Journal of Engineering Science and Technology
5th EURECA 2015 Special Issue March (2017) 82 - 96
© School of Engineering, Taylor's University

THE APPLICATION OF IMAGE PROCESSING FOR IN-STORE MONITORING

KAH HENG LAU*, WEI JEN CHEW

School of Engineering, Taylor's University, Taylor's Lakeside Campus,
No. 1 Jalan Taylor's, 47500, Subang Jaya, Selangor DE, Malaysia
*Corresponding Author: henglau_93@hotmail.com

Abstract

Merchandisers and store owners are increasingly paying attention to in-store monitoring systems to help improve sales and profit by finding efficient and convenient ways to identify their customer's interest. In this paper, the application of image processing through video surveillance for in-store monitoring was proposed as the solution for understanding a customer's shopping behaviour as well as monitoring the attendance of employees in real time. This system will be able to differentiate between a customer and an employee walking into a store and will only track the movements of a customer in the store. An integrated system involving face detection, face recognition and human tracking algorithms was proposed. A database of training images of known faces was generated to test the face detection and face recognition algorithms while a video which recorded people moving around in a confined space was used as the input for the human tracking algorithm. Viola-Jones was implemented for the face detection technique, eigenface was used to separate out the customers from the employees while background subtraction and Kalman filter techniques were used in the human tracking algorithm. Results show that the proposed human tracking algorithm is able to differentiate the customers from the employees and is able to provide accurate data for a customer's shopping behaviour study like the amount of time spent at a particular location in a store.

Keywords: In-store monitoring, face detection, face recognition, human tracking.

1. Introduction

Customers today have many options when it comes to buying a product. Therefore, it has become increasing competitive for a store to know what their target customers want to help determine which product is popular and marketable at the moment. This helps the store stock up on popular items which helps increase sales. One method is to observe the behaviour of the customer in a store to help identify their shopping patterns or determine a suitable location to place a merchandise based on their customer’s habits.

Traditionally, frequent point-of-sale visiting is often implemented in a store to help retrieve qualitative information of customer behaviours but this can be time consuming. However, minimising the frequency of the visits will compromise the company’s ability to determine their customer’s behaviour since they will only be getting minimal level of information and insight. To study the customer’s interest and shopping behaviour, surveys are usually conducted. However, they are often time consuming and the data collected is qualitative, which is hard to analyse. Furthermore, by looking only at the best-selling products, the current customer’s interest can be known but the possible potential of other products may be missed.

Therefore, in this paper, the application of image processing for in-store monitoring is proposed as the solution for understanding customer’s shopping behaviour. The advantages of using image processing method is that it offers shorter data collection and processing time as well as the ability to track customers without them being aware of it. This helps a store collect honest and unbiased opinions which will be more accurate that filling up a survey form.

Popa et al. [1] proposed a camera system that is able to determine the shopping behaviour of a person. This system captures various locations of the store and tracks the movement of a shopper. It helps categorise the shopper either as being goal orientated or disorientated. The main objective is to create an automatic assessment tool that is able to determine the preferences of a shopper and help merchandisers determine the best locations to place an item or which item is well received.

Sicre et al. [2] proposed a video surveillance system that is able to recognise the behaviour of a person based on how this person interacts with a merchandise. At the point of sale, which is where the product is located, the authors concluded that there is six possible behaviours that can be shown by a shopper. These behaviours are enter, exit, interested, interaction, stand by and inactive. Basically, in each scene which corresponds to a merchandise, the behaviour of the shopper would be identified to determine the level of interest of that shopper.

Liciotti et al. [3] also tried to determine the behaviour of shoppers by observing their interaction with the products. A RGB-D camera was installed near the shelf of a product and based on how a shopper interacts with the products, that encounter will automatically be classified as either positive, negative or neutral. This is to determine whether the customer picked up an item, repositioned an item or if nothing was taken respectively.

Vildjiounaite et al. [4] wanted to determine a shopper's interest based on that person’s physical behaviour in a store so that personalised offers and recommendations can be provided. However, activities like taking items, examining items or scanning for items may require computationally expensive analysis of video streams. Hence, Vildjiounaite et al. [4] proposed a low cost depth sensor based human tracking system that is able to track the trajectories of customers to help predict their shopping behaviour and their location in the store.

Kanda et al. [5] also determined a potential customer by retrieving information about their behaviour based on their walking trajectories as well as their speed.

By applying a clustering technique, they were able to program a robot to target services to potential customers efficiently.

From these works surveyed, it can be observed that the systems built uses video surveillance to determine the level of interest of a customer to a product by automatically determining the behaviour of the customer towards the product. These systems either have multiple cameras focusing on different areas of a store or tracks the movement of the customer in the store. The actions of the shopper will be tracked and observed and any moving person will be assumed to be a customer. However, this assumption may not be accurate since in a store, besides having customers walking around, there are also staff members that also moves around the store. These staff members could be restocking the products or attending to any enquiries by the customers. By not differentiating between customers and employees, the previous systems proposed may not be obtaining accurate readings to determine the popularity of a product.

Hence, in this paper, an in-store monitoring system that is able to differentiate between a staff member and a customer is proposed. Instead of having multiple cameras to observe individual products, this proposed system will use a two camera system to monitor the whole store to minimize on cost and complexity since there will not be a need to observe multiple cameras or making sure that the data obtained does not overlap from one location to another.

The proposed in store monitoring system is further elaborated in Section 2 while Section 3 discusses about the results obtained. Finally, the conclusion is presented in Section 4.

2. Proposed In-Store Monitoring System

Currently, most systems proposed does not clearly address how the customers are separated out from the employees. It is just assumed that the people interacting with the products are the customers, which can make the data collected inaccurate. Therefore, in this paper, a video surveillance based system which combines face detection, identity classification, face recognition and human tracking is proposed. The surveillance system will be based on two cameras, one to capture the face of a person entering the store and one more to have a bird’s eye view of the whole store from the ceiling. The first camera will be used to classify the identity of the person walking in and after that, based on his identity, the second camera will be used to track this person’s movement in the store.

Face detection is a computer vision technique to detect human faces in digital images while facial recognition is based on feature extraction on a face from facial components such as eyes, nose, and mouth. This will assist in the differentiation of a customer from an employee. Video based human tracking works by tracking a moving object against a non-changing background. This will be used to track only the customers to help determine their shopping behaviour.

To show how the proposed system operates, a flowchart of the overall configuration of the system developed is illustrated in Fig. 1. At the beginning of the system operation, the input video frame is first captured from the pre-setup camera located at the entrance of the simulated environment. It is assumed that the store has only one entrance and everyone will come in through there. The video frame is then fed into the face detection algorithm to produce a preprocessed output frame with face detected. The output frame will be compared with the training images in the database and goes through the identity classification process. The identification process determines whether the subject of the output frame is an employee or customer. If the subject is determined as an employee, the system will update the attendance of the recognized employee for the day and will not track his movements in the store. If the subject fails to match with any of the faces in the employee database, then this person will be classified as a customer and the human tracking algorithm will start working. In the counting algorithm, one or more region of interest (ROI) is set up to output the number of people passing through a region and the staying time in that region.

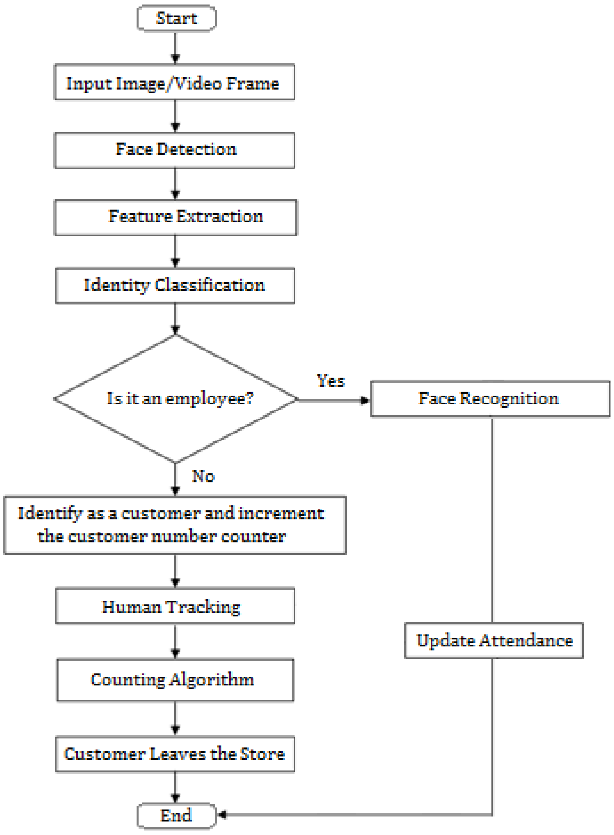


Fig. 1. Proposed in-store monitoring system.

2.1. Face detection

The flowchart of the face detection process is shown in Fig. 2. Since the system will be using a video camera as the input source, the system will capture the image by taking a snapshot of a video frame as shown in Fig. 3(a). From the captured image, integral image can be created by computing the value of each pixel in the rectangle feature from Haar filter [6]. Next, the integral image is then passed through a series of weak classifiers based on the Adaboost training algorithm [6]. Adaboost is an efficient feature selector where each round of weak classifier selects optimal features given by the previous selected features.

Cascaded classifier is then used to focus on the region of interest only to speed up the detection process. The output of the cascaded classifier is a successfully detected face [6]. Once the position of the detected face is determined, a rectangle box or bounding box is drawn to locate the face in the output image using the information such as the upper left corner pixel coordinate, [x, y], width and height of the box as shown in Fig. 3(b). To pre-process the output image, it will need to be converted to grayscale and normalized by resizing it to the size of 100 by 100 as shown in Figs. 3(c) and (d). The image conversion is needed to reduce the computation speed while image normalization is required so that same sized images are used for image matching at the face recognition stage.

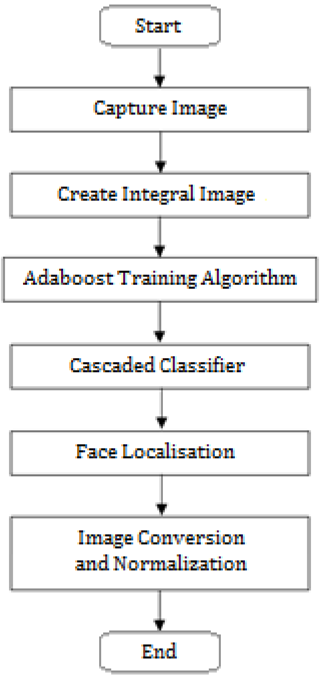


Fig. 2. Viola-Jones [6] face detection algorithm flowchart.

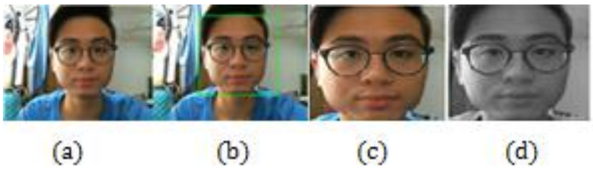


Fig. 3. (a) Captured image. (b) Output image with bounding box. (c) Cropped image. (d) Cropped image converted to grayscale.

2.2. Identity classification and face recognition

After the detection of a face, the next step is to determine if that person is an employee of the store or a customer. This is achieved by comparing the input image with a face database of the employees. Only if that person is an employee, then the person will be identified using face recognition.

First, a set of face images of known faces is collected to create an employee face database. Next, to create eigenfaces, the average of faces in the face space, N where each face images is converted into vectors, Γn, is calculated [7]. Each face's difference from the average face is computed and formed a covariance matrix (C) for the database [7]. From the result of the computation, eigenfaces can be obtained from the eigenvectors of C. PCA is used to reduce the large dimensional space and to find the top eigenfaces [7].

After an employee database is setup, the input face image is fed into the face recognition algorithm and that image will undergo feature extraction. Conventionally, the information from the feature extraction is used to compare with the top eigenfaces in the database and by using the nearest neighbour classifier system, the identity of that person in the input image can be determined. However, since the input image can either be an employee or a customer while the database that the face recognition system is using only consists of employee's faces, the nearest neighbour classifier will be wrong when the input image is a customer. Therefore, a threshold is also added to the face recognition system to help determine if the nearest neighbour is an employee or customer. The nearest neighbour value should be much lower for someone in the database compared to someone who is not. Therefore, by adding this threshold, the system will be able to determine if the input image is an employee or customer. A person with the nearest neighbour value above the threshold will be considered as a customer while those who are below the threshold value will be considered as an employee and their identity will then be determined by choosing the database image with the nearest neighbour value. Figure 4(a) shows an example of an input image that was determined as a customer while Fig. 4(b) shows an example of an input image that was determined as an employee. For input images determined as an employee, the identity of the employee is also shown based on the nearest neighbour value obtained from the database.

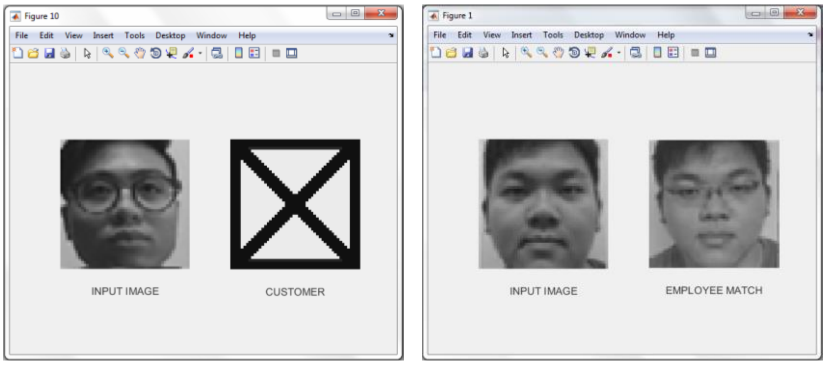


Figure 10
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE
CUSTOMER
Figure 1
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE
EMPLOYEE MATCH

(a) **(b)**

Fig. 4. (a) Input image of a customer. (b) Input image of an employee.

2.3. Human tracking

According to the flowchart shown in Fig. 5, the proposed human tracking stage is divided into two parts, one is the detection of moving objects and the other is motion-based tracking. Background subtraction algorithm based on Gaussian mixture models [8] is used for moving object detection. This technique can detect foreground objects from a video frame from the stationary camera by comparing it to a background model to determine whether individual pixels belong to the background or foreground. Foreground is represented by pixels with a value of 1 whereas background is represented by pixels with a value of 0. To eliminate noise, morphological operation is implemented in the foreground mask and the holes are filled to form a blob. Connected pixel or so called blob in the foreground mask is considered as a moving object. Since the video is captured indoor, moving objects are most likely to be humans walking around. Lastly, blob analysis is applied on the blobs and their characteristic such as area, centroid, and bounding box are obtained.

Motion-based tracking involves detection of moving objects in each frame and associating the detection of the same object over time. Kalman filter [9] is introduced to deal with the motion of the object's track. It predicts the track's location in each frame and corrects the estimate of the object's location with new detection. In any given frame, some of the detections are assigned to tracks but other detections and tracks still remain unassigned. The assigned tracks will be updated according to corresponding detection while unassigned tracks remain invisible and unassigned detection is given a new track. If the unassigned tracks remain invisible for a specific number of frame which exceeds a threshold, the tracks will be deleted.

For the counting algorithm, a region of interest (ROI) is set up for assigned tracks to trigger the counter. Once the counter is triggered, the label of the assigned track and the number of frames that assigned track stays within the ROI are recorded.

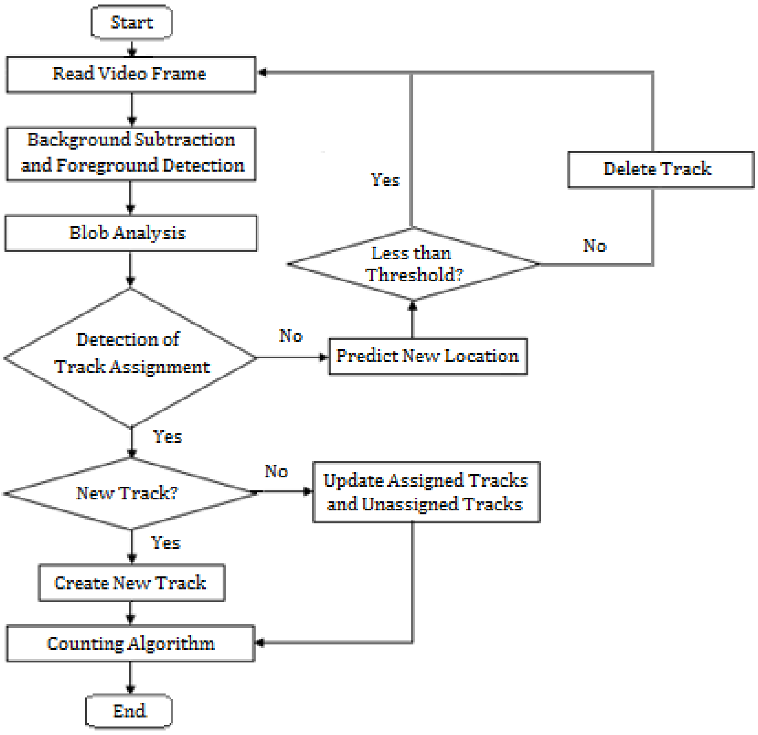


Fig. 5. Human tracking algorithm.

3. Results and Discussion

In this section, the results for each of the three aforementioned algorithms are presented. Before testing these algorithms, a database of known faces is created. This is to mimic having a database of employee’s faces. After that, the face detection, identity classification and recognition algorithms are evaluated by first taking a frame grab from a video and detecting the face in each frame. Then, the identity classification and face recognition algorithm are tested by observing how high the classification and recognition rates are. Lastly, the human tracking algorithm is tested with a recorded video from a webcam. The results of detecting moving people and counting the number of people moving in and out a region of interest (ROI) are analysed and discussed.

3.1. Face detection results

A set of raw images of 11 subjects each having 4 images taken by camera are shown in Fig. 6(a). Each image in the series represents a mixture of subjects without spectacles, subjects with spectacles, subjects with different facial expression such as a smile and subjects with slight head tilt. After the images are fed into the face detection algorithm, the output images with green bounding boxes showing the detected faces are displayed in Fig. 6(b). All images have the face detected successfully and this gives a 100% face detection rate. This shows Viola-Jones [6] face detection algorithm is robust and has a high detection rate. To produce the final training images for the database, the detected face must be cropped from the output images shown in Fig. 6(b), converted to grayscale and normalized to the size of 100 by 100.

3.2. Identity classification and face recognition results

To test the identity classification and face recognition algorithm, five people were chosen as employees while six people were chosen as customers. Three images for each employee were used for training. The training was done with 15 training images and the selected images not used for training were used to test the algorithm. Firstly, an average face is calculated by subtracting the mean of all training images from themselves. Then, the eigenvectors of the correlation matrix is calculated and the weight of each training images are stored into the eigenspace. Now the selected image is also subtracted from the mean and compared to find the nearest match in the eigenspace. Figure 7 shows the training images used, Fig. 8 shows the testing images used and Fig. 9 shows the results obtained from the identity classification and face recognition algorithm. The proposed system was able to correctly differentiate an employee from a customer by checking the nearest neighbor value obtained from the eigenspace. Those with values above the set threshold were considered as customers while those with values below the threshold were considered as employees. If the input image was an employee, then the database image with the minimum distance in the feature space with the input image will be considered as the identity of that person. Since the result gave a 100% correct identity classification and face recognition results, this shows that the proposed system is able to differentiate between the customers and employees as well as being able to determine the identity of that employee.

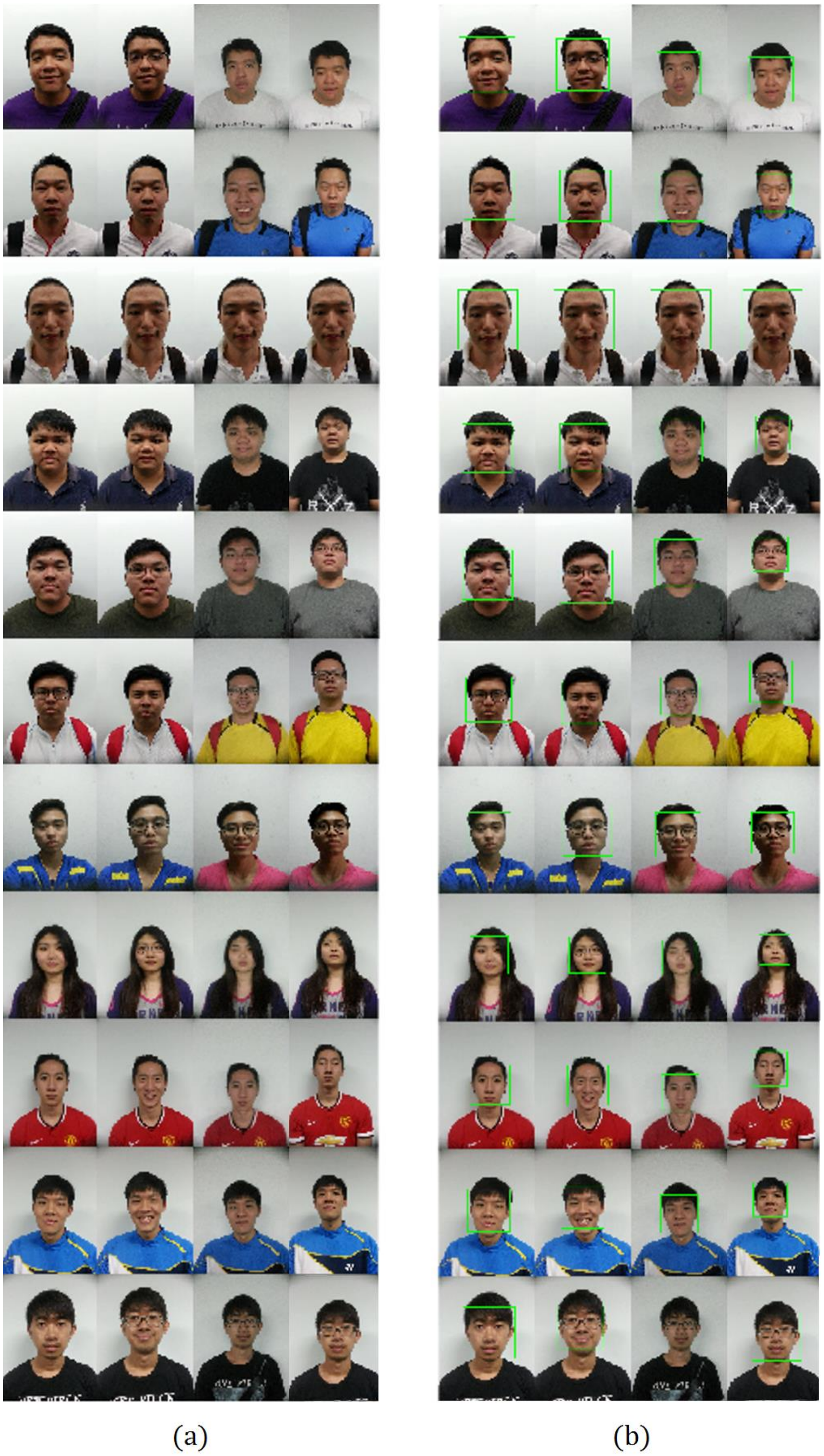


Fig. 6. (a) Input images. (b) Output images with detected faces.

Fig. 7. Images used for the employee database.

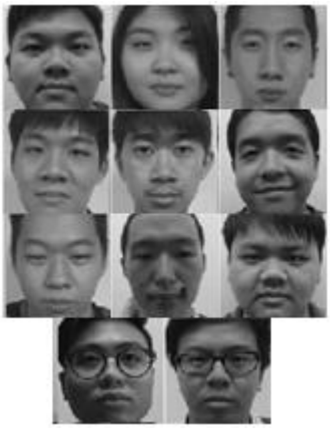


Fig. 8. Input images used for testing.

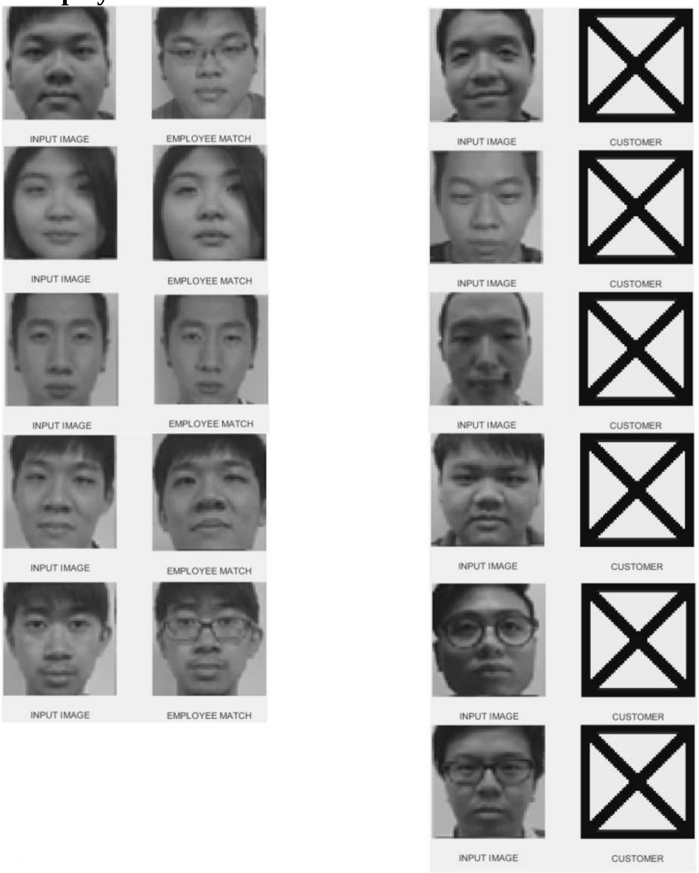


Fig. 9. Results of identity classification and face recognition algorithm tested with all subjects.

3.3. Human tracking algorithm results

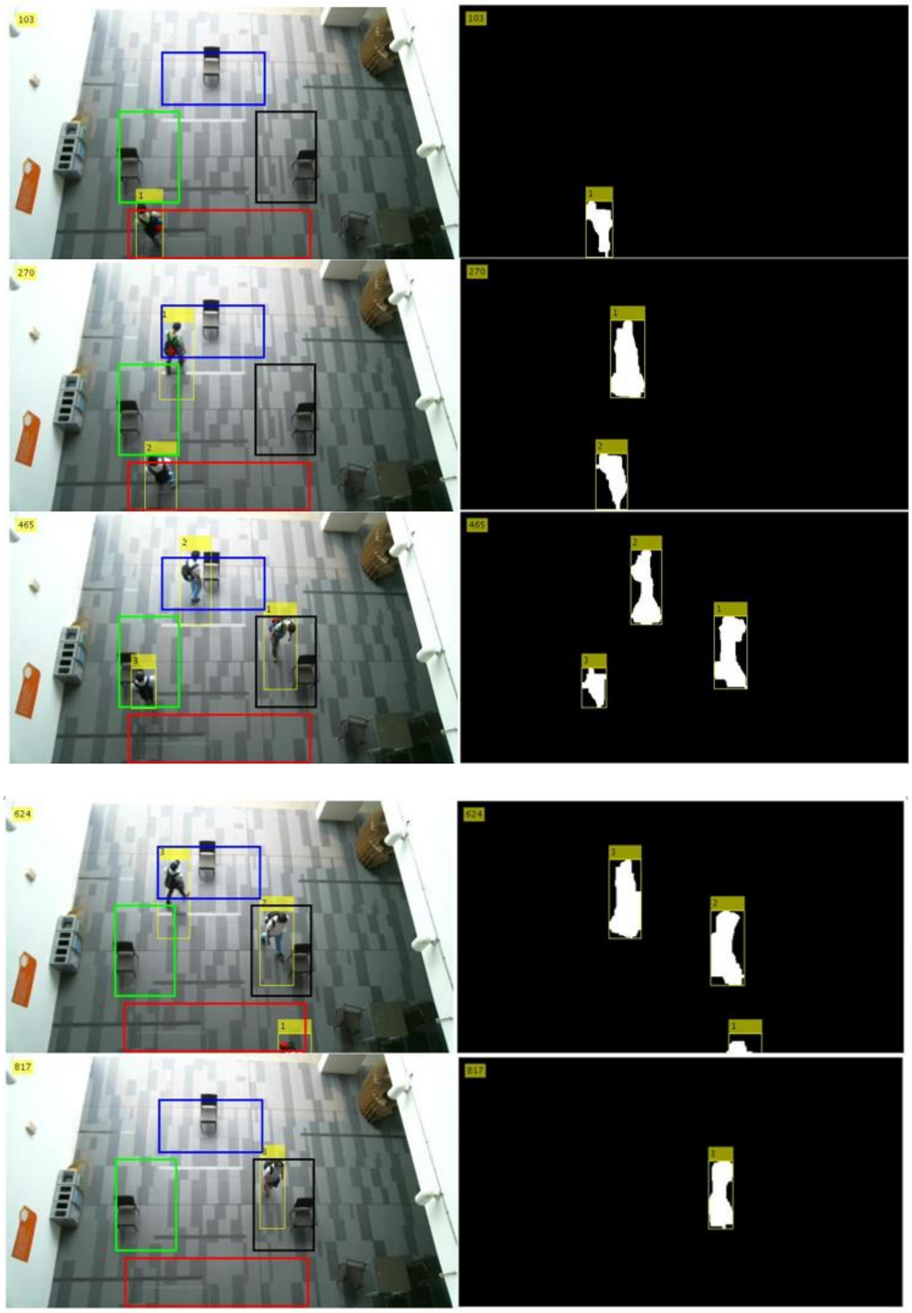
The human tracking algorithm is tested using a video with the size of 640 by 360. The video is captured at a location with fairly good illuminance and a few pedestrians as subjects of interest. The video is viewed as original frame and foreground mask as shown in Figs. 10(a) and (b) respectively. This video is used to test the tracking algorithm with and without the employee excluding algorithm. For better tracking result, the motion of the subject in this video scene is planned accordingly. 4 different regions of interest (ROI) in the form of bounding boxes are drawn on the original video frame as shown in Fig. 10. Red, green, blue, and black bounding boxes indicate ROI 1, ROI 2, ROI 3, and ROI 4 respectively. At the start of the video, the first person will appear from the bottom of the video frame, entering bounding box from ROI 1, to ROI 2, to ROI 3, then to ROI 4, lastly back to ROI 1, and then leaves the video scene at the bottom of the frame. This motion is followed by second subject and third subject.

Figure 10 shows the video frames of different subject entering different ROI. At frame 103, subject labelled number 1 was entering the video scene and was in ROI 1. At frame 270, Subject 2 appears from ROI 1 and was about to enter ROI 2 and Subject 1 was in ROI 3. At frame 465, Subject 3 was in ROI 2, Subject 2 was in ROI 3 and Subject 1 was in ROI 4. At frame 624, Subject 1 was leaving the scene through ROI 1, Subject 2 was in ROI 4 and Subject 3 was in ROI 3. At frame 817, Subject 1 and Subject 2 have left the scene and Subject 3 reached ROI 4. Table 1 shows the data collected from the counting algorithm without employee excluding algorithm which shows the staying time of the particular subject in a particular ROI in term of number of frames.

To determine the accuracy of the proposed human tracking algorithm, the results obtained was compared with data obtained by visually inspecting the same video. The amount of time in seconds each subject stays in an ROI was recorded in Table 2. Since the video observed was 15 frames per seconds, hence the number of frames each subject was in each ROI was also obtained.

By comparing the results from Tables 1 and 2, it was observed that the number of frames each subject spend in each ROI was quite similar between visual observation and the proposed human tracking algorithm. The highest difference in number of frames is 16 while the lowest difference is 1 frame. Therefore, this shows that the proposed algorithm is quite accurate and can be used to automatically determine the amount of time spent in an area of a shop.

Next, the employee excluding algorithm was tested by setting Subject 1 to be an employee. The results obtained was recorded in Table 3. By comparing the result from Table 1 and Table 3, the data collected from Subject 1 in Table 1 is not shown in Table 3. However, the count of the number of frames for Subject 2 and Subject 3 does not change in both tables. This shows that the employee excluding algorithm works accordingly without affecting the frame counting of other subjects.



(a) (b)

Fig. 10. Tracking results of the human tracking algorithm with the test video. (a) Original frame. (b) Foreground mask. (Frame number: 103, 270, 465, 624, and 817 respectively)

Table 1. Counting algorithm results of test video without employee excluding algorithm.

ROI	Subject	Count
1	1	52
2	1	86
1	2	41
2	2	114
3	1	104
1	3	48
2	3	132
3	2	85
4	1	124
4	2	121
3	3	131
4	3	124

Table 2. Visual observation of subjects in the test video.

ROI	Subject	Time (seconds)	No. of Frames
1	1	3	45
1	2	2	30
1	3	3	45
2	1	6	90
2	2	8	120
2	3	8	120
3	1	8	120
3	2	5	75
3	3	8	120
4	1	8	120
4	2	8	120
4	3	8	120

Table 3. Counting algorithm results of test video with Subject 1 excluded.

ROI	Subject	Count
1	2	41
2	2	114
1	3	48
2	3	132
3	2	85
4	2	121
3	3	131
4	3	124

The video taken is to simulate the situation in a store where the proposed in-store monitoring system is to be installed. Having a first point of contact in RO1 is similar to having only one entrance to the store. Everyone going into the store will need to go through RO1 first. They would be identified using face detection and facial recognition to determine if they are an employee or not. If they are an employee, the system will choose not to track them, as simulated above by Subject 1. Also, since the employee has been recognised, his identity can be used in an attendance tracking system for the store. If the person is identified as not an employee, the movement of this person will be tracked throughout the store until this person leaves the store, again through RO1. At the end of the day, the frequency of each area visited can be tallied up to determine which location is highly frequented by customers. From the results obtained, it shows that the system proposed was able to identify the location that are frequently visited by customers. This can help a store identify which product are popular and influence their stocking decisions.

4. Conclusions

In conclusion, the proposed system integrated Viola-Jones algorithm [6] and eigenfaces method [7] using principal component analysis as the face detection and face recognition algorithms respectively to differentiate between employees and customers. Furthermore, background subtraction and foreground detection technique incorporating Kalman filter [9] are used in the human tracking algorithm to provide relevant data for customer’s shopping behaviour study. In this project, Viola-Jones algorithm gives a 100% face detection rate, providing fast and accurate input to the face recognition part of the system. The identity classification and face recognition algorithm using eigenfaces is able to help classify a person walking in as an employee or a customer. The human tracking algorithm is then able to track multiple moving customers accurately to help determine which area in a store is more frequented. The novelty of this project is that the customer tracking feature of this project is able to differentiate between customers and employees in a store to provide better data collection for more accurate customer’s shopping behaviour studies.

References

1. Popa, M.C.; Rothkrantz, L.J.M.; Yang, Z.; Wiggers, P.; Braspenning, R. and Shan, C. (2010). Analysis of shopping behavior based on surveillance system, *2010 IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC'2010)*, Istanbul, Turkey.
2. Sicre, R.; and Nicolas, H. (2010). Human behaviour analysis and event recognition at a point of sale, *Fourth Pacific-Rim Symposium on Image and Video Technology (PSIVT)*, Singapore.
3. Liciotti, D.; Zingaretti, P.; and Placidi, V. (2014). An automatic analysis of shoppers behaviour using a distributed RGB-D cameras system, *2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, Senigallia, Italy.
4. Vildjiounaite, E.; Mäkelä, S.; Järvinen, S.; Keränen, T.; and Kyllönen, V. (2014). Predicting consumers' locations in dynamic environments via 3D sensor-based tracking, *Eighth International Conference on Next Generation Mobile Apps, Services and Technologies (NGMAST)*, Oxford, United Kingdom.
5. Kanda, T.; Glas, D.F.; Shiomi, M.; Ishiguro, H.; and Hagita, N. (2008). Who will be the customer? : a social robot that anticipates people's behavior from their trajectories, *UbiComp '08 Proceedings of the 10th international conference on Ubiquitous Computing*, Seoul, Korea, 380-389.

6. Viola, P.; and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features, *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Kauai, Hawaii, USA.
7. Turk, M.A.; and Pentland, A.P. (1991). Face recognition using eigenfaces, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, USA.
8. Stauffer, C.; and Grimson, W.E.L. (1999). Adaptive background mixture models for real-time tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, USA.
9. Welch G.; and Bishop, G. (2006). An introduction to the Kalman filter, *Technical Report: TR95-041*, Department of Computer Science, University of North Carolina, USA.