HKIE THE HONG KONG INSTITUTION OF ENGINEERS 香港工程師學會

A robust and efficient automated product recognition system based on scenario design methodology and computer vision techniques

Wai Tong Fung and Kin Man Lam

Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China

ABSTRACT

Automated product recognition systems should minimise the online product verification time while keeping product recognition robust to human interventions, including misplaced products, rotated products and empty shelves. A predefined template of each product is generated offline, and scale and location invariant binary local features are employed for recognition. Three types of Sampling Bounding Mask (SBM) are defined and combined with RGB/LUV colour histograms, and matched with a planar level of a shelf by using the Earth Mover's Distance, thereby reducing the computation time by 50%. Eight comprehensive product views are used to define five product scenarios on a shelf. The SBMs can be used to accurately distinguish the scenarios in a region at a level of shelf. By identifying four scenarios where a shopkeeper does not need to take any action to immediately reallocate misplaced products on shelves, the computation time is reduced by 80%. Six invariant keypoints and feature computation methods are compared for the recognition of product shelf items in different scenarios. The proposed approach speeds up the process by 200% to 300% for product matching. By integrating SBM view detection, the accuracy of product recognition in the front-view with rotations ranges from 90% to 95% using six different feature descriptors.

KEYWORDS Computer vision; product recognition; product counting; planogram matching; online product verification; unnamed store
CONTACT Wai Tong Fung torinfung8@gmail.com
Received 8 February 2020

1. Introduction

Numerous consumer products are being purchased or removed from shelves. Part of an online store-front management system is to monitor various abnormal scenarios of products in real time, and a long computation time to recognise all products one by one is not a practical approach. Indeed, many customers will pick up products from and put products back on shelves every day. Probably one customer will take away products, i.e. put them into a shopping basket, less than 5% of the time and put products back, i.e. examine and put them back on the product shelf over 95% of the time. The "put back on" action causes the orientation of a product to become slightly changed, such that the front view of most items is not perfectly shown. The shopkeeper may not need to take any action to fix the product's orientation so that its front position is shown at any time.

1.1. Shelf scenarios analysis

Most of the recent research focuses on identifying three scenarios, which are defined such that all products are placed on a shelf with their front view showing: put-it-right, misplacement without identification, and out-of-stock by roughly counting the quantity. Generally, they do not consider other realistic scenarios, such as slightly rotated, slightly shifted, innermost or inverted placement of items, as shown in Figures 1(a) and 1(b).

(a)

(b)

Figure 1. (a) One of the retail shelves shows the real scenarios of inverted items, rotated items, innermost items and misplaced items but less items showing their front; and (b) the other retail shelf shows the real scenarios of more rotated items, innermost items, misplaced items and out-of-stock items.

1.2. Product brand recognition

Brand recognition was proposed in some recent papers (Kleban et al., 2008; Varol et al., 2014; Cleveland et al., 2017), which cannot distinguish isomorphism products (Tonioni and Stefano, 2017), even if the products are of the same brand and same shape, but with different subtypes, as shown in Figure 2. Previous research fails to consider the variable orientation information of each product. On the shelves, products of the same brand possibly have different shapes, sizes, colours, types and patterns of label. All of these same or similar products almost appear adjacent to each other on the same level of shelf.

Figure 2. Verification of isomorphism products.

1.3. New proposed research methodology of shelf-scanning analysis

In order to reduce the computation time of online product verification, the processing of thousands of local invariant feature matchings for each product recognition with shelf images by computer vision can be reduced. In this paper, a new high-speed shelf-scanning method is presented to match the segmented spatial graph of each product on a level of shelf. The spatial graph-based detection is the summation of the colour histogram of all products and matching by Earth Mover's Distance (EMD) (Grauman and Darrel, 2004) on each level of shelf. The new proposed methodology of different Sampling Bounding Masks (SBMs) integrated with EMD can indicate the rough localisation of abnormal scenarios of the product. It is not necessary to process the accurate product recognition of all products on shelves and much more manpower can be saved by not taking any action. Most product shelf management focuses on frontal product recognition while our work focuses on complete 360 degree product recognition to handle misplacement by customers. The multiple views covering 360 degrees requires lots of training samples with different viewpoints. As deep learning requires many training samples per view, the high number of products and views as well as the high number of training samples per view make deep learning expensive in the process of obtaining and training sufficient training samples. Our approach only requires one to two samples per view in order to achieve high accuracy which is not easy via deep learning.

2. Related work

Future robotic systems can be built to enable cooperative distribution stock monitoring, cooperative RFID inventory control, cooperative robot formations and human-robot assistance (Katić and Rodić, 2010). Advanced computer vision and RFID approaches are the main recognition techniques in intelligent shopping malls. A shopping service robot that has product recognition and grabbing ability has been proposed (Cheng et al., 2017). The SURF feature method is used for generating feature points and the flood fill algorithm (Asundi and Wensen, 1998) is used to find the grabbing plane of the product (Cheng et al., 2017). Only feasible results are presented in this work. A mobile robot designed to facilitate automatic detection of Shelf Out of Stock (SOOS) situations was proposed (Paolanti, 2017a). Using visual and textual features extracted from Deep Convolutional Neural Networks (Paolanti, 2017b), F1-scores of 0.8 and 0.74 were obtained, respectively. An approach for finding missing items by RFID was proposed by Ferreira (2010). The remote sensing of a large number of product shelf items with a large number of antennas poses interference problems, which makes accurate detection of products difficult. Typical performances from four RFID approaches are found to give 60% precision and 98% recall. A robotic system with thousands of sensors is placed on a shelf whereby the setup and maintenance costs are very high.

2.1. Repeated product layout matching approaches

Other work on planogram compliance is based on the detection of recurring patterns where recurring patterns are detected by the split and merge approach (Liu et al., 2016). In their work, product layout is extracted from an input image by means of unsupervised recurring pattern detection and matched via graph matching. There is a comparison between template-based matching and their recurring pattern method. The template-based method yields accuracies of 40% to 95% in different product recognition settings. The recurring pattern method gives accuracies of 81% to 96% in different product recognition settings. Recurring pattern detection is based on graph matching of the front view to improve the detection speed and there is no need for product template images for training sets. The bounding box is merged with the recurring pattern of each product on the shelf. These two methodologies were mainly proposed to define the boundary grouping of the repeated product patterns, which can increase the recognition accuracy and improve the recognition speed dramatically.

Other recent work proposes to solve the issue of product recognition in store shelves as a sub-graph isomorphism problem (Tonioni and Stefano, 2017) whereby each product is auto-localised by eight cardinal directions. Several detectors and descriptors are constructed based on different local invariant features from the front view of products on a shelf. The feature extraction methods evaluated include SIFT, SURF, ORB, AKAZE, BRISK, etc. As the local features of the above algorithms cannot provide for high accuracy, the graph-isomorphism is added to correct the errors from the local features by integrating the geometry and layout of similar items on the product shelf. Better results of the local invariant features are obtained from BRISK with 78% precision which can be raised to 98% precision using graph isomorphism. BRISK + SURF has 66% precision which can be raised to 97% precision. Sub-graph isomorphism can also be added to the current approach to further refine accuracy though many of the local features in the current approach already exceed their standard feature-based approach.

By using spatial continuity of brand and metric size, the co-occurrence of the products and the adjacency relations between the products can be used to improve product recognition (Baz et al., 2017). The Hidden Markov Model (HMM), SVM and conditional random fields (CRF) are compared. Integrating context into CRF improves the accuracy to 79.86% compared to the accuracy of SVM of 68.45% and accuracy context-integrated HMM of 78.02%.

In related work on planogram compliance (Saran et al., 2015), the Hausdorff metric for occupancy computation in regard to product shelf images is proposed. The occupancy detection has an accuracy of 36/36, while product identification has an accuracy of 23/25.

2.2. Local invariant feature extraction approaches

There is grocery detection for the visually impaired (Winlock et al., 2010) which uses a multiclass naive-Bayes classifier, SURF descriptors, windows probability distributions and colour histograms. For the 52 items tested, they found 17 easy items with 100% accuracy, eight moderate items with 1 to 100 false positives and difficult-to-classify hard items with more than 100 false positives on up to one thousand images per item.

Recent work on object recognition has used local binary invariant features combined with colour histograms (Phan et al., 2013). However, in supermarket product recognition, products from a product line with different flavours, e.g. different variants of a particular brand of shampoo, often vary slightly in wording and have a uniform colour design which makes the variants of a product difficult to differentiate using a colour histogram. Thus, invariant feature-based methods are the main focus of this work.

Four feature selection methods were selected in this work. First the scale-invariant feature transform (SIFT) was selected (Lowe, 2004). In this method, first there is a scale-space extrema detection using the difference of Gaussians which is an approximation to Laplacian of Gaussian. Local extrema in scales are selected as potential keypoints. The keypoint candidates are then further refined using contrast selection. Orientation is then assigned to keypoints. The keypoint descriptor is a 16x16 neighbourhood which is then divided into 16 subblocks with a size of 4x4 using eight bin orientation histogram results in 128 bin feature values. The second selected method is the speeded-up robust features (Bay et al., 2008). In SURF the LoG is approximated with a Box filter. The feature description in SURF uses wavelet responses in the horizontal and vertical directions. Both 64 features and 128 features are available. The third selected method is the binary robust independent elementary features BRIEF method (Calonder et al., 2010). It is based on the keypoint location method of centre surround extremas for real-time feature detection and matching censure (Agrawal, 2008). The fourth selected method is the ORB (Rublee et al., 2011) which is an efficient alternative to SIFT and SURF. It is based on the high-speed corner detection FAST keypoint detector (Rosten and Drummond, 2006; Rosten et al., 2010) with a binary descriptor based on BRIEF.

Another novel keypoint detector and feature selector is BRISK (Leutenegger et al., 2011). BRISK has lower computational cost than other algorithms. Comparisons with other feature algorithms using large image datasets reveal that it has very good performance accuracy. The high speed results from a scale-space FAST-based detector combined with a bit-string descriptor from intensity comparisons retrieved by dedicated sampling of each keypoint neighbourhood. The KAZE is a novel multiscale 2D feature detection and description algorithm in nonlinear scale space (Alcantarilla et al., 2012). KAZE detects and describes 2D features by using nonlinear diffusion filtering. Blurring is locally adaptive to enable noise reduction and retain object boundaries better than other algorithms. The KAZE algorithm is further speeded up in the improved AKAZE algorithm (Alcantarilla et al., 2013). A Fast Explicit Diffusion (FED) numerical method embedded in a pyramidal framework improves the speed of feature detection in nonlinear scale spaces. A Modified-Local Difference Binary (M-LDB) descriptor is also adopted to exploit gradient information from the nonlinear scale space. The resulting algorithm is scale and rotation invariant and has low storage requirements.

In some feature matching work, brute force matching is sometimes suggested. In this work, the matching method for the features is the Fast Library for Approximate Nearest Neighbours (FLANN) (Suju and Jose, 2017). The FLANN performs approximation that is sufficient for most practical applications with orders of magnitude faster than the exact brute-force method (Muja and Lowe, 2014). After the matches are found, they are further filtered by the Lowe's ratio test using a ratio of 0.7. The minimum number of matches is often set to be 10 in many applications.

Other work on planogram maintenance is the heuristic approach for product occurrences (Emanuele et al., 2014). The algorithm involves morphological operation, template matching and histogram comparison. The user must manually select the most significant label of the product to find in the shelf image and thus the approach is not fully automatic.

There is work on recognising groceries using both images from the Web as well as store-capture images (Merler et al., 2007). Three recognition/detection algorithms are used including object colour histogram matching, SIFT matching and boosted Haar-like features. Different products are found to have better performance in each of the above algorithms without any particular algorithm having clear superiority.

In the automated system for semantic object labelling work (Cleveland, 2017), a template-based approach together with the four feature extractors, SIFT, DSIFT, HOG and SURF are evaluated. A multi-scale SIFT descriptor at multiple scales is proposed as the detection camera is located on a robot and the robot can navigate freely which leads to a large variation in distance between the camera on the robot and the shelf product. Furthermore, the DSIFT (Bosch et al., 2007) algorithm is also found to be favourable, where a higher resolution SIFT feature with smaller steps than standard is used to generate more keypoints. In that work, the DSIFT has the highest performance of 60-70% precision and recall, SIFT has 55-60% precision and recall, and SURF and histogram of gradients (HOG) have precision and recall of below 50%. In the current design, high-resolution cameras are mounted in fixed locations on a shelf which enables very high quality and a high quantity of keypoints to be generated; thus, such multiple scale SIFT and DSIFT descriptors are not necessary.

In earlier work on retail product recognition proposed (Varol, 2015), the support vector machines (SVM) with a Gaussian radial basis function and various feature detectors including HSV colour histogram and SIFT are proposed. The performance of the algorithms is 69% precision and 90% recall. With constraints of height and width, the performance can be boosted to 81% precision and 94% recall, respectively.

In other work using image analytics to monitor retail store shelves, the histogram of gradients and the bag of words approach are proposed (Marder et al., 2015). The accuracy rate of their approach is 87.4%. There are 4% of products that are not detected at all and 8.6% of products are misclassified.

2.3. Neural network approaches

Deep learning has also been used for product recognition (Goldman and Goldberger, 2017). The deep learning network has been assisted by conditional random field (CRF) and deep class embedding (Chu and Cai, 2018). Comparison with convolutional neural network (CNN) (Krizhevsky et al., 2012), recurrent neural network (RNN) (Schuster and Paliwal, 1997), long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997; Huang et al., 2015) and other CRF methods are evaluated. Most deep learning and CRF-based methods give an error of 13-15%. The proposed CRF+deep class embedding deep learning method has an error rate of 11%. In other work using convolutional neural networks (Jingsong et al., 2016), 35 categories are selected and for each 100 images are taken and split into training 80 images and 20 testing images. CaffeNet has an accuracy of 67.8%, AlexNet has an accuracy of 76.3% and the Finetune approach has an accuracy of 80.4%.

In another deep learning approach to planogram compliance, convolutional neural networks (CNN) are evaluated with three different training dataset designs (Chong et al., 2016). Their work only focused on the recognition of eight types of products: cans, tap drills, axes, etc., without the actual brand or product label. The accuracy ranges from 36-100% for one type of training, and 56-100% for another type of training.

Advanced robotics and computer vision used for shelf analytics, deep learning and computer vision techniques are compared in (Agnihotram et al., 2017). Using computer vision techniques of bounding box detection, noise removal, histogram of oriented gradients, block normalisations and SVM yields an accuracy of 71%. Using a deep learning approach yields an accuracy of 83%.

Work to improve the estimation of product amount is proposed (Higa and Iwamoto, 2018). Methods to calculate changes of products on shelves including “product taken” and “product replenished” are developed. The measurement of success is taken as an error margin of 0.1 which allows error in detecting one object out of ten. On different shelves, the success rate is from 51.4-100%.

Current approaches range from heuristic computer vision-based techniques to scale invariant keypoint features and deep learning approaches. However, almost all of these works are based on direct front view approaches and ignore cases of side view and back view. Furthermore, a lot of the approaches focus on box-like objects which have planar surfaces that are easier to handle using affine transformation and classical computer vision techniques. In the case of complex-shaped objects, for example, shampoo which is often a non-circular bottle without much symmetry, such objects can produce features that change rapidly and are difficult to be handled by invariant features. Most of the feature-based computer vision approaches only have limited accuracy which could be problematic in real-world applications. The deep learning approaches have higher accuracy but require a lot of training samples. Generating enough training samples for the various scenarios in a real-world supermarket can be very time-consuming and expensive.

In summary, a new design for the APRS which has the following novel characteristics is desired. First, the APRS must be robust to human interactions. To tackle this, a scenario design methodology is established to analyse scenarios and handle scenarios with high accuracy and efficiency. Second, the APRS must be robust to product placement. For example, the training product placed on a middle shelf should be recognisable without any extra training and be recognisable when placed on a top shelf.

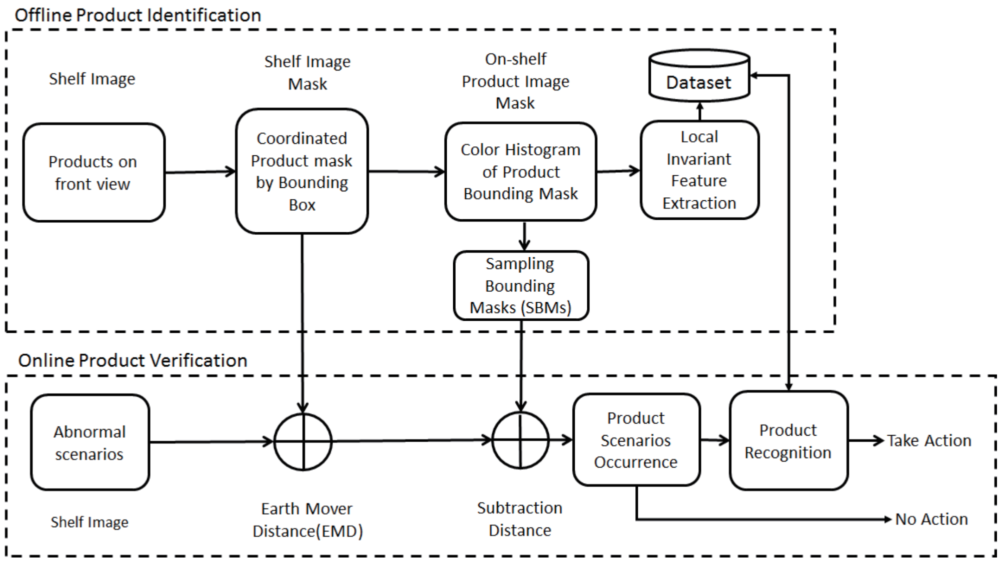


Figure 3. Online verification and offline identification of products on shelves.

Thus, a small training effort and expense is needed in setting up the APRS because the training sample size is much reduced. Third, the APRS must be able to handle complex object recognition. In this work, we focused on non-box-like, non-circular products such as shampoo as box-like objects have been studied widely. Fourth, APRS must have high accuracy. This is easy to understand as mistakes in the system are costly to the business and may often take lots of human effort to correct. Lastly, the APRS should have high efficiency in recognising products and scenarios.

3. Proposed methodologies

In the scenario-based methodology, shelf management is divided into several components in order to maximise its accuracy and efficiency. The system is composed of a fast online product recognition module and a slower offline product recognition module. The online product verification model makes use of fast colour histogram computation distances between a product ground-truth template and the shelf scene to be verified. EMD is selected as a robust measure, which is least affected by lighting and shifts in the position of products. This module can output the response of whether an intervening action should be taken.

In Figure 3, the offline product identification consists of training products from the front view and optionally other views of the product. Bounding box and template masks are generated from each template. The Sampling Bounding Masks (SBMs) are applied to both training and testing images before colour histogram computation. The product recognition is based on the colour histogram distances and the local invariant feature distances. The new algorithm proposed for online product verification can identify the rough localisation of variant orientation products by EMD on each level of shelf. Several SBMs are defined to distinguish the abnormal scenarios of product orientation. Some scenarios cannot be distinguished need to be processed by product recognition. Finally, the computation time and manual labour time can be significantly reduced.

Online shelf verification of the abnormal scenarios is shown in Figure 4, and the scenarios are defined as follows:

1) Corrected item
2) Misplaced item
3) Out-of-stock item

In addition to the above listed scenarios, there are other possible scenarios such as:

4) Slightly rotated items (Total of six rotated views)
5) Inverted item
6) Innermost item
7) Slightly shifted item

For 2) misplaced items, it is also proposed to find the true product item label by product recognition. This indicator can be shown to the shopkeeper who places the product in the correct location with the correct price label on its shelf. For 4) slightly rotated items, it is also proposed to use eight rotated views consisting of front and back views of the item in Figure 5. The rotated scenarios are examples of scenarios that do not require any immediate action by the shopkeeper. Finally, only items that are processed by product recognition or found to be in the out-of-stock scenario will be notified to the shopkeeper to take action.

Figure 4. Investigate the computation time and recognition accuracy of 40 spaces of product item location on a shelf simultaneously.

Figure 5. Definition of eight product views consisting of front and inverted views, from left to right; Front view (front-B), front rotated 45° anti-clockwise (front-A), rotated 90° (side-A), back rotated 45° clockwise (back-C), back view (back-B), back rotated 45° anti-clockwise (back-A), Rotated 270° (side-B) and front rotated 45°clockwise (front-C).

3.1. Offline product identification on a shelf

After all the products are placed on shelves, offline product identification is processed for a total of 40 products from shelf images captured by high-resolution motion cameras as shown in Figure 6. Product images with background removed and the associated product masks are shown in Figure 7 and Figure 8, respectively. The initial training samples are processed to produce an accurate template by the following operations. 1) The morphological openings with standard structuring elements is applied (Rougherty, 1992). 2) Connected component labelling analysis with an 8-connectivity configuration is applied (Klaiber et al., 2015). Suitable threshold object sizes are selected to reflect actual product and noise objects from the background. Figure 8 shows the results after the connected component analysis.

Figure 6. Offline shelf image captured by a high-resolution motion camera.

Figure 7. Clear background colour and product with coordinates on a shelf.

Figure 8. Product masks.

In order to obtain fast detection of objects and their respective positions, a masked colour histogram approach is suggested. Due to the variations of lighting that have a huge effect on the colour histogram, various histogram methods are evaluated to minimise the lighting effect. In this work, the use of CIELUV colour space is proposed to minimise the effect of histogram variations due to lighting.

The CIELUV colour space is designed to be perceptually uniform and is calibrated from the CIEXYZ colour space. CIELUV colour space is calibrated and verified to adhere well to human visual perception. The original colour image is mapped to an LUV colour histogram. Figure 9 shows two sample images represented by UV components from the LUV colour histogram. The quantisation of the histogram uses 4-bit; thus, there are 16 levels in each of the LUV components, resulting in 4,096 (16x16x16) histogram bins.

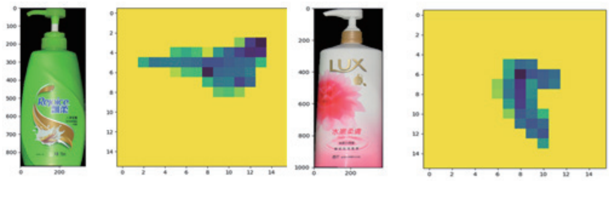


Figure 9. Pseudo colour of UV components represented as colour histogram.

3.2. Product recognition on shelves

Six algorithms are selected for feature extractions including SIFT, SURF, ORB, BRISK, KAZE, AKAZE. First scale-invariant feature transform (SIFT) is selected, and there is scale-space extrema detection using difference of Gaussians which is an approximation to Laplacian of Gaussian. Local extrema in scales are selected as potential keypoints. The keypoint candidates are then further refined by using contrast selection. Orientation is then assigned to keypoints. The keypoint descriptor is a 16x16 neighbourhood which is then divided into 16 sub-blocks with a size of 4x4. Using an 8-bin orientation histogram results in 128-feature values. The second selected method is the speeded-up robust features (SURF), and the LoG is approximated with a Box filter. The feature description in SURF uses Wavelet responses in the horizontal and vertical directions. Either 64-feature or 128-feature are available. The third selected method is the ORB (Rublee et al., 2011) which is an efficient alternative to SIFT and SURF. It is based on the high-speed corner detection FAST keypoint detector (Rosten and Drummond, 2006; Rosten et al., 2010) and a binary descriptor based on BRIEF. The fourth method is Brisk (Leutenegger et al., 2011). The fifth and sixth methods are KAZE and AKAZE (Alcantarilla et al., 2012; Alcantarilla et al., 2013). The matching method for the features is a Fast Library for Approximate Nearest Neighbours (FLANN) (Suju and Jose, 2017; Muja and Lowe, 2014). After the matched templates are found, they are further filtered by the Lowe's ratio test using a ratio of 0.7. The keypoints of the testing product image are compared with the keypoints of the training product image, and the maximum number of keypoint matches between the training image and the testing image are selected as the label of the testing image. To understand the performance of the different algorithms, the average number of keypoints is selected for each view and each algorithm of the testing image. The SURF algorithm has the biggest number of keypoints so the computation time is the slowest. The AKAZE algorithm has the smallest average number of keypoints. In terms of the different views of the product, the back view has the greatest number of keypoints as there are many text descriptions and patterns on the back. The smallest number of features is found on the side views. The number of keypoints is very important in considering the recognition speed of different scenarios.

3.3. Online fast product verification on shelves

In regard to supermarkets, most of the different scenarios are defined and mentioned in Figure 5. In order to solve the above problem, the 3D-LUV colour histogram, EMD distance measures and another novel sampling technique are introduced in this section. Figure 10 shows the high-resolution image by motion camera that illustrates the seven scenarios, as shown in Figure 4. The product template of the normal product shelf is obtained. The real-world images are captured and compared with the normal product template. We can see in Figure 11 that the product template defines masked regions which will be compared using multiple distances of the 3D-LUV colour histogram with EMD. The masked regions of each product obviously show all the abnormal scenarios mentioned in Figure 4. One example of the background is shown in the masked region while the product is missing, which is the 3) scenario item, as mentioned in Figure 4.

Figure 10. Online abnormal scenarios in the shelf image captured by a high-resolution motion camera.

Figure 11. Multiple distance comparison of offline template and online product within masked regions by EMD and a 3D-LUV colour histogram.

EMD is an advanced distance often formulated to compare two images. It can be interpreted as two image distributions which are two different ways of piling up a certain amount of dirt over the region (Rubner et al., 2000). Another application of EMD in computer science is the comparison of two greyscale images that may differ due to dithering, blurring or local deformations. In this case, the region is the image's domain, and the total amount of light (or ink) is the "dirt" to be rearranged. EMD is widely used in content-based image retrieval to compute distances between the colour histograms of two digital images. The EMD algorithm depends on a signature-generation process for each histogram. The algorithm below is the python code for generating the EMD signature, sig, from the RGB 3D histograms, hist, with bins equal to the number of bins in each colour component. The first axis of the signature is ordered as a projection of each of the colour spaces where bins2 equals to bins*bins. The total number of rows in the signature is bins*bins*bins. The second axis of the signature has four components. The histogram value is stored in the first component. The r, g, b values are stored in the second to fourth components respectively. The pseudocode for generating the EMD signature is as follows:

```
for r in range(0, bins):
   for g in range(0, bins):
       for b in range(0, bins):
          sig[r*bins2 + g*bins+b, 0]= hist[r,g,b]
          sig[r*bins2+g*bins+b, 1]= r
          sig[r*bins2+g*bins+b, 2]= g
          sig[r*bins2+g*bins+b, 3]= b
```

The real-world scenario is introduced in the next section. While the EMD distance is more distinct it is also much slower than the BHA distance and the COR distance.

3.3.1. New proposed Sampling Bounding Masks (SBMs)

The four SBMs with colour distances can be applied to localise all variant orientation products on each level of shelf at one time as shown in Figure 12. There is no need for any action to be taken among the scenarios of 1) corrected item, 5) Innermost item, 6) inverted item and 7) slightly shifted item because the products are still located in the right position with the right price label on the shelf. The 3) scenario is also known which is a background. At the end, only the 2) scenario is processed for product recognition and the product label will be identified. It is not necessary to recognise all products on a shelf even if they are different, the same or isomorphism products. This scenario is only reported to a shopkeeper who takes immediate action after receiving the indication alert via the Internet.

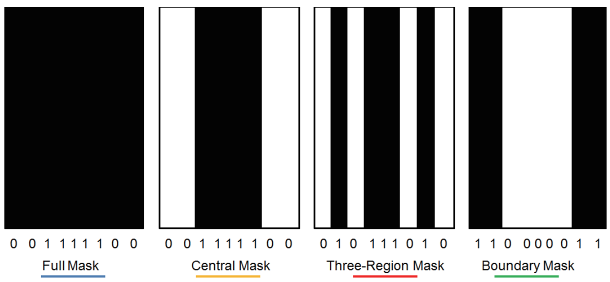


Figure 12. Proposed four SBMs of the product; Full mask [11111111] indicated by a blue line, central mask [00111100] indicated by an orange line, boundary mask [11000011] indicated by a green line and 3-region mask indicated by a red line.

In order to handle missing cases and rotated cases with high accuracy, a novel boundary mask approach is proposed to find the scenarios, as shown in Figure 4. The novel design is based on the colour constancy of most products in order to capture the geometry of the product and maximise the ability to detect the same colour product from different angles, as shown in Figure 5. Figure 12 shows different SBMs that can achieve better product differentiation for missing and rotated products. The SBM is applied to the product template mask using a Boolean AND operation. The position with logical ONE will be extracted for the 3D LUV colour histogram so that the computation time will be reduced by 50%.

4. Implementation and experiments

The accuracy performance of our algorithm on these datasets by offline product identification is provided in Figure 13(a) and online product verification is provided in Figure 13(b).

(a)

(b)

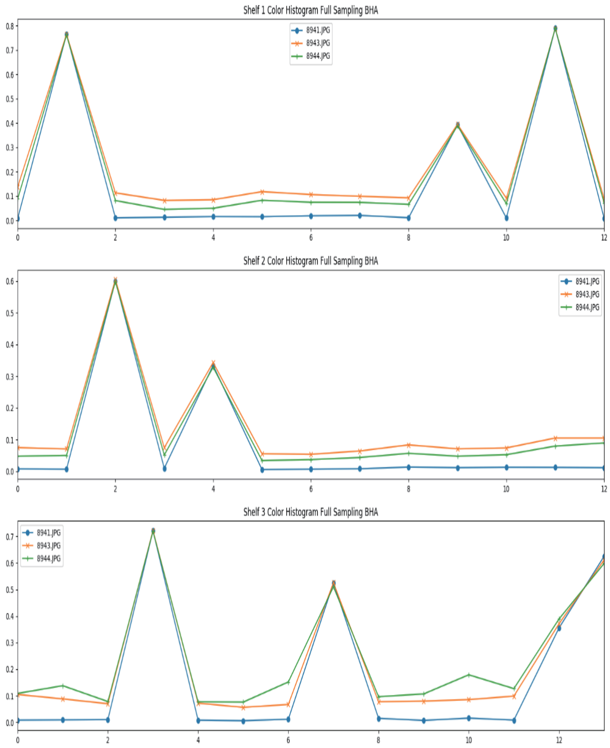
Figure 13 (a) Recognition accuracy of 40 products' shelf image after offline product identification; and (b) the scenario occurrences for online product verification of a shelf image.

4.1. Colour planar layout matching

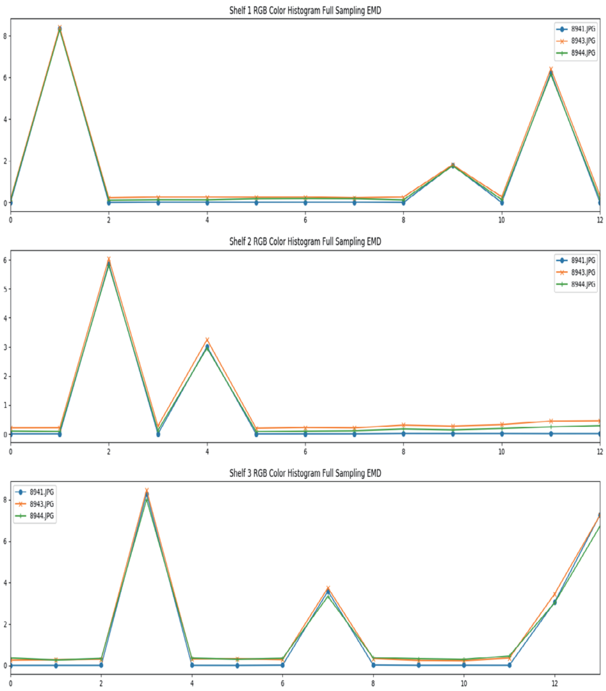
In order to compare two different colour histograms, conventional methods include the cosine distance, Bhattacharyya distance and correlation distance are adopted. The Bhattacharyya distance measures the similarity between two distributions. The correlation distance between two histograms is also widely used and has a maximum value of 1 and a minimum value of zero. Figure 14(a) shows the comparison of the colour histogram distance using the Bhattacharyya distance in LUV colour space. Figure 14(b) shows the comparison of the colour histogram distance using the Bhattacharyya distance in RGB colour space. Three images are compared with large lighting differences (see Figures A1, A2 and A3). The distances between identical objects and foreign objects are not as distinct as in the case of the Bhattacharyya distance. The distances of sample products under different lighting conditions only changes mostly less than 0.1 units while differences between different objects are 0.8 to 0.4 which shows that different objects are possible to be detected under such a scheme.

Figure 14(c) shows the comparison of the colour histogram distance using the correlation distance in LUV colour space. Figure 14(d) shows the comparison of the colour histogram distance using correlation distance in RGB colour space. Figure 14(e) and 14(f) shows the comparison result of the EMD of the colour histograms in LUV and RGB space, respectively. The EMD results are very distinct and foreign objects that differ from the template are associated with very big distances while objects that are the same as the template are represented by very small distances.

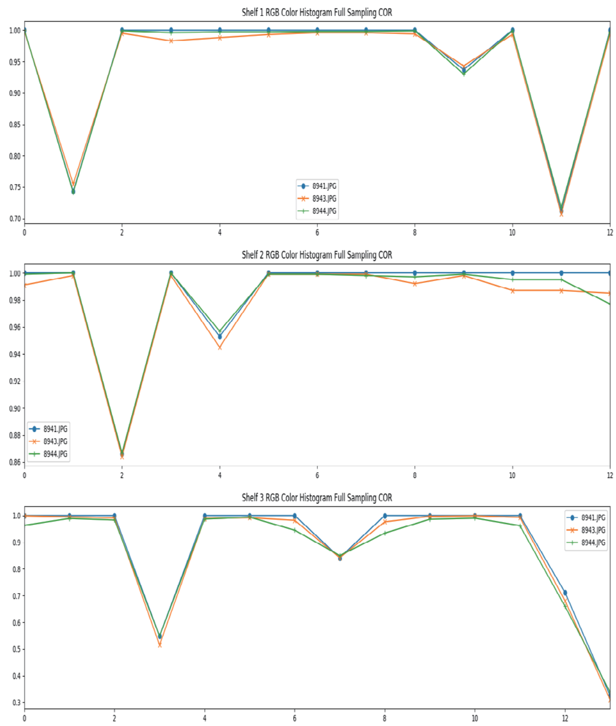
(a)



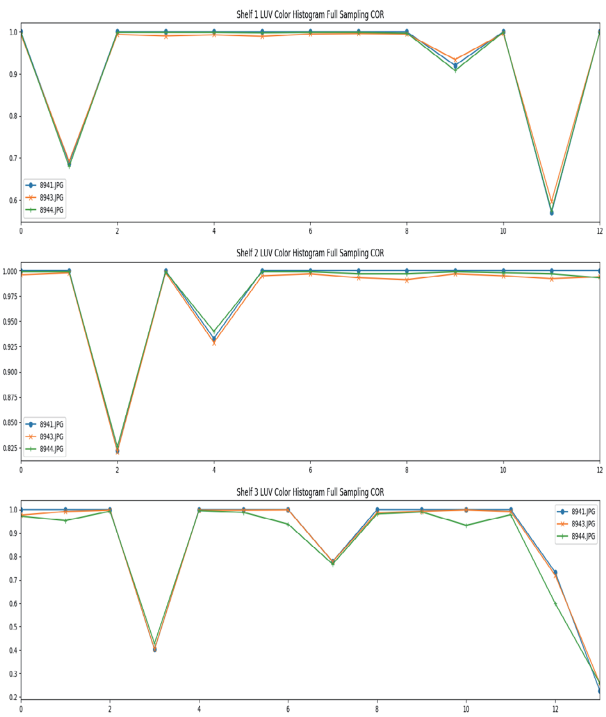
(b)


Shelf 1 RGB Color Histogram Full Sampling EMD
Shelf 2 RGB Color Histogram Full Sampling EMD
Shelf 3 RGB Color Histogram Full Sampling EMD
8941.JPG
8943.JPG
8944.JPG

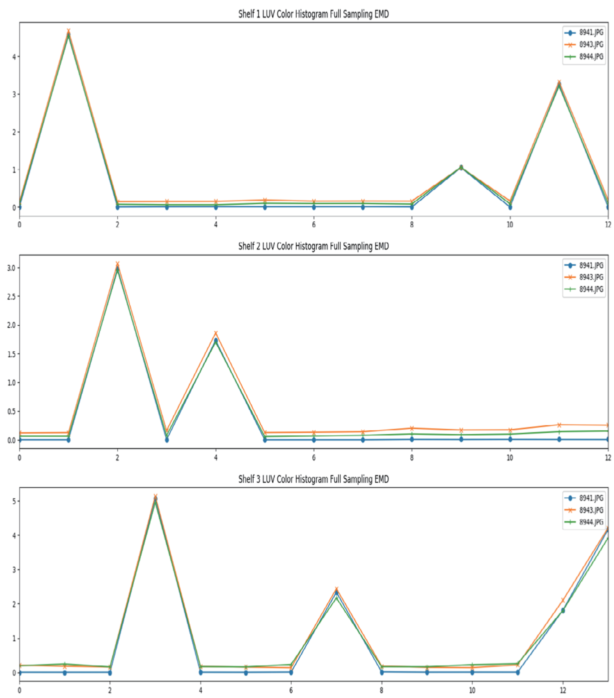
(d)


Shelf 1 RGB Color Histogram Full Sampling COR
Shelf 2 RGB Color Histogram Full Sampling COR
Shelf 3 RGB Color Histogram Full Sampling COR
8941.JPG
8943.JPG
8944.JPG

(c)


Shelf 1 LUV Color Histogram Full Sampling COR
Shelf 2 LUV Color Histogram Full Sampling COR
Shelf 3 LUV Color Histogram Full Sampling COR
8941.JPG
8943.JPG
8944.JPG

(e)


Shelf 1 LUV Color Histogram Full Sampling EMD
Shelf 2 LUV Color Histogram Full Sampling EMD
Shelf 3 LUV Color Histogram Full Sampling EMD
8941.JPG
8943.JPG
8944.JPG

(f)

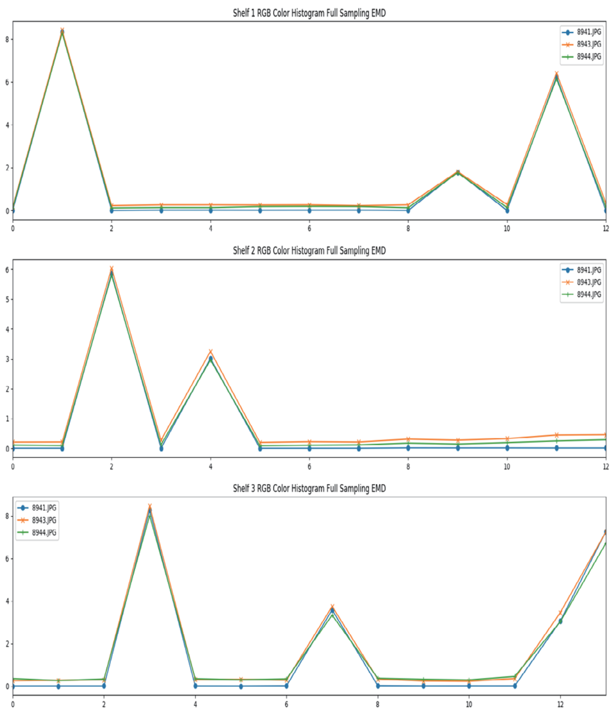


Figure 14. (a) Different lighting luminosity and dithering of the 3D-LUV colour histogram with the Bhattacharyya distance on shelves; (b) different lighting luminosity and dithering of the 3D-RGB colour histogram with the Bhattacharyya distance on shelves; (c) different lighting luminosity and dithering of the 3D-LUV colour histogram with correlation distance on shelves; (d) different lighting luminosity and dithering of the 3D-RGB colour histogram with correlation distance on shelves; (e) different lighting luminosity and dithering of the 3D-LUV colour histogram with EMD distance on shelves; and (f) different lighting luminosity and dithering of the 3D-RGB colour histogram with EMD distance on shelves.

In order to better evaluate the different colour spaces and the three different colour histograms comparison method, average normal object distances (ND) over three shelves and abnormal object average distances (AD) over three shelves are computed for each colour space and each distance measurement. Table 1(a) and Table 1(b) summarise the results. The three images Figures A1, A2 and A3 have large variations of lighting. The Bhattacharyya distance can differentiate abnormal objects from normal objects quite well and the LUV colour space performs better than the RGB. The correlation distance does not differentiate abnormal objects with distinct distances. Lastly, the EMD distance has the biggest distance between abnormal objects and normal objects. It is almost two times larger than the Bhattacharyya distance.

Table 1. (a) Average abnormal distance of three shelves in the LUV colour space, average normal distance of three shelves and ratio of the AD/ND; (b) Average abnormal distance of three shelves in the RGB colour space, average normal distance of three shelves and ratio of the AD/ND.

(a)

	Abnormal distance (AD)	Normal distance (ND)	AD/ND
EMD, A1	2.90	0.08	34.73
EMD, A2	2.98	0.24	12.23
EMD, A3	2.86	0.19	14.95
BHA, A1	0.57	0.028	20.44
BHA, A2	0.58	0.10	5.72
BHA, A3	0.58	0.093	6.19
COR, A1	0.68	0.99	0.69
COR, A2	0.69	0.99	0.70
COR, A3	0.67	0.98	0.68

(b)

	Abnormal distance (AD)	Normal distance (ND)	AD/ND
EMD, A1	5.18	0.13	40.46
EMD, A2	5.32	0.39	13.56
EMD, A3	5.09	0.31	16.22
BHA, A1	0.58	0.034	17.01
BHA, A2	0.59	0.12	5.02
BHA, A3	0.59	0.11	5.35
COR, A1	0.75	0.99	0.75
COR, A2	0.74	0.99	0.75
COR, A3	0.74	0.98	0.75

Figure 15 shows the result of applying the EMD distance to the SBM masked colour LUV histogram. The different object distance ranges from 3.5 to 6 units while identical object distances are very close to zero, showing a good signal to noise ratio.

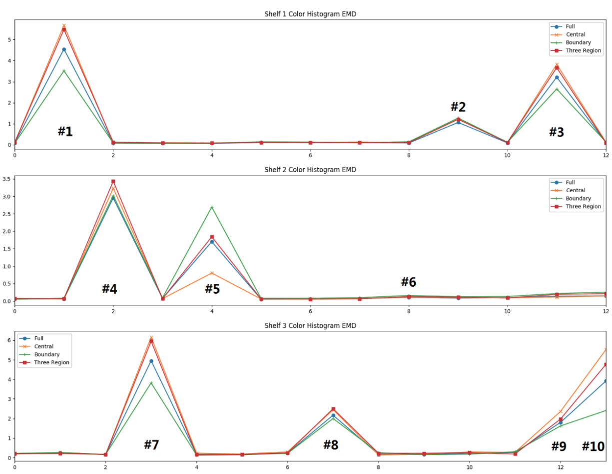


Figure 15. The measurement results obtained by Earth Mover Distance (EMD) between the offline and online 3D-LUV histogram per bounding box of each product on each level of shelf.

Table 2. The outcome of three Sampling Bounding Masks (SBMs).

#	Scenarios on shelf	Localisation	Total items	Identification on shelf level	Outcome by shelf level	Outcome by product
1	Corrected item (front view)	All others	30	EMD is almost equal to zero	Yes	N/A
2	Identification of misplaced items	#1, #3 and #4	3	Very close and higher EMD of 2 masks; Central and 3-region masks	No	Find actual items
3	Removed item then out-of-stock item	#7	1			Out-of-Stock
4	Rotated items					
	90° Rotated item	#5	1	The highest EMD of Boundary mask and the lowest EMD of Central mask	Yes	N/A
	45° Rotated item	#2	1	Very close EMD of 3 masks; Central, 3-region and Boundary masks	Yes	N/A
5	180° Inverted item	#8	1			
6	Innermost item	#6	1	Cannot distinguish	No	
7	Shifted items	#9 and #10	2	1st item; Very close and higher EMD of 3 masks; Central, 3-region and Boundary masks 2nd item; Very close and higher EMD of 2 masks; Central and 3-region masks	No	#10 shifted item #9 removed item
Total items on shelf			40			
Need further processing of items			6	#1,#3,#4,#7,#9 and #10		
No need for further processing of items			34			

Figure 16. A total of seven abnormal scenarios are proposed and shown in the testing image during online product verification.

In Table 2, the result of spatial distance measurement is shown to be able to find all possible variant product localisations on the shelf in Figure 16, where #1, #3, #4 and #7 are wrong or removed items, #2 and #8 are either 45°or 180° inverted items, #5 is a 90° inverted item, #6 is a non-detected item, #9 and #10 are shifted or removed items, and others are all corrected items. The full SBM is no longer to be applied and can save 50% of the computation time required for full shelf scanning.

The level classifier can identify no action for 34 items out of 40 items on the shelf, and only six items to be further processed for advanced product recognition. Regarding the computation time, 85% can be saved compared with advanced product recognition conducted one by one.

Table 3(a), 3(b) and 3(c) shows the average normal object and abnormal object distances computed using the different SBMs on three different sample Figure A1, A2 and A3. There is a large lighting variation in Figure A1, A2 and A3 as reflected in the large change in normal object distances. The central bounding mask is found to have a better signal to noise ratio compared to the full mask. The signal to noise ratio in all images remains good for the four different masks even under large variations of lighting.

Table 3. (a) Average abnormal distance of three shelves from Figure A1, sum of average normal distance of three shelves and ratio of the AD/ND; (b) average abnormal distance of three shelves from Figure A2, sum of average normal distance of three shelves and ratio of the AD/ND; and (c) average abnormal distance of three shelves from Figure A3, sum of average normal distance of three shelves and ratio of the ND/AD.

(a)

	Normal distance (ND)	Abnormal distance (AD)	ND/AD
Full	2.90	0.08	34.73
Central	3.58	0.09	37.49
Boundary	2.40	0.07	31.61
Three regions	3.39	0.09	36.52

(b)

	Normal distance (ND)	Abnormal distance (AD)	ND/AD
Full	2.93	0.24	12.04
Central	3.62	0.26	13.90
Boundary	2.42	0.27	9.122
Three regions	3.51	0.29	12.09

(c)

	Normal distance (ND)	Abnormal distance (AD)	ND/AD
Full	2.86	0.19	14.95
Central	3.56	0.31	16.63
Boundary	2.32	0.21	11.28
Three Regions	3.37	0.21	15.70

4.2. Product recognition and verification with SBMs

In this part, the second part of the SBM-based approach is discussed and analysed. Figure 13(a) shows one sample test image to be classified. There are totally 72 images each with 40 products giving a total of 2,880 product samples. Three main approaches for feature matching of products will be discussed and analysed in this section. There are three steps for product recognition: single front view of product recognition, multiple-view of product recognition and SBM-assisted multiple-view product recognition. For all of the three analysed approaches, only one front view of all products on a shelf is selected for training samples. One sample of the training image is shown in Figure 17. The location of the product in the testing shelf image is shown to be different from the training shelf image. This is to show the ability of algorithms to classify images even if the product is misplaced onto different shelves. The initial product training samples can be obtained from various sources including product web templates, Grabcut in Opencv and image processing software. The initial product training samples are also processed to produce an accurate template through the following operations: 1) The morphological openings with standard structuring elements is applied. and 2) Connected component labelling analysis with 8-connectivity configuration is applied.

Figure 17. Example of a training shelf image.

Each shelf image contains 38 products; using one training image from the front view, the total number of training product images is 38. The total number of test product images is 2,432. The labels of product view on a shelf are defined and shown in Figure 5.

Table 4. Average number of keypoints for each view and each algorithm.

	SIFT	SURF	ORB	BRISK	KAZE	AKAZE
back-A	923.76	1042.87	495.97	679.32	422.45	339.71
back-B	1528.53	1429.37	500.00	1190.50	572.32	531.42
back-C	548.79	790.34	471.11	377.00	308.08	233.68
front-B	567.42	927.61	499.68	692.13	548.13	444.03
front-A	374.50	651.03	474.53	423.26	371.92	258.05
front-C	526.95	785.21	494.76	590.29	455.61	346.66
side-A	195.79	383.92	355.45	145.26	193.11	105.47
side-B	223.29	390.24	346.79	160.39	200.39	103.97
overall mean	611.13	800.07	454.79	532.27	384.00	295.38

Figure 18 shows the accuracy of the six algorithms in recognising the different views in product images by only training from the front view. First, the front-B results are typical of normal or optimal supermarket scenarios where the products are nicely arranged. Very high accuracy is obtained in this approach where AKAZE exhibits the best performance, followed closely by ORB. However, the performance of other front views, such as front-A and front-C which are rotated 45 degrees drops dramatically. It is also noted that the back-B (back direct view) gives slightly better results than the side view because the back-direct view contains the same product logos that are recognised from the front-B training sample. This shows the limitations of these algorithms in terms of rotation invariance. It is quite obvious that such low accuracy will be a significant challenge in a real-world scenario and a more advanced approach is necessary.

Another important consideration is the speed of the invariant feature matching algorithm. The matching speed is dependent on the number of keypoints generated by the invariant feature algorithms. Table 4 shows the total number of keypoints generated by each algorithm for each view. The SURF algorithm generated the largest number of keypoints compared to other algorithms and was also found to be much slower than other algorithms in the matching stage. The AKAZE algorithm has the smallest number of keypoints and also runs two times faster than the SURF algorithm.

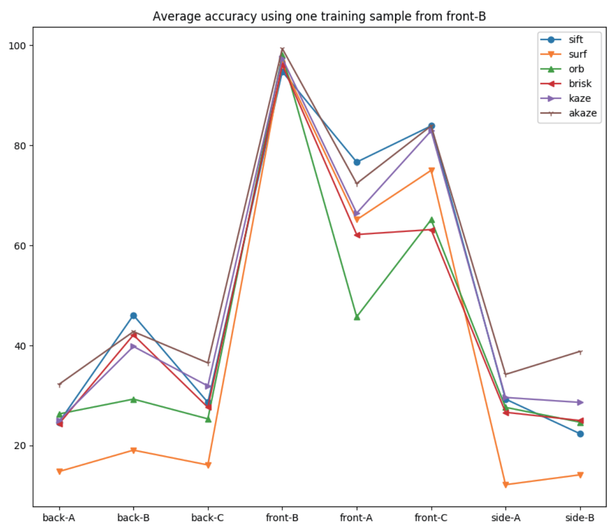


Figure 18. Average percentage accuracy for each algorithm using one image sample from front-B (front view) for training.

4.2.1. Multiple views of product recognition

One training product image for each view of each product was selected as a training product sample. Each shelf image contained 38 products, using one training image from each view, so the total number of training product images was 304. The total number of test product images was 2,432. Six algorithms were selected for feature extractions, including SIFT, SURF, ORB, BRISK, KAZE, AKAZE.

The result of accuracy comparison is shown in Figure 19. The algorithms that gave the highest accuracies are the KAZE and AKAZE algorithms, followed by the ORB algorithm. In terms of accuracy of the different views, the back view had the highest accuracy, followed by front view. The side view had the lowest accuracy. Furthermore, considering both the accuracy and computation time, the algorithms of KAZE and AKAZE are recommended for advanced product recognition.

Our work scenario was similar to the highest unconstrained product recognition rate attained of 81% using the feature method of BRISK+SURF. In our approach for frontal view using BRISK or SURF, the accuracy was both above 91% and the SIFT approach had an accuracy of 97%.

The side view scenario had the lowest accuracy due to the following two facts. First, our dataset focuses more on difficult products with a non-rectangular shape. Second, the side view is often smaller and less descriptive of the product than the front or back view. However, our system still manages an accuracy of between 58% and 85% for the side view scenario.

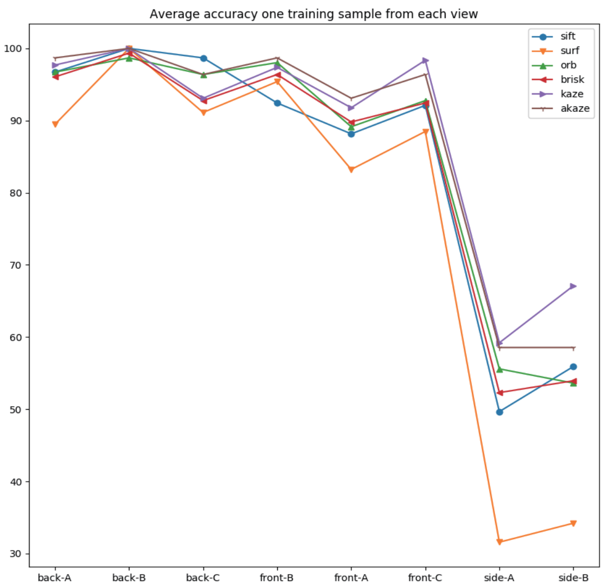


Figure 19. Average percentage accuracy for each view and each algorithm.

4.2.2. SBM-assisted multiple views of product recognition

As there are many possible views and thus many training samples for each product, a good way to speed up the processing and accuracy is to first determine the view and then use the feature detection and matching algorithm only on a fixed view. The SBM method provides one way to determine the possible view to speed up and improve the accuracy of feature matching. Figure 20 shows the percentage of greater accuracy by which each algorithm is improved for a reduction in computation time of 50%.

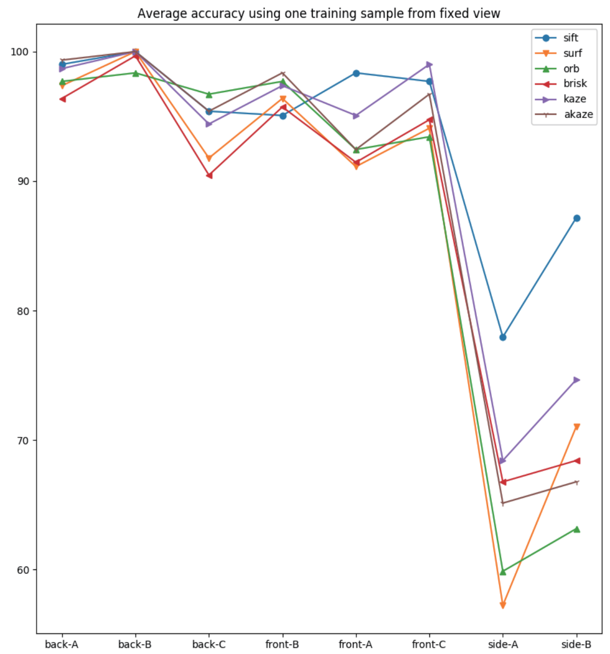


Figure 20. Average percentage accuracy after integrating view information into each fixed view and each algorithm.

5. Conclusion

In this paper, a novel approach for the development of a robust and efficient Automated Product Recognition System (APRS) is proposed and analysed. The robustness of our method to human interactions and product placement position on a shelf is due to the design of several effective modules. The proposed Sampling Bounding Mask (SBM) for online product verification has been demonstrated to be efficient even if the products are placed at different orientations on a shelf. The colour histogram based on different colour spaces and EMD are shown to eliminate illumination variations, as well as a certain degree of positional shift of products. Detailed experiments and analysis show that traditional invariant-based feature recognition methods have much lower accuracy and computation time when products are subjected to human interaction in various scenarios. The spatial layout of the scene of a level of a shelf is matched to the shelf-level template by EMD for online product detection and verification on the shelf. A total of 12 scenarios, which represent most of the orientations of a product on a shelf, are pre-defined. Only one scenario, where an item is misplaced, needs to be processed by local invariant feature detection for product recognition. Thus, about 80% of the computation time can be saved in product recognition. The recognition accuracy can be increased by up to 90%, when the orientations of a product are determined by a scenario-based approach before applying the six invariant feature recognition algorithms including SIFT, SURF, ORB, BRISK, KAZE, AKAZE. In practice, the proposed algorithm can achieve an accuracy of 80% under abnormal product scenarios, where the shopkeeper need not take any action.

Lastly, the APRS based on a designed scenario can achieve high accuracy and efficiency in regard to item recognition on shelves, with only a little training needed.

Appendixes

Figure A1. Abnormal scenario image 1.

Figure A2. Abnormal scenario image 2.

Figure A3. Abnormal scenario image 3.

Notes on contributors

Dr Wai Tong Fung received his B.Eng. (Hons) in Computer Engineering from the City University of Hong Kong, his M.Sc. in Information Engineering from The Chinese University of Hong Kong, and his Doctor of Engineering from The Hong Kong Polytechnic University. He has held various research and engineering positions in the industry and is currently an entrepreneur. His professional service includes: President of the Hong Kong Association for the Advancement of Science and Technology, Deputy Director of the Institute of Enterprise Management Innovation of the Chinese Academy of Management Sciences, Vice Chairman of the Hong Kong Federation of Invention and Innovation, Visiting Professor at the School of Business of the Chinese Academy of Management Sciences, and member of the Hong Kong and Macau Advisory Committee of the China Guangzhou Science and Technology Association.

Prof Kin Man Lam received his Associateship in Electronic Engineering with distinction from The Hong Kong Polytechnic University in 1986, his M.Sc. degree in communication engineering from the Department of Electrical Engineering, Imperial College of Science, Technology and Medicine, London, U.K., in 1987, and his Ph.D. degree from the Department of Electrical Engineering, University of Sydney, Australia, in 1996. Since 1996, he has been with The Hong Kong Polytechnic University, Hong Kong. He is currently a full professor in the Department of Electronic and Information Engineering. He is a Senior Member of the IEEE.

References

[1] Agnihotram G, Vepakomma N, Trivedi S, Laha S, Isaacs N, Khatravath S, and Kumar R (2017). Combination of Advanced Robotics and Computer Vision for Shelf Analytics in a Retail Store. In *2017 International Conference on Information Technology (ICIT)* pp. 119-124.

[2] Agrawal M, Konolige K and Blas M (2008). Censure: Center surround extremas for realtime feature detection and matching, In *European Conference on Computer Vision*, Springer, Berlin, Heidelberg, pp. 102-115.

[3] Alcantarilla P., Bartoli A., and Davison A (2012). Kaze features. In *Computer Vision–ECCV 2012*, Springer, pp. 214–227.

[4] Alcantarilla P, Nuevo J and Bartoli A. (2013). Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, In *British Machine Vision Conference (BMVC)*, Bristol, UK.

[5] Asundi A and Wensen Z (1998). Fast phase-unwrapping algorithm based on a gray-scale mask and flood fill, *Applied optics*, 37(23), pp. 5416-5420.

[6] Bay H, Ess A, Tuytelaars T and Van Gool L (2008). Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3), pp. 346-359.

[7] Baz I, Yörük E. and Çetin M (2017). Retail product recognition with a graphical shelf model, *25th Signal Processing and Communications Applications Conference (SIU)*, pp. 1 – 4.

[8] Bosch A, Zisserman A. and Munoz X(2007). Image classification using random forests and ferns, in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2007, pp. 1–8.

[9] Calonder M, Lepetit V, Strecha C, and Fua P (2010). Brief: Binary robust independent elementary features, In *European conference on computer vision*, Springer, Berlin, Heidelberg, pp. 778-792.

[10] Chaves T, Buchmann L, and Böhm K (2010). Finding misplaced items in retail by clustering RFID data, In *Proceedings of the 13th International Conference on Extending Database Technology*, pp. 501-512.

[11] Cheng C, Chen C, Liang J, Tsai T, Liu, C, and Li T. (2017). Design and implementation of prototype service robot for shopping in a supermarket. In *2017 International Conference on Advanced Robotics and Intelligent Systems (ARIS)*, pp. 46-51.

[12] Chu W and Cai D (2018). Deep feature based contextual model for object detection. *Neurocomputing*, 275, pp.1035-1042.

[13] Cleveland J, Thakur D, Dames P, Phillips C, Kientz T, Daniilidis K and Kumar V (2017). Automated system for semantic object labeling with soft-object recognition and dynamic programming segmentation. *IEEE Transactions on Automation Science and Engineering*, 14(2), pp. 820-833.

[14] Chong T, Bustan I and Wee M (2016) Deep Learning Approach to Planogram Compliance in Retail Stores, machine learning report, Stanford University.

[15] Emanuele F, Contigiani M and Ribichini G (2014). A heuristic approach to evaluate occurrences of products for the planogram maintenance, *2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA)*.

[16] Goldman E and Goldberger J (2017). Large-Scale Classification of Structured Objects using a CRF with Deep Class Embedding, *arXiv preprint arXiv:1705.07420*.

[17] Grauman K and Darrel T (2004). Fast contour matching using approximate earth mover's distance, *Proceedings of CVPR 2004*.

[18] Higa K, and Iwamoto K (2018). Robust Estimation of Product Amount on Store Shelves from a Surveillance Camera for Improving On-Shelf Availability, In *2018 IEEE International Conference on Imaging Systems and Techniques (IST)* pp. 1-6.

[19] Hochreiter S. and Schmidhuber J (1997). Long short-term memory, Neural computation, vol. 9, no. 8, pp. 1735–1780.

[20] Huang Z, Xu W and Yu K (2015). Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint *arXiv:1508.01991*.

[21] Katić D and Rodić A (2010). Intelligent multi robot systems for contemporary shopping malls. *IEEE 8th International Symposium on Intelligent Systems and Informatics*, pp. 109 – 113.

[22] Klaiber M Bailey D, Baroud Y and Simon S (2015). A Resource-Efficient Hardware Architecture for Connected Components Analysis, *IEEE Transactions on Circuits and Systems for Video Technology*, 26 (7): pp.1334–1349.

[23] Kleban J, Xie X and Ma W (2008). Spatial pyramid mining for logo detection in natural scenes, *IEEE International Conference on Multimedia and Expo* pp. 1077 – 1080.

[24] Krizhevsky A, Sutskever I and Hinton G (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processings NIPS, pp. 1097-1105.

[25] Leutenegger S, Chli M, and Siegwart R (2011). BRISK: Binary robust invariant scalable keypoints. In *2011 IEEE international conference on computer vision (ICCV)*, pp. 2548-2555.

[26] Li J, Wang X, and Su H (2016). Supermarket commodity identification using convolutional neural networks, In *2016 2nd International Conference on Cloud Computing and Internet of Things (CCIOT)*, pp. 115-119.

[27] Liu S, Li W, Davis S, Ritz C and Tian H (2016). Planogram Compliance Checking Based on Detection of Recurring Patterns, *IEEE MultiMedia*, Volume: 23 , Issue: 2, pp. 54 – 63.

[28] Lowe D (2004). Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, 60(2), pp. 91-110.

[29] Marder M, Harary S, Ribak A, Tzur Y, Alpert S and Tzadok A (2015). Using image analytics to monitor retail store shelves, *IBM Journal of Research and Development*, 59(2/3).

[30] Merler M, Galleguillos C and Belongie S (2007). Recognizing groceries in situ using in vitro training data. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8.

[31] Muja M, and Lowe. D (2014). Scalable nearest neighbor algorithms for high dimensional data. *IEEE transactions on pattern analysis and machine intelligence*, 36(11), pp. 2227-2240.

[32] Paolanti M, Sturari M, Mancini A, Zingaretti P, and Frontoni E (2017a). Mobile robot for retail surveying and inventory using visual and textual analysis of monocular pictures based on deep learning, In *2017 European Conference on Mobile Robots (ECMR)*, pp. 1-6.

[33] Paolanti M, Kaiser C, Schallner R, Frontoni E, and Zingaretti P (2017b). Visual and textual sentiment analysis of brand-related social media pictures using deep convolutional neural networks, In *International Conference on Image Analysis and Processing*, Springer, Cham, pp. 402-413.

[34] Phan D, Oh C, Kim S, Na I, and Lee C (2013). Object recognition by combining binary local invariant features and color histogram, In *Pattern Recognition (ACPR) 2nd IAPR Asian Conference on* pp. 466-470.

[35] Rosten E, and Drummond T (2006). Machine learning for high-speed corner detection, In *European conference on computer vision*, Springer, Berlin, Heidelberg, pp. 430-443.

[36] Rosten E, Porter R, and Drummond T (2010). Faster and better: A machine learning approach to corner detection, *IEEE transactions on pattern analysis and machine intelligence*, 32(1), pp. 105-119.

[37] Rougherty E (1992). An Introduction to Morphological image processing, Society of Photo Optical.

[38] Rublee E, Rabaud V, Konolige K, and Bradski G (2011). ORB: An efficient alternative to SIFT or SURF. In Computer Vision (ICCV), *2011 IEEE international conference on*, pp. 2564-2571.

[39] Rubner Y, Tomasi C and Guibas L (2000). The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2), pp. 99-121.

[40] Saran A, Hassan E, and Maurya A (2015). Robust visual analysis for planogram compliance problem, In *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, pp. 576-579.

[41] Schuster M. and Paliwal K., Bidirectional recurrent neural networks, *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[42] Suju D., and Jose H (2017). FLANN: Fast approximate nearest neighbor search algorithm for elucidating human-wildlife conflicts in forest areas,In *Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*.

[43] Tonioni and Di Stefano (2017). *Product recognition in store shelves as a sub-graph isomorphism problem*. University of Bologna, Department of Computer Science and Engineering (DISI).

[44] Varol G, Kuzu RS and Akgül YS (2014). Product placement detection based on image processing *22nd Signal Processing and Communications Applications Conference (SIU)*, pp. 1031 – 1034.

[45] Varol G, and Kuzu R. (2015). Toward retail product recognition on grocery shelves. In *Sixth International Conference on Graphic and Image Processing (ICGIP 2014)*,Vol. 9443.

[46] Winlock T, Christiansen E, and Belongie S (2010). Toward real-time grocery detection for the visually impaired, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops* pp. 49-56.