithm (Apriori). Advantages of this system include being simple, user-friendly, and lightweight. In addition, information can be drawn directly from the operational database. The main goal of [4] is to discover the association between items to help store owners figure out the best layout for their stores. A simple Apriori algorithm is used to find the frequent itemsets and then apply the Lift measure for the association between sets to discover the rules between items. Before mining the association rules, the items of the store are divided and then grouped into sub-categories. This is useful because, in many cases, some items are considered substitutes.

Association rule mining in [5] is based on quantitative correlation coefficient (Pearson's correlation coefficient). It allows us to extract linear relations between two item sets. The suggested method starts by grouping transactions by the desired criteria (specific periods, for example), making it worthwhile to discover relevant rules within a given time frame. After the data has been grouped, a quantitative correlation analyzer is used to derive relations between item sets. Results of this implementation show more patterns detected in comparison to traditional association rule mining. In addition to discovering periodic customer demand which helps owners introduce offers and balance the supply and demand for products.

A study described in [6] proposes a model for finding periodically repeated patterns in a database. A pattern-growth algorithm (GPF-growth) was proposed to detect these patterns. A given pattern is considered frequent when it occurs within a defined maximum period, and its support exceeds the user-defined minimum. Thus, when it comes to analyzing data from the customers' transactions finding repeated patterns yields information related to regularly purchased products. The works described in [7], [8] apply the principles of association rule mining algorithms such as Apriori and FP-growth. Nevertheless, unlike these algorithms, the proposed approach focuses on the time of transactions. As a result, a new type of pattern is defined, which is called transitional patterns. The frequency of such patterns changes noticeably over a period of time. First, frequent patterns in the database are discovered by the use of an association rule mining algorithm such as Apriori or FP-growth. After that, an algorithm called TP-mine is introduced for the purpose of transitional mining patterns. With the use of this algorithm, it is possible to extract points in time where a frequency of an item-set has changed negatively or positively.

The aim of a study described in [9] is to introduce a method for analyzing a customer's basket data and deriving relevant association rules between its items. The approach relies on minimum spanning trees (MSTs). Using a minimum spanning tree for representing relations between items allows us to consider those strongly correlated items, which limits the search

space for the association rule mining. Another advantage of this approach is that not only can we extract strong association rules, but also, we can find items that tie other items together. In the study described in [10], an algorithm is proposed for detecting changes and trends in transactions data. It relies on association rule mining and the prediction of rules that may change at a later time. The algorithm works by tracking rules with high confidence, and the result of this algorithm is rules that will have higher confidence in the next period and rules which will have lower confidence. Apriori algorithm is used to mine rules over many time frames, and later each of the rules is given a score for each time frame to help track changes in confidence level over time. The approach provides great insight that could help prepare for upcoming association rules changes and find outlier rules that may apply in few time frames. However, the main disadvantage of this method is that it is heavily skewed based on the manually selected threshold values.

The main point of a study described in [11] is to understand the purpose behind customers' visits to a store. This is done through data mining transactions data and identifies the products in each transaction, and based on these products, the customers' purpose can be understood. The approach uses clustering techniques to segment different visit types, and the purpose behind these visits is later identified based on the types of products they contain. K-means clustering is used to segment visits in this research. The contribution of a study described in [12] is a framework that helps in predicting sales changes based on weather data. Two models are deployed to help with the prediction of the change in product demand. One of them is short-term, and the other is for long-term predictions. A LASSO Poisson regression model was used for predicting the impact of weather on customer demands.

A method for recommending the next purchase for customers is detailed in [13]. They have named the Co-Factorization model over sequential and historical purchase data (CFSH). Transactions data is mined to discover sequential purchasing patterns. From these patterns, two matrices are constructed. One for sequential and one for historical customer behaviors, these two matrices are both factorized to predict the best recommendation for the next customer visit. A study described in [14] suggests a method for segmenting customers based on their lifestyle. The segmentation is done with the help of analyzing a retailer's transactional database. The approach relies on a variable clustering algorithm (VARCLUS), resulting in clusters of related items. The lifestyles for each cluster are identified by looking into the type of items in each cluster. Each customer is then assigned to a lifestyle based on which cluster's item types did they purchase the most.

The contribution of [15] is a framework for product feature characterization and predicting the customer's most preferred specifications by analyzing purchase history data. A trained neural network model was used with the input as a matrix of all possible combinations of a product specification. The output is the predicted customer satisfaction rate. Predicting the customers' preferred characteristics in a product can guide a manufacturer to develop a product based on the customers' needs. A study described in [16] introduces a recommendation system that's based on collaborative filtering and data mining techniques. Customers are segmented based on their values

using the RFM analysis. The used clustering algorithm is K-means. To discover the best recommendation for a customer, the cluster that they belong to is identified first. Then, the Apriori algorithm is applied to transactions that belong to customers in the same cluster to extract the top associated items.

Another study described in [17] proposed a method to discover frequent itemsets with high value to the sellers. Association rule mining is applied on transactional databases considering the FM (frequency, monetary) values for the transactions. The result is association rules which have high revenue potential. These works highlight many important algorithms and techniques in the field of data analysis in order to find helpful information from the customer's transaction data such as products that have more or less sold, compatible products, time statistics, and more information, which we will mention it in the coming sections.

The main advantage of the proposed method in this paper is that it considers user-centric and item-centric recommendations to discover the correlation between customers and products, allowing for better recommendations.

B. Literature Related to Motion Tracking

Work described in [18] describes a computer vision system in real-time; this system is designed for an electronic billboard and recognition and track customers, the provide a piece of demographic information about these people. This information is used to update the current advertisement on current products to fit customer needs. The information provided by this system about customers includes age, the number of customers, how much time customers sets in front of the billboard. Another study described in [19] introduced a new system for monitoring and tracking the number of customers in some open places like a shopping center and museums hall. Two main techniques are used to develop this work first one is the laser scanner, and the second one is a single camera. To combine the information that extracts from these two tracking devices, Bayesian methods are used. The advantage of this approach is that it combines two techniques and uses them in one field of tracking people. However, the disadvantage of this approach is that it is inaccurate information is from the laser scanner. Work described in [20] provides an integrated system based on an RGB-D camera. This system can monitor customer behavior inside the shop environment. In addition, discover the interaction between customer and products through analysis of the recorded videos and classify this interaction into three types: pick up the product, pick up the product, and bring it back to the shelf and no interaction between customer and product.

The study described in [21] provides a system which distinguishes a variety of customer behavior opposite the products: incurious, taking a look, passing the body near shelf, touch the product, take product then return it to a shelf and take it and put into shopping cart, which gives us a hint that the customer has an interest in products. The given system is dependent on the orientation of head and body and arm action, which divided those into eight directions to estimate whether the customer searches to products or looking to the shelf. Then, a semi-supervised learning method was applied to

improve the training dataset and generate the file that contains accurate data. Work described in [22] showcases a system for tracking and identifying customers' interactions with on-shelf products inside a retail store. To accomplish the tasks of this system, an RGB-D camera is installed in a vertical position to capture the area of a store shelf. The Water Filling algorithm is used to recognize customers' interactions. Interactions such as a customer picking up a product off the shelf, putting a product back on the shelf, and customers grouping formation can be evaluated. The data collected from these interactions makes it possible to construct an intensity map on a shelf to show where interactions happened. Some of the most prevalent obstacles that a system like this might face are camera position, person's clothing, occlusion, body pose changes, diversity in product shapes, and constantly changing background. Some key factors make the implementation of this system viable, such as its affordable cost and ease of installation and maintenance. A study described in [23] proposes a framework for people detection in a store environment with the help of essential CCTV (Closed-circuit Television Camera) security cameras. The approach aims to achieve accurate customer detection from video recorded by the store's security cameras. An SVM (Support Vector Machine) binary classifier is used to classify each frame, where it returns positive if a person is detected or negative if not. Detected data is then recorded and mapped with coordinates to signify a customer's location within the video frame. This data is later mapped into a heat map to show the highest traffic areas within the store. The heat map visualization is based on Kernel Density Estimation (KDE). Advantages of this approach are the availability of CCTV cameras in many stores and an unambiguous visual representation of the exciting store regions.

Work described in [24] suggested CREEN system, which is an intelligent mechatronic system for help customers to search or to get the products which they want within the retail environment. This system works to reduce the needing to put helping icons or maps within the storing center. This system's primary function helps people move within the store by forecasting the probability that customers will attract positions that are analyzed in front of a shelf of products. There are installed tools on the cart that include information about the location of elements and a location map for elements, which in turn leads the customer to the product he wants to pick. This system aims to develop a robot that searches for the best location to put the product within the store. Also, it can be used for blind and older adults. A study described in [25] presents the VMBA problem, which is based on set up cameras on store carts and considers three main stages, which include: the interaction, location, and the scene that is taken from the camera. Which in result goal to inference the behavior of customers within the retail environment. Furthermore, by merging these behaviors with information exported from the analysis, the market cart-like (transactions) gives retail store owners the management of areas and shopping strategies. Work described in [25] provides a new methodology for tracking, which is based on the visual attention focusing for customers WVFAO for many persons and which is used on understanding human behavior. This approach is based on capturing the attention of wandering to external ads and considers Bayesian network (HDBN) to discover the number of people in scene, body, head location (Direction), the interactions, and WVFAO

their own. In addition, this research includes the way to design the model WVFOA.

As a result of previous research, we can list many techniques in the field of tracking the motion of people and the study of their behavior. These techniques will provide our work the complete information about customers' behavior within the store, such as information about customers' (age, gender), customer count and region of gathering people, etc. In addition to that, the interactions between customers and products like how much time customers spend in front of a product, if interested or not. This information considers the main thing in our work which we will mention this necessary in the following steps.

C. Literature Related to Emotion Detection

A study described in [26] explains that a strong structure is a design to recognize the three-dimensional position and local emotional expressions of the face, such as eyelid movements and mouth movements, which given a significant role in explaining the face emotion expression, by using RGB-D camera and the Kalman Extended filters. Advantages of this method include high 3D data with the information of color and density, high accuracy, and full automation. Disadvantages include that it is sensitive to light cases and gets confused by shadows. Another study described in [27] study the face recognition depending on FTFA-DLF. FTFA-DLF can merge deep learning features and handcraft features extracted from the nose, eye, and mouth regions. It used to help deep learning features by adding deep learning and handcraft features to the objective function layer to get better facial recognition performance based on the LFW dataset. Advantages include color information, and the descriptor computation strategy does not affect performance, high accuracy, and is fully automated. Disadvantages include missing pixels and brightness have a harmful effect and data storage.

Work described in [28] study an emotion recognition system depending on a MATLAB environment into a MATLAB Simulink environment that is able to recognize facial expression automatically in real-time. The label and the dataset used in this study build-up from the videos. The facial recognition system built on the programmable array and the camera sensor uses through this study can recognize facial emotion in actual time at a frame rate of 30. Advantages include FPGA devices being updatable, fully automated, high-performance, and FPGA no need cost compared to ASLCS. Disadvantages include Data storage and camera angle. Work described in [29] presents some algorithms for facial expression recognition grin detection. The algorithms depend on deep machine learning a [CNN], the main goal of this network is to select one of the six types of emotions using the [CMN] MultiPie database, by training through this database parallel on a large number of independent flows on [GPU]. Advantages include accuracy in image detection problems and being fully automated. Disadvantages include being very slow to train and needing to use a lot of data and data storage and camera angel training.

A study described in [30] suggests a way for emotion recognition-based facial components. It is complete in extracting the local features of the mouth and eye from each frame using GW with selected orientations and scales. The

express that features on to classifier for detecting of the face scope. From detection, each pixel on through the face. Finally, select and recognize the emotion of the face by using the Adboost algorithm. Advantages include improved performance compared to the current techniques because of its new features and not affected by changes in colors lighting. Work described in [31] applies the concepts of deep learning for detecting facial expression, so they proposed a facial expression monitoring system [PFEMS] in addition to design a convolutional neural network model using TensorFlow into two parts: a) Validation tools and b) Training model for data training after extracts the facial images from the video frames into the facial detector then detecting this image by using CNN to monitor the emotion state from the six universal emotions [angry, disgust, happy, surprise, sad and fear].

A study described in [32] introduces a method for the classification of facial expressions, which is based on a fuzzy logic model. The approach relies on a supervised machine learning method, which uses pre-existing databases that contain images of faces labeled with their respective emotions. A set of fuzzy logic rules are generated from the given data using the FURIA algorithm. These rules are generated based on the cosine values of the essential triangles plotted on the most critical points on a face. Each generated fuzzy rule presents some conditions, which are then used to classify different emotion types. To start the emotion recognition, first, a face must be detected. For this task, the DLIB toolkit was used for real-time face tracking and provide a set of 68 landmarks from the face detected. Some triangles are then overlaid over these points, and the cosine value is calculated for all the vertices of the triangles. The values are later passed to the generated fuzzy rules to recognize the emotions expressed in the image. The main advantages of the proposed method include working with image files, pre-recorded video, and live webcam, and it can classify emotions in real-time and detect multiple faces simultaneously. Limitations of this study include it can only classify six basic emotions, the model used is influenced by the quality of the data used in training it, and the model detects most intensive emotions better than less intensive emotions.

While Cisco has afforded different and promising solutions towards improving the shopper experience using IoT and big data analytical approaches [33], [34] (i.e. depending on monitoring and analyzing the shopping path and their movements), their focus stayed on studying horizontally a massive set of data without paying much attention to specific products. Nevertheless, this paper differs in that our approach focuses on detecting a particular product (i.e. out-of-demand store-items) based on the shopper's interaction with a product vertically.

III. APPROACH

This section presents our proposed association rule-mining approach, which consists of three main components (i.e., data analysis, motion detection, and emotion detection), illustrated in Figure 1. We explain these components in some detail in the following subsections.

A. Data Analysis

This component is concerned with analyzing data collected from stores' transactional databases using association rule

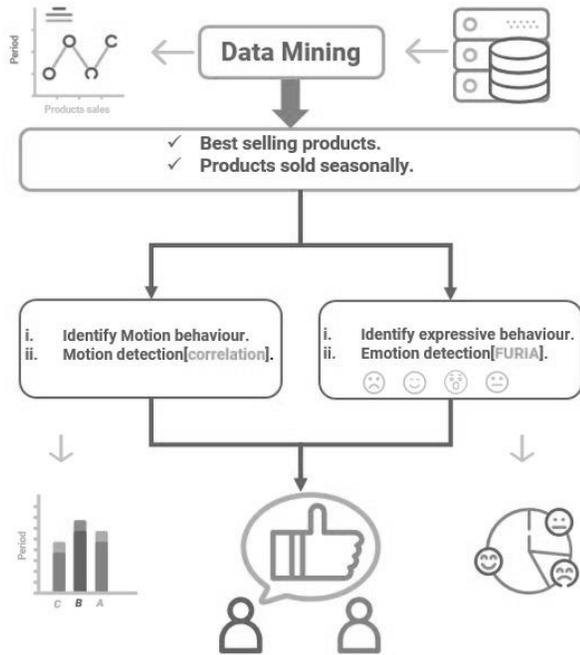


Fig. 1. The Top Level Workflow of our Proposed Approach, Illustrating the Interaction between the three Main Components (i.e. Data Analysis, Motion Detection, and Emotion Detection).

mining. Through this analysis, business owners can discover knowledge pertaining to associated store items that could be bought together, customers’ periodic purchasing patterns, product performance, and other sound patterns. In addition to analyzing the data from the customers’ transactions, this component supports analyzing video data collected from motion and emotion observation tools.

There are different algorithms for association rule mining, but the two widely used are Apriori and FP-Growth. We choose to implement the FP-Growth algorithm as it performs better with larger datasets. In addition, the FP-Growth algorithm relies on the divide-and-conquer approach and only performs two full I/O scans of the database, making it more suitable for more large databases. Concerning the tool used, we have implemented the FP-Growth algorithm using PyFpgrowth¹.

More in detail, to use PyFpgrowth, one needs to encode data from transactional structures into a 2-dimensional list object, where each sub-list contains the items of a single transaction. To generate the association rules, one also needs to specify the minimum values for support, confidence, and lift. By specifying these values, the best rules can be generated while controlling their size. The output from this algorithm is a set of association rules, each with its specific values for support, confidence, and lift.

Figure 2 describes the results generated from the association rules from a sample dataset used in our experiments using

¹A Python implementation of the Frequent Pattern Growth algorithm <https://pypi.org/project/pyfpgrowth/>

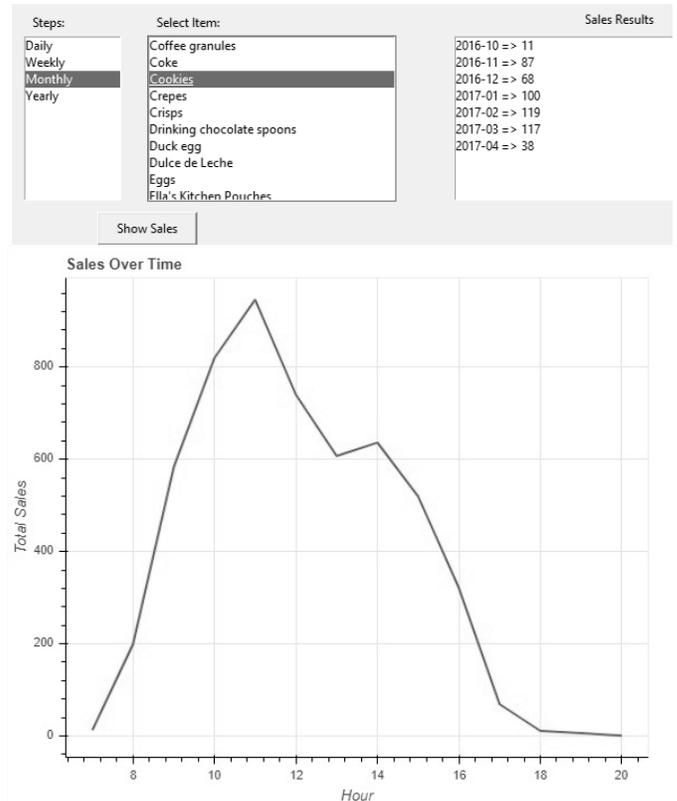


Fig. 2. Examples of Association Rules Generated by FP-Growth Algorithm.

the FP-Growth algorithm with the configurations (Minimum support = 0.05, minimum confidence = 0.2, minimum lift = 0.4). The aim here is to infer the total sales of a selected product in specific periods (weekly, monthly). This is essential to detect the change in customer behavior and their inclination towards some products. By using the same previous dataset of bakery sales, we can view a list of unique products present in the transactions. The next step is to select one of the products from the list and the period steps to show the total sales. The output of choosing the product 'cookies' and choosing the monthly steps shows the total sales of the product for each month recorded in the dataset.

The ability to discover relations between items and to show the change in customer inclination to buy a particular product is helpful for this project as they provide us with critical information about the customers’ behavior. Although sometimes the reasons behind these changes in customer behavior may not be apparent to the business owners. Hence, we suggest that business owners need to further inspect the reasons behind the ambiguous changes using the proposed methods for customer tracking and emotion classification, discussed in the next subsections.

B. Motion Detection

Store video analytics is an essential technique to understand the customers’ behavior accurately. With the help of this component, we can collect information related to customers’ region of interests (ROIs), customer count, density maps, customer’s interaction with on-shelf products, and general information about the customer such as their age and gender. The system

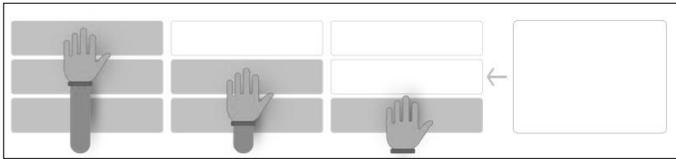


Fig. 3. Classification of Customers' Interest based on Hand Movement.

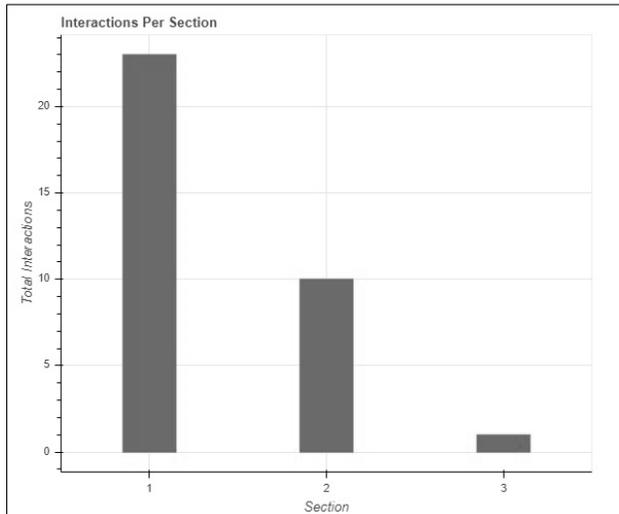


Fig. 4. Result of the Interaction Per Section between the Customer and the Selected Product.

also records data related to the customers' interactions with the products on the shelf; this is done by applying a motion detection algorithm that we implemented. We concluded that the best way to capture the customers' interactions is through a camera that is installed above the shelf. The camera shall only record the area which is very close to the shelf.

The used algorithm works with the real-time video that is captured from the camera and divides every captured video frame into three equal sections, as shown in Fig. 3. The main reason behind this separation is to provide us with a way to classify the customers' interactions based on how close they are to the shelf. The algorithm will then detect the interactions (motion) that happen at each section separately and record the section number and the time when the interaction happened.

This data can help business owners understand the types of interactions that their products get. For example, if we compare the total number of interactions recorded in Section 1 (closest section to the product) and the total sales of that product, we could detect cases where products might get a lot of interaction/interest from customers but a low number of those customers purchasing the product. Thus, we can conclude that there might be something we can improve with that product.

The main goal of the algorithm is to detect and record motion-captured from real-time video. The execution starts by defining a list that contains three objects of the Frame class. The Frame class has a constructor that defines two main properties for each frame. First is the content property, which will be assigned the image captured from a video source. Second is the stoppage frames property that will start as 0

initially but should later be incremented accordingly. Finally, we create a second list that should hold the three corresponding reference frames (background frames) during execution. These frames are only initialized once.

Execution continues by reading the first frame from the video source (camera). The frame will then be split into three equal (in height) frames. Each one of these subframes will be assigned to the three previously declared frame objects. We then iterate over each of the frame objects to detect movement separately for each section. The iteration over a frame object starts by processing the frame to be ready for movement detection. Frame processing includes: (1) converting the frame to grayscale and (2) applying Gaussian blur on the frame.

The next step is to check if the corresponding reference frame is initialized. To determine if there is a movement that occurred between our current frame and its reference frame, we need to:

- 1) Calculate the absolute difference between them and assign it to a new frame.
- 2) Apply threshold on the subtracted image with a chosen threshold value.
- 3) Dilate the threshold image.
- 4) Find contours in the dilated image.
- 5) Calculate the total area of the contours found.

Next, check if the total area of contours exceeds the set minimum. Setting the minimum value depends on how large the object one needs to track (hand in this case). If the minimum is exceeded, we check if the stoppage frames for the current frame if it exceeds a set value (15) that means that an object has entered the frame, and we can record the current time and the section related to the movement and we have to reset the stoppage frames back to 0 because the frame currently has movement.

If the total area of contours did not exceed the minimum value, we could increment the stoppage frames for the object. The execution will then continue for the remaining two frames. After that we begin executing the same steps for the next frame of the video source. Fig. 4 displays the interaction per section between the customer and the selected product.

C. Emotion Detection

Advancement in the field of deep learning and computer vision allows to capture the facial expressions of a customer and accurately predict their displayed feeling. This can be helpful for getting information about the customers' feelings toward certain products. Understanding how the customer feels opens up many questions about why they feel a certain way in a given moment. Through understanding how the customers feel, we can propose solutions to enhance their shopping experience further. The system analyzes video recordings of people and detects their facial expressions, then record and store that data. The emotions detected from peoples' faces can be helpful in trying to understand their behavior and to evaluate their satisfaction rate. A business must keep the customers' satisfaction to retain their loyalty.

The input is taken from a video camera. At the start, the algorithm splits the video into a list of frames, and each frame constructs an image frame. Then a face must be detected within

the image frame. This is done with the DLIB toolkit. Then cover the face that is discovered with a rectangular shape. The toolkit will also plot some points and form triangles on significant areas in the face (eyes, nose, and mouth) detected and calculate the cosine values for all vertices. Values are then passed to some fuzzy rules to determine the emotion detected from the frame. Testing was done using a video recording of one person. The length of the video was 42 seconds, and the recorded instances in the results were 450. The result is recorded in an Excel file with their respective time stamps and emotion classifications. Then Represent the results in a byte chart to make them more understandable by the client. The recorded result contains 450 instances where each one has a precise time stamp and a value for each emotion recorded as shown in Figure 5.

Data on peoples' expressions could be further analyzed to understand the meaning behind them. For this paper, we found a reliable way to collect such data, but this information is handy to business owners because they need to know the customers' feelings and understand their causes to adapt their business to improve the customers' satisfaction. In our case, we consider that the customer needs to touch the product to know the price.

Time	Msec	Stamp	Happy	Sad	Surprise	Fear	Disgust	Anger	Neutral
00:00:12	967	0.11.967	0.00	0.00	0.00	0.00	0.00	0.00	1.59
00:00:12	0	012.000	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	33	0.13.033	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	67	012.067	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	100	012.100	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	133	021.200	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	167	012.233	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	200	012.267	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	233	012.300	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	267	012.333	0.00	0.00	0.00	0.17	0.00	0.00	1.81
00:00:12	300	012.367	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:12	333	012.400	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:12	367	012.433	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:12	400	012.467	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:12	333	012.500	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:12	467	012.533	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:13	500	012.567	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:13	533	012.600	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:13	567	012.633	0.00	0.00	0.00	0.35	0.00	0.00	1.99
00:00:13	600	012.667	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	633	012.700	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	667	012.767	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	7000	012.800	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	767	012.833	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	800	012.867	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	833	012.900	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	900	012.933	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	933	012.967	0.00	0.00	0.00	0.52	0.00	0.00	1.81
00:00:13	967	013.00	0.00	0.00	0.00	0.56	0.00	0.00	1.81
00:00:13	0	013.033	0.00	0.00	0.00	0.56	0.00	0.00	1.81
00:00:13	33	013.067	0.00	0.00	0.00	0.56	0.00	0.00	1.81
00:00:13	67	013.100	0.00	0.00	0.00	0.56	0.00	0.00	1.81
00:00:13	100	013.133	0.00	0.00	0.00	0.56	0.00	0.00	1.81
00:00:13	133	013.167	0.00	0.00	0.00	0.56	0.00	0.00	1.81

Fig. 5. A Screenshot Illustrating the Generated Output from the used Emotion Detector.

IV. EXPERIMENT AND DISCUSSION

A. Questionnaire

We prepared a questionnaire² to identify customers' emotional reactions against the items with exorbitant prices, out-of-date products, and so on. 150 people took the questionnaire

²https://docs.google.com/forms/d/e/1FAIpQLSdSRhmV9esbHdCSZQ6n\CvDMm2e-JK4Q-M_FyYQ89VEC7J5xJg/viewform

(90 males and 60 females). The mean age of the participants was 22 years. The responses were statistically analyzed to find the customers' emotions against the design and price of the items. Based on the results of the questionnaire, we classified emotions into six categories, which include Happy, Sad, Surprise, Fear, Disgust, and Anger.

B. Experimental Setup

We experimented with a grocery store in which some products were facing sales problems, as reasoned from the transaction database of the store. We picked these products for our experiment. We set up the proposed system for 30 days to analyze the behavior of customers towards the product. During the experiment, 3000 customers visited the store in total, with an average 100 customers per day. Out of these, 2500 customers interacted with the products monitored with an average of 83 customers per day.

C. Results

TABLE I. SUMMARY OF RESULTS DEPENDING ON BOTH THE PROPOSED PREDICTOR AS WELL AS THE QUESTIONNAIRE RESULTS.

Emotion	Motion		
	Excited (>50%)	Hesitant (>50%)	Careless (>50%)
Happy	Price/Placement	Design	Placement
Sad	Placement	Placement	
Surprise	Price	Design	
Fear	Price	Design	
Disgust	Design	Placement	
Anger	Placement	Placement	

Table I summaries our findings based on the measures from both prediction tasks by our proposal as well as from questionnaires. Moreover, we report on seven different scenarios representing customer behavior towards items that are out-of-demand, as follows:

Case 1: This case is an intersection between Excited and Careless. If no motion is recorded, then no emotion can be predicted, and the problem can be addressed by correcting the placement.

Case 2: This case represents the emotional happiness for more than 50% of customers towards the selected product and the interaction between customer and product (i.e. excited, hesitant, careless). If the customer feels happy and there is an interaction between the customer and product, but the product's sales are still low, then the problem is in the product price, which needs to be improved.

Case 3: This case represents the emotional happiness for more than 50% of customers towards the selected product and the interaction between customer and product (i.e. excited, hesitant, careless). If the customer feels happy and there is no interaction between customer and product, but there are movements in front of the product, the problem is in the product design, which needs to be improved.

Case 4: This case represents the emotional sadness for more than 50% of customers towards the selected product and the interaction between customer and product (i.e. excited, hesitant, careless). If the customer feels sad and there is an interaction between customer and product, then the problem is in the product placement, which needs to be corrected.

Case 5: This case represents the emotional sadness for more than 50% of customers towards the selected product and the interaction between customer and product (i.e. excited, hesitant, careless). If the customer feels sad and there is no interaction between customer and product, but there are movements in front of the product, then the problem is in the product placement and needs to be corrected.

Case 6: This case represents the emotional surprise for more than 50% of customers towards the selected product and the interaction between customer and product (i.e. excited, hesitant, careless). If the customer feels surprised and there is an interaction between customer and product, but the product's sales are still low, then the problem is in the product price, which needs to be improved.

Case 7: This case represents the emotional surprise for more than 50% of customers towards the selected product and the interaction between customer and product (i.e. excited, hesitant, careless). If the customer feels surprised and there is no interaction between customer and product, but there are movements in front of the product, then the product design problem needs to be corrected.

V. CONCLUSION AND FUTURE SCOPE

Mining approaches based on video data can serve in identifying a store's performance and production by gaining insight into what needs to be proceeded to further enhance customers' experience, leading to increased business profits. To this end, we have proposed an association rule mining approach, depending on video analytic techniques, for detecting store-items that are likely to be out of demand. Our approach is developed upon motion-tracking and facial emotion expression methods. Results of our conducted experiments yielded seven different scenarios representing customer behavior towards out-of-demand stores' items.

Regardless of the high computational costs associated with the mining process, challenges to ideally apply our approach require overcoming: illumination conditions, complex people activities, crowded areas, constantly changing backgrounds, occlusion, and ineffective camera placements. We plan to tackle these challenges in the future besides conducting other experiments on a more comprehensive real-world dataset to validate the concept of our approach.

ACKNOWLEDGMENT

We are grateful to Wasim Ahmad  for his advice on motion-tracking and emotion analysis methods, and their use in association rule-mining..

REFERENCES

- [1] I. E. Olatunji and C.-H. Cheng, "Video analytics for visual surveillance and applications: An overview and survey," *Machine Learning Paradigms*, pp. 475–515, 2019.
- [2] S. Chandramana, "Retail analytics: Driving success in retail industry with business analytics," vol. Volume 7, 08 2017.
- [3] K. P. Nagwekar, K. S. Chowdhary, G. M. Shejwal, and V. N. Alone, "Data visualization and data mining for retailers," *International Journal for Innovative Research in Science and Technology*, vol. 3, no. 11, 2017.
- [4] K.-I. Ahn, "Effective product assignment based on association rule mining in retail," *Expert Systems with Applications*, vol. 39, p. 12551–12556, 11 2012.

- [5] C. F. Cheung and F. Li, "A quantitative correlation coefficient mining method for business intelligence in small and medium enterprises of trading business," *Expert systems with applications*, vol. 39, no. 7, pp. 6279–6291, 2012.
- [6] K.-C. Lin, I.-E. Liao, T.-P. Chang, and S.-F. Lin, "A frequent itemset mining algorithm based on the principle of inclusion-exclusion and transaction mapping," *Information Sciences*, vol. 276, pp. 278–289, 2014.
- [7] R. U. Kiran, J. Venkatesh, M. Toyoda, M. Kitsuregawa, and P. K. Reddy, "Discovering partial periodic-frequent patterns in a transactional database," *Journal of Systems and Software*, vol. 125, pp. 170–182, 2017.
- [8] Q. Wan and A. An, "Discovering transitional patterns and their significant milestones in transaction databases," *IEEE transactions on knowledge and data engineering*, vol. 21, no. 12, pp. 1692–1707, 2009.
- [9] M. A. Valle, G. A. Ruz, and R. Morrás, "Market basket analysis: Complementing association rules with minimum spanning trees," *Expert Systems with Applications*, vol. 97, pp. 146–162, 2018.
- [10] M. Kaur and S. Kang, "Market basket analysis: Identify the changing trends of market data using association rule mining," *Procedia computer science*, vol. 85, pp. 78–85, 2016.
- [11] A. Griva, C. Bardaki, K. Pramatar, and D. Papakiriakopoulos, "Retail business analytics: Customer visit segmentation using market basket data," *Expert Systems with Applications*, vol. 100, pp. 1–16, 2018.
- [12] G. Verstraete, E.-H. Aghezzaf, and B. Desmet, "A data-driven framework for predicting weather impact on high-volume low-margin retail products," *Journal of Retailing and Consumer Services*, vol. 48, pp. 169–177, 2019.
- [13] P. Wang, J. Chen, and S. Niu, "Cfsh: Factorizing sequential and historical purchase data for basket recommendation," *Plos one*, vol. 13, no. 10, p. e0203191, 2018.
- [14] V. L. Miguéis, A. S. Camanho, and J. F. e Cunha, "Customer data mining for lifestyle segmentation," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9359–9366, 2012.
- [15] J. Zhang, A. Simeone, P. Gu, and B. Hong, "Product features characterization and customers' preferences prediction based on purchasing data," *CIRP Annals*, vol. 67, no. 1, pp. 149–152, 2018.
- [16] F. Rodrigues and B. Ferreira, "Product recommendation based on shared customer's behaviour," *Procedia Computer Science*, vol. 100, pp. 136–146, 2016.
- [17] C.-H. Weng, "Revenue prediction by mining frequent itemsets with customer analysis," *Engineering Applications of Artificial Intelligence*, vol. 63, pp. 85–97, 2017.
- [18] I. Haritaoglu and M. Flickner, "Attentive billboards: Towards to video based customer behavior understanding," in *Sixth IEEE Workshop on Applications of Computer Vision, 2002.(WACV 2002). Proceedings. IEEE*, 2002, pp. 127–131.
- [19] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Multi-modal tracking of people using laser scanners and video camera," *Image and vision Computing*, vol. 26, no. 2, pp. 240–252, 2008.
- [20] D. Liciotti, P. Zingaretti, and V. Placidi, "An automatic analysis of shoppers behaviour using a distributed rgb-d cameras system," in *2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA)*. IEEE, 2014, pp. 1–6.
- [21] J. Liu, Y. Gu, and S. Kamijo, "Customer behavior classification using surveillance camera for marketing," *Multimedia Tools and Applications*, vol. 76, no. 5, pp. 6595–6622, 2017.
- [22] E. Frontoni, P. Raspa, A. Mancini, P. Zingaretti, and V. Placidi, "Customers' activity recognition in intelligent retail environments," in *International Conference on Image Analysis and Processing*. Springer (Berlin, Heidelberg), 2013, pp. 509–516.
- [23] T. Katanyukul and J. Ponsawat, "Customer analysis via video analytics: customer detection with multiple cues," *Acta Polytechnica Hungarica*, vol. 14, no. 3, pp. 187–207, 2017.
- [24] M. Paolanti, D. Liciotti, R. Pietrini, A. Mancini, and E. Frontoni, "Modelling and forecasting customer navigation in intelligent retail environments," *Journal of Intelligent & Robotic Systems*, vol. 91, no. 2, pp. 165–180, 2018.

- [25] V. Santarcangelo, G. M. Farinella, A. Furnari, and S. Battiato, "Market basket analysis from egocentric videos," *Pattern Recognition Letters*, vol. 112, pp. 83–90, 2018.
- [26] Y. Dong, Y. Wang, J. Yue, and Z. Hu, "Real time 3d facial movement tracking using a monocular camera," *Sensors*, vol. 16, no. 8, p. 1157, 2016.
- [27] Y. Li, Z. Lu, J. Li, and Y. Deng, "Improving deep learning feature with facial texture feature for face recognition," *Wireless Personal Communications*, vol. 103, no. 2, pp. 1195–1206, 2018.
- [28] S. Turabzadeh, H. Meng, R. M. Swash, M. Pleva, and J. Juhar, "Facial expression emotion detection for real-time embedded systems," *Technologies*, vol. 6, no. 1, p. 17, 2018.
- [29] L. Ivanovsky, V. Khryashchev, A. Lebedev, and I. Kosterin, "Facial expression recognition algorithm based on deep convolution neural network," in *2017 21st Conference of Open Innovations Association (FRUCT)*. IEEE, 2017, pp. 141–147.
- [30] P. I. Rani and K. Muneeswaran, "Emotion recognition based on facial components," *Sādhanā*, vol. 43, no. 3, pp. 1–16, 2018.
- [31] J. Hu, "A personal facial expression monitoring system using deep learning," Ph.D. dissertation, 2019.
- [32] K. Bahreini, W. Van der Vegt, and W. Westera, "A fuzzy logic approach to reliable real-time recognition of facial emotions," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 18 943–18 966, 2019.
- [33] "Cisco analytics for retail," https://www.cisco.com/c/dam/en_us/solutions/industries/docs/retail/analytics-retail-aag.pdf, accessed: 2021 November.
- [34] "Retail, digitize and secure your entire retail environment for a more connected experience," <https://www.cisco.com/c/en/us/solutions/industries/retail.html>, accessed: 2021 November.