Visual-optical positioning and configuration of non-identified electronic pricing tags on retail shelves

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

Firstly, a new positioning methodology for electronic pricing tags (EPTs) is proposed, performed by a Visual-Optical Mobile Module (VOMM) on the shelf. The module is composed of a monocular mobile camera and an infrared laser emitter built into a precise X-Y-Z angle divider driven by subdivision control of a stepping motor for Simultaneous Localization and Mapping (SLAM). The inter-tag distance between two EPTs can be down to a minimum of 2 mm. Secondly, the battery-operated EPT consists of an infrared receiver and a low-cost MCU with LC-display. All EPTs can be positioned and automatically assigned to a specific ID. Hence, neither Radio Frequency Identification (RFID) nor a beacon is needed to tag the products. All the off-the-shelf products can be freely placed anywhere on the shelves. Thirdly, all individual EPTs can be auto-configured by a VOMM via Line-of-Sight (LoS) communication. The initial database simply records product identification and description, regular price, promotion price and sale period. Finally, the rotatable VOMM is installed on the ceiling, and a 3D floor plan can be generated to show the coordinates of products and EPTs. Shelves and zones can be generated automatically by using visual recognition and the information from the database according to the position of all EPTs on the shelves.

KEYWORDS Computer vision; optical communication; object positioning; object configuration; non-identified electronic pricing tag; retail shelf management; unmanned store

CONTACT Wai Tong Fung torinfung8@gmail.com

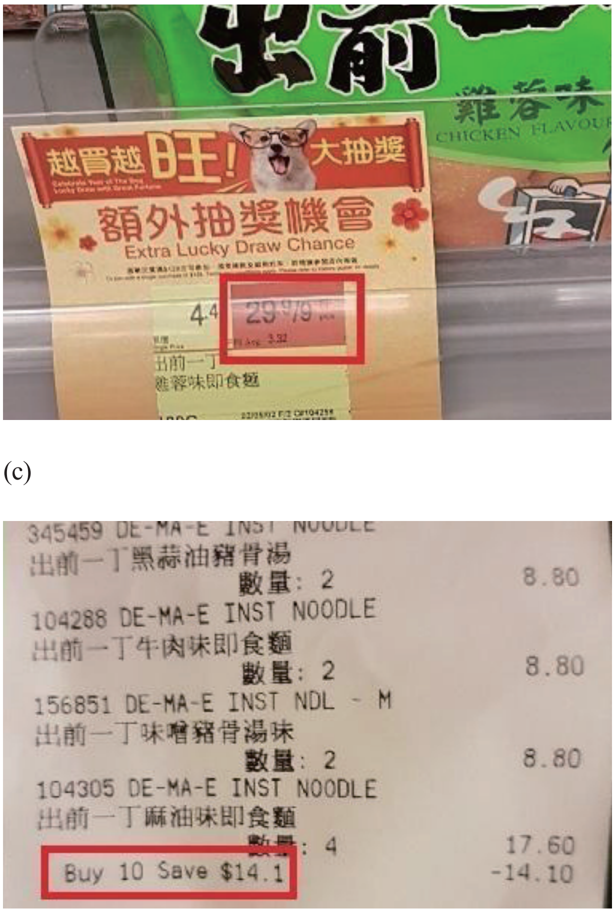
Received 7 February 2020

1. Introduction

Recently, supermarkets have rapidly adopted unmanned stores, supported by Internet-of-Things (IoT) (Guo and Li, 2014), to reduce operation costs and cater for the needs of individual customers as shown in Figure 1. However, there might be defects due to printed price labels shown on the retail shelves at unmanned stores being different from the records at the checkout counter. One example of the defects is the discount rates shown in Figures 1(b) and 1(c). The operator may mix up different price labels or put an incorrect price label on a shelf, and mistakes can easily be made on the labels of daily special promotional items with a promotional offer for buying several discounted items together. In the scenario, as shown in Figures 1(b) and 1(c), the customers need to go back to that shelf and pick up one more piece, then return to the cashier, or even more serious is when customers realise what has happened only after they have arrived at home.

(a)

(b)



(c)

Figure 1. (a) Printed pricing labels on the shelves in an unmanned store; (b) the pricing paper label indicates the erroneous discount rate ($29.90/9 pcs), as shown on planogram; and (c) the printed shopping receipt ($8.80*3+$17.60-14.10 = $29.90/10 pcs) by a cash register.

1.1. Automatic stock management by robots

Automatic monitoring of stock management is based on the technologies of a mobile robot integrated with a coordinate system of the floor to report products that are out of stock for a floor plan in a retail store, and the mobile robot counts the products placed on the shelves (Kejriwal, 2015) by a visual product recognition unit. The other approaches use multiple robots equipped with RFID (Katic and Rodic, 2010) and a 2D laser scanner (Cheng et al., 2017), and loaded with a visual product recognition unit, for customer navigation and inventory control management. The robots can navigate and capture exact pictures of the products in the store, and the multi-robot system can support various tasks in the store, for example, product ordering, product placement and category management. However, this system needs to be pre-set, with the respective locations of all products corresponding to their RFIDs on each shelf or in each area. The robots can read a built-in barcode or a built-in RFID tag in each of the products in the shelf area, and use visual processing to identify product gaps on the shelves. This research has the advantage of using visual product recognition in place of manual operation, which automatically configures placement of all the products anywhere in the store by the use of EPTs. However, the EPTs do not involve the interaction between the products on the shelf and the multi-robot system, and also the expensive and inefficient robots used in the system will disturb the customers during their navigation process. For inventory control management, one or more RFID-arrays are assigned to each shelf, and a fast automatic flight built-in RFID reader, a Lidar and a high-resolution camera are provided in the visual recognition unit that can manage the products on the stock shelf (Beul et al., 2018).

1.2. Mounting EPTs on the shelf and RFID tagged products for stock management

The EPT system was proposed by Evans et al. (1996) in early 1996. An identification of each product on a floor plan was pre-defined in a coordinate system. With the data on zones in the floor plan, the shelves standing on the zones, rows and columns of the shelves, and the products, the EPTs are arranged adjacent to the respective products, with all information contained in the EPTs pre-stored in a network database. The operator manually places all products to be pre-allocated in the formulated zones and on the shelves. The information displayed by the EPT can be auto-configured by a radio antenna moving on the ceiling. The remote pricing (Blau, 2004) module can show the prices on a LC display, and is equipped with a RFID reader which monitors the stock for products with unique RFID tags on the shelves and a Wi-Fi transceiver which communicates with the terminals. However, it is very costly to match the hundreds of EPTs on the shelves with thousands of RFID tags built into the products and to perform product verification on the shelves manually.

1.3. Auto-configuration of EPTs with products on shelves

Smart LaBLEs (Beacon) (Harris et al., 2016) was designed as a BluetoothTM Low Energy (BLE) module to auto-configure the pricing labels to display up-to-date production information about the nearest products. Each beacon has a new mechanism for unification of identifiers (Rykowski, 2017), via two-way wireless communication with multiple channels for broadcasting. Thus, it is not necessary to determine the exact location or physical coordinates of each product corresponding to each EPT. The traditional ranging technique of a Received Signal Strength Indicator (RSSI) can find the location of an object within a certain range. In the experiments of linear or circular sensing between the beacons/tags and the antenna device, the RSSIs of several beacons are measured to have very close results in different orientations in the range. Hence, all the same tagged products can be placed in a unidirectional manner behind the device for scanning the

EPTs on the shelves, and monitored regarding their EPTs within the range. Actually, this approach using the latest Bluetooth technology is very close to the method of Evans et al. (1996) using the traditional RFID technology, but the cost of a BLE beacon is much higher than that of a RFID tag. Another approach is that an EPT is equipped with both BLE and RFID readers, and a RFID tag is affixed to the product. The cost of tags can be minimised for thousands of products. The monitoring function performed by the robot can be replaced by EPTs to make cost savings, but there is a need to create a mapping database of location and identification between all the EPTs and all the tagged products.

1.4. New proposed research methodology of fully automatic retail inventory and pricing system

The above-mentioned papers discussed the design and implementation of supermarket stock management, but the floor plan is not fully computerised by locally mapping the products into zones, shelves, and EPTs. Hence, a new methodology of auto-positioning and auto-configuration of EPTs through optical communication and visual recognition is proposed. One EPT can represent all the freely placed products without any preset identification on the shelf. The system can automatically computerise all non-identified EPTs, the shelves and the zones, where the non-identified EPTs are located on the 3D floor plan. The instantaneous information regarding the EPTs can be displayed and shown on a mobile phone of an individual customer during his or her navigation of the floor plan. The supermarket can not only provide various special offers, including special discount rates, buy one get one free, special gifts, etc., according to their shopping behaviour as analysed by IoT with a Point of Sale (PoS) System, but also connect to inventory control management and payment gateway systems. An example shelf processed by our proposed system is shown in Figure 2(a), which is substantially comparable to the retail shelf shown in Figure 2(b), and the EPTs affixed to the shelf of Figure 2(a) are shown in Figure 2(c).

(a)

(b)

(c)

Figure 2. (a) Example retail shelf; (b) supermarket retail shelf; and (c) electronic pricing tags (EPTs).

2. Related work

This section presents the major works that address the limitations of localisation of and communication with the products or EPTs on retail shelves via several media. The cost comparison with the use of frequency bands for SLAM is shown in Table 1.

The barcodes or QR code tags are the cheapest solution for product identification, but have a narrow scanning angle and a very short scanning range (<20 cm). Moreover, the barcode or QR code technologies cannot align the scanning angles of the tags one by one if there is a pile of tagged products located on the shelves. Thus, these technologies cannot be applied to remote inventory management of tagged products on the shelves. Hence, it is quite popular to use RFID tags built into the products or RFID devices for low-cost inventory management. An active or passive RFID tag is pasted on each product, so that a pile of products can be localised and counted within the inventory boundary.

A wide range of sensors, including light, infrared, pressure, push pin sensors or the like, are located inside the shelves, and also one or more cameras are mounted on a store roof to visually recognise a plurality of images of the shelves to count the in-stock products and to determine and

Table 1. Use of frequency bands with cost for Simultaneous Localisation and Mapping (SLAM).

Cost	Types of electronic tag	Media	Frequency range	Localisation	Communication
Low	Barcode / QR Code	Optics	300 GHz	Indoor	Uni-direction
	NFC / RFID	Radio	100 KHz~1 GHz	Indoor	Bi-direction
	Infrared (IR)	Optics	300 GHz	Indoor	Bi-direction
	Ultrasonic	Ultrasound	20 MHz~50 MHz	Indoor/Outdoor	Uni-direction
	NBIoT	Radio	900 MHz (GSM)	Indoor/Outdoor	Bi-direction
	BLE IoT	Radio	2.4 GHz	Indoor/Outdoor	Bi-direction
	Wi-Fi	Radio	2.4G Hz~5.6G Hz	Indoor/Outdoor	Bi-direction
High	Microwave	Magnetic	300 GHz~300 MHz	Indoor/Outdoor	Bi-direction

verify whether the product is out of stock (Vargheese and Dahir, 2014). However, the sensors used for this purpose are very costly. The application of a monocular RFID reader with several RFID tags for indoor localisation is a low-cost approach, and Boontrai et al. (Boontrai, 2009) proposed a location estimation method based on the angulation technique to determine the reading range, as well as providing the step of equipping the UHF-band (902-928 MHz) RFID reader with a linear polarised antenna and attaching the RFID tags to the wall. In their experiments, the inter-tag distance was limited to 60 cm, and the measurement distances between the tags and the reader were 100 cm and 150 cm, respectively. The application of the monocular reader/antenna is out of consideration of cost. However, the maximum location estimation error in the measurements is 50 cm. As a result, this approach can only find an application in object clusters on the wall.

Recently, NBIoT and BLE IoT beacons pasted on a device have been very popular, as IoT can collect the data anywhere. The following evaluates and analyses various approaches configured by different technologies for positioning, mapping, and recognition of objects with the major consideration of reducing costs.

2.1. Localisation and tracking by sensing arrays in the environment

2.1.1. RFID approach

Multi-Antenna-RFID Readers are used for robotised self-localisation by an indoor navigation system (Takahashi, 2013). Textile carpets with an array of RFID tags are installed on the floor. The robot is constructed to be equipped with multiple RFID readers/antennas to detect the RFID tags for localisation during the navigation process. The average position-estimation error amounts to 1.3 cm.

2.1.2. RFID or wireless communication with the RSSI approach

Joni et al. (Jamsa, 2010) described multi-RF readers, which are fixed in terms of their location to transmit and receive active RFID tags built into the robot. The distance between the reader and the tagged robot can be measured by RSSI and the angle of arrival (AoA) detects the direction of the tags relative to the reader. The tagged robot can be located and tracked by the RSSI and the AoA through intersections of circles predefined on the floor. A 2D indoor positioning method was discussed by Savochkin (2013), where the method is also operated on an array of RFID tags on the floor. However, the method is characterised by the use of a monocular RFID reader/antenna in combination with a trilateration method based on RSSI, with the mean localisation error of around 31 cm. Gimpilevich et al. (2013) proposed using multiple RFID readers/antennas in combination with an integral read rate method based on RSSI, with the mean localisation error of around 27.2 cm.

According to Lu et al. (2014), the wireless sensor arrays constructed on the floor have the average distance error of about 1.2 m for robotised navigation. After the sensor arrays are integrated with a SVM-WKNN method for robotised navigation, the SVM-WKNN can compensate for the non-linear signal measurement of RSSI for linear distance measurement. This improvement results in an improved average distance error of about 8%. In accordance with Tobias et al. (2012), UWB radar with a bat-type sensor configuration forms the sensor array on the floor: a simple mobile antenna array with one transmitter and two receivers. Corners, walls and edges can be detected by geometrical propagation path models. The frequency response function was calculated from a frequency in the range of 4.5 GHz to 13.4 GHz for an array comprising three double-ridged horn antennas. The distance measurement is more accurate, and 20 cm approximately. According to Ma et al. (2017), the inertial sensor-array is localised on the floor for human indoor navigation; the position of a person is measured by a magnetic field, which has a positioning error of about 1.8 m.

The identification of the ongoing activities of people is tracked by a smart home system (Fortin-Simard, 2012), where several passive RFIDs are tagged nearby the appliances that represent the ongoing activities, and multiple RFID readers/antennas are located on the floor for localisation and tracking of people based on elliptical trilateration and fuzzy logic. The method was tested with multiple filters, depending on the RSSI generated by the antennas. The activities of the tracked people are localised somewhere in elliptical regions following the principle of trilateration. The average accuracy of +/- 14.12 cm was obtained in the test of people tracking localisation. This paper compares the precise localisation of the method with other methodologies based on Ultrasonic, ZigBee and RFID. Most of the research has proposed either using multiple transmitters or multiple receivers for object localisation. Some of them can be used for object positioning, but are very costly for installation. The installation cost is directly proportional to the floor area and will be significant for an increased floor area.

2.2. Ultrasound or optics with coordinate system approaches

Multiple microphone sensors/receivers were proposed (Murata et al., 2014) to be embedded in an environment according to a predefined coordinate system on the floor. A near-ultrasonic transmitter is fixed to a moving object. The coordinates of the moving object can be derived by acquiring the distance between a sound transmitter and each receiver, and the relative time delays at each receiver to be detected. The average of the coordinate error is smaller than 12 cm for object tracking. Low-cost optical sensing arrays (Salomon et al., 2006) are allocated by the coordinate system and communicated with a robot for self-localisation in a large environment. The position-estimation error is over 70 cm. Using multiple cameras or optical sensors can identify the object localisation, but may have limitations in some circumstances. If the object is invisible or two objects overlap each other, object localisation cannot be performed.

2.3.Localisation and tracking with hybrid sensing approaches

2.3.1. Radio-optical approach

A radio-optical beaconing system was introduced by Ashok et al. (2015). The beacon transmitter has two built-in parts: an infrared emitter that carries no bits of information for accurate positioning through AoA, and RF packets via the 900 MHz radio channel with the CSMA communication protocol to transmit information including identification, IR pulse calibration, and synchronisation for angle and range estimation at the receiver side. The receiver comprises several photodiodes to form a receiving array, which can achieve a wide coverage range for receiving IR signals. The ranging error is about 40 cm.

2.3.2. Radio-vision approaches

Radio-vision (R-V) is based on an indoor positioning and identification system for people tracking (Wang and Cheng, 2011). The positioning blocks of the system are defined by RF sensing arrays with RSSI allocated on the floor and stored in a database. The absolute coordinates determined by an object recognition unit via web cameras and the rough position of tagged people determined by the sensing arrays are calculated individually and combined. This combination has a better accuracy than the RF sensing arrays with RSSI only in terms of people positioning and tracking. The tracking of RF-tagged medicine bottles on a table using a R-V module and a vision system was proposed by Hasanuzzaman et al. (2011). The R-V module is a built-in part of the RFID antenna-reader, monocular camera and USB interface. The vision system can separate medicine bottles from other moving objects, and the object must be in the range of the RFID antenna.

The vision system integrated with ultrasonic sensors can define the range of a target and distinguish several targets from one another over the same target range (Chou and Wykes, 1997). The system consists of a single camera, two ultrasonic transmitters and one ultrasonic receiver. Since the geographical environment of a supermarket is not suitable for ultrasonic sensors to be mounted on the floor of each shelf, which also requires a high setup cost, this positioning method cannot be applied to stock management.

2.4. Other approaches for identification and localisation of objects

A single Kinect scanner comprises a built-in infrared camera, a VGA camera and an infrared laser emitter, and two optical cameras are used in combination with a static Leap Motion Controller for image mapping of the indoor plane (Kusari et al. 2014). The point-plane accuracy is about 10-11 cm.

One of the low-cost approaches for rough localisation of ID-tagged objects is that an optics-frequency device is used to search for the objects within the visible range, but the AoA is too wide to allow LoS communication (Hong et al., 2013). The inter-tag distance is too wide between the ID-tagged objects. The multiple emitters transmit signals to search for and detect the ID-tagged objects for any user identification in the meeting room. A motion camera with emitters can display the rough locations of the ID-tagged user.

According to Guan et al. (2018), a testing environment for object positioning and a tracking system with multiple transmitters were designed by using visible light LEDs mounted on the ceiling, and a smart car built-in industry camera can automatically drive on the floor. The proposed indoor dynamic positioning and tracking algorithm using optical flow and Bayesian forecasting can detect in real time the LEDs and position the camera. The accuracy of positioning the moving camera, at a speed of up to 48 km/h, can amount to 0.86 cm. Visual object recognition by a monocular camera provides a low-cost solution for rough location and tracking of each product on the shelves. The survey of several product recognition approaches by the use of remote sensing technologies based on shelf images was summarised by Melek et al. (2017). The object recognition approaches for book-spine image matching by CDSIFT and C-Color-SIFT colour detectors (Fowers, 2012) and the supervised deep learning for scene text reading (Yang, 2017) were presented to bookshelf management based on the standard off-the-shelf configuration in a library. The outcomes of book management on the bookshelves are classified as the scenarios of book remaining, book removed and book misplaced.

The other visual object recognition approaches use the methodology of image analytics to monitor products on the shelves. The multiple detection strategies of products recognition consist of vote mapping, Histograms of Oriented Gradients (HOG) and Bag-of-Words (BOW) (Marder et al., 2015). The inventory management system can receive alerts regarding insufficient quantities, missing products, and misplaced items. It is not necessary for books on the shelves in a library and products on the shelves in a supermarket to have built-in RFID tags, beacons or the like.

All the above methodologies involving localisation, positioning and tracking are summarised in Figure 3. The features and costs associated with these methodologies are tabulated and compared in Table 2. The RFID, beacon and ultrasonic approaches transmit non-directional signals, but can be designed to provide a RF array-based architecture, which can define the precise positioning of objects within the intersection range. However, these kinds of approaches are very costly for installation. Either beacons or optical solutions are selected to communicate with EPTs for price configuration. The beacons are quite expensive because of the built-in Bluetooth, NBIoT, which are applied to valuable products for IoT anywhere. Furthermore, they might not be suitable for Fast-Moving Consumer Goods (FMCGs) in a supermarket.

The low-cost visual approach to hardware is very useful to roughly localise different appearances of objects within a certain range, and can replace the RFID or beacon ID tags in the objects. In other words, the appearance of the objects is the same, but different ID-tags in the objects can be individually in communication with and detected by the wireless beacons. However, the beacons cannot be distinguished from one another in regard to their accurate locations, if they are located within the coverage range using visual recognition. A variety of emitters, sensors, and transceivers usable in the market for localisation, positioning, and communication of objects are summarised and shown in Figure 3.

Figure 3. Types of transmitter/sensor/receiver for localisation, positioning and communication of objects.

In this paper, a new low-cost solution to the monitoring controller, and a rotatable infrared laser emitter with a monocular mobile camera, listed as Items 14 and 15 in Table 2 are proposed. They can provide auto-positioning and auto-configuration of non-identified EPTs representing all non-identified products that can be arbitrarily placed on the shelves. In the following sections, the methods are described in detail and their performance is evaluated.

3. Proposed approaches

3.1. A new visual-optical-based positioning methodology

In the proposed method, each product has one or more identification images, instead of a tagged identification readable by RF or beacon technologies. A pile of products can be arbitrarily placed on the shelves by an operator, which can reduce the labour intensity. The monitoring controller can visually recognise and match each non-identified EPT with its corresponding product. Hence, the identification of each EPT can be assigned in real time, and therefore can eliminate the up-front investment cost to fit and record all tagged products manually.

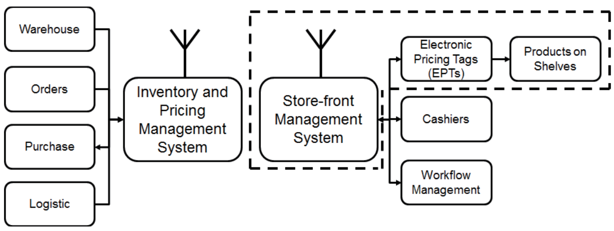


Figure 4. The storefront management system connected to EPTs and products, as well as to the inventory and pricing management system via Wi-Fi.

This paper proposes a new application of the visual-optical approach for a low-cost storefront pricing management system. The pricing management system, after completion, can connect to EPTs and products on the shelves. The proposed model is illustrated inside the dotted lines in Figure 4. Each product is assigned a printed price

Table 2. Summary of the costs of different methodologies in terms of communication, positioning and tracking functions.

	Types of sensing application	References	Media	Localisation (Roughly)	Positioning accuracy	Comm-unication	Costs: Hard ware	Costs: Install -ation	Costs: Identifi-cation mapping
Sensing arrays approach									
1	RFID Array (Tx/Rx)	Yasutake et al. Y2013 [12]	Radio	Yes	No	No	Low	High	High
	With RSSI	Joni et al. Y2012 [13], Yu et al. Y2013 [15], Xiaoqing et al. Y2014 [16]	Radio or wireless	Yes	No	No	High	High	High
	With AoA	Savochkin et al. Y2014 [14]	Radio	Yes	No	No	High	High	High
	With geometry	Tobias et al. Y2012 [17]	Radio	N/A	Yes	No	High	High	High
	With trilateration	Dany et al. Y2012 [19]	Radio	Yes	No	No	High	High	High
2	Set of inertial sensors (Tx/Rx)	Zixiang et al. Y Y2017 [18]	Magnetic field	N/A	Yes	No	High	High	N/A
3	Set of ultrasonic (Tx)	Shotaro et al. Y2014 [20]	Ultra-sound	N/A	Yes	No	Low	Low	High
4	Set of beacons (Zigbee, Z-wave, Wi-Fi, BLE-IoT, NB-IoT....)	Albert et al. Y2016 [8]	Radio	Yes	No	Yes	High	High	Low
Monocular type of sensor approach									
5	Single RFID transmitter	Boontrai et al. Y2009 [11]	Radio	Yes	No	No	Low	Low	Low
6	Set of infrared transmitters	Wei et al. Y2013 [27]	Optics	Yes	No	Yes	Low	Low	High
Hybrid approach									
7	Wireless and infra-red	Ashwin et al. Y2015 [22]	Wireless + Optics	N/A	Yes	No	Low	High	High
8	RFID and camera	C.S. Wang et al. Y2011 [23], Faiz et al. Y2011 [24]	Radio + Vision	N/A	Yes	No	Low	High	High
9	Infrared laser and camera	Arpan et al. Y2014 [26]	Optics + Vision	Yes	No	No	Low	Low	N/A
10	Visual light and camera	W. Guan. et al. Y2018 [28]	Optics + Vision	N/A	Yes	No	Low	Low	N/A
Tx means transmitter, Rx means receiver									
This paper proposes a method for non-identification of EPT auto-positioning									
11	Rotatable infrared laser emitter		Optics	N/A	Yes	Yes	Low	Low	Low
This paper will propose a method for non-identification of EPT auto-configuration									
12	Camera and infrared laser emitter		Optics + Vision	N/A	Yes	Yes	Low	Low	Low

label, which displays an accurate description, and its price and place on the shelf. Furthermore, the printed price label is replaced by an EPT, described as follows.

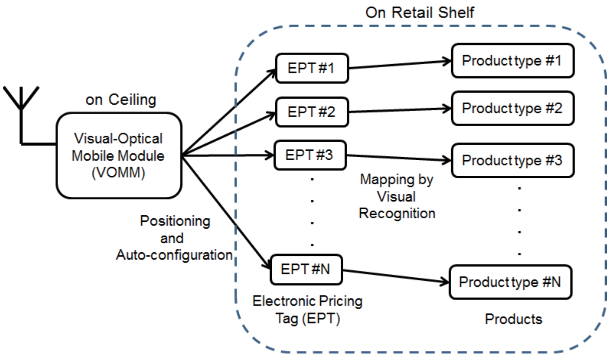
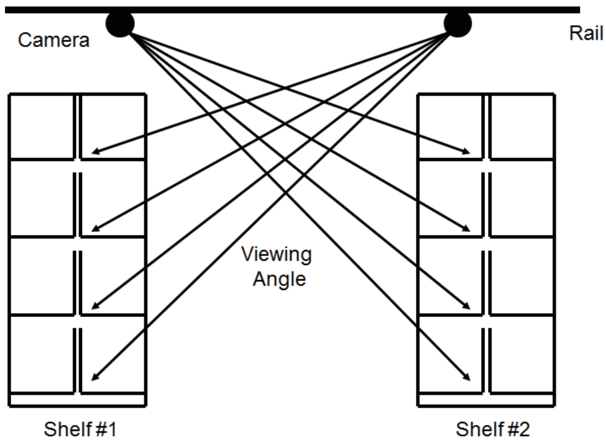


Figure 5. Block diagram of the Visual-Optical Mobile Module (VOMM) for a full store-front pricing management system.

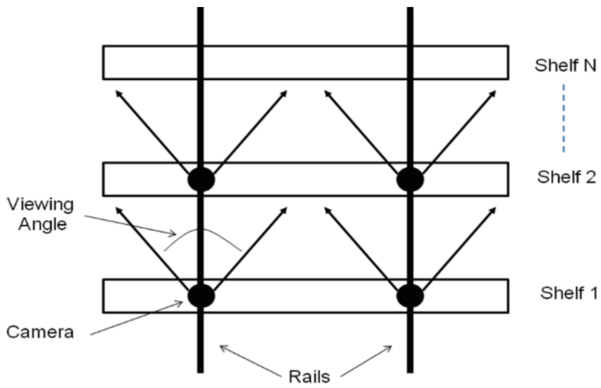
A complete store-front pricing management system is shown in Figure 5. The Visual-Optical Mobile Module (VOMM) is recommended to be mounted on the ceiling in order to more easily monitor the goods placed on the shelves. Rails can be installed on the ceiling and above the shelves, so that the VOMM can move along the rails, as illustrated in Figures 6(a) and 6(b). The VOMM moving on the rails is rotatable and controlled by a servo motor, so that the number of VOMMs required for mounting on the ceiling is minimised. The motion of each VOMM is programmable, according to which zone and to which shelf the VOMM is allocated. A VOMM consists of a monocular mobile camera and a rotatable infrared laser emitter to perform visual recognition, auto-positioning, and auto-configuration of all the non-identified EPTs. Each EPT on the shelves can represent the nearest isomorphic products by visual recognition. Thus, no RF tag or beacon is required to be built into the products before placing them on the shelves. It is not compulsory for the identification of the products to be pre-stored in a database manually. Moreover, the products can be arbitrarily placed anywhere on the shelves and the EPT can accurately show all the correct information, such as product description, price, promotional period, or the like. Consequently, the labour intensity and human error caused by the operator can be minimised. The coordinates of the shelves and the zones are obtainable by projecting the coordinates of all EPTs on each shelf and all the isomorphic products are represented by one EPT. The overall geometrical coordinate system can be shown on a 3D floor plan using a mobile phone for customer navigation in the supermarket.

(a)



Front view

(b)



Top view

Figure 6. (a) An example VOMM auto-configures the EPTs and recognises the products on the shelves; and (b) example rails are mounted on the ceiling, along which the cameras move

3.2. Design of the Visual-Optical Mobile Module (VOMM)

The VOMM, shown in Figure 7, consists of two functional parts: 1) a mobile phone module with a monocular camera, a BLE and Wi-Fi transceiver; and 2) a rotatable infrared laser emitter module, which connects to the mobile phone module via a Bluetooth transceiver. The infrared laser emitter is rotatable, on top of X-Y stepping motors, controlled by a microprocessor (MCU) for Simultaneous Localization and Mapping (SLAM). For power-saving purposes in normal operations, a low-power MCU is employed to operatively communicate with the mobile phone module via a BLE module. The built-in monocular camera in the mobile phone provides visual recognition of both the products and EPTs on the shelves, to detect and recognise them according to their identification images, so as to obtain their rough location.

The Wi-Fi and BLE transceivers use a standard module with a private communication protocol pre-installed in a personal computer or smart phone to instantly access the automatic stock management system, if necessary.

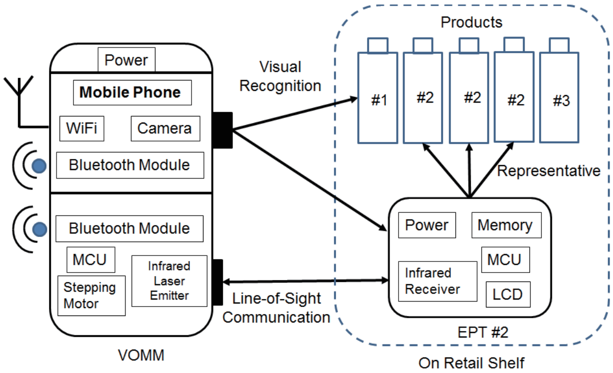
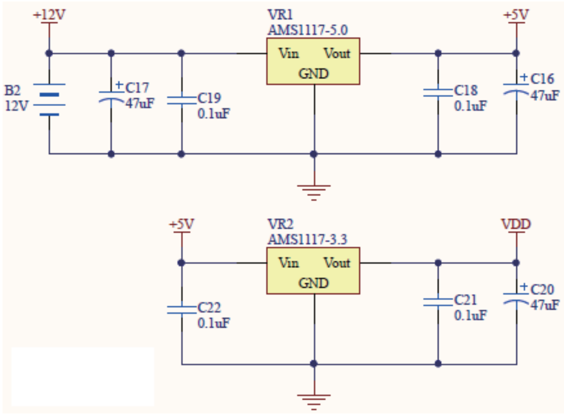


Figure 7. Hardware and communication configuration of the Visual-Optical Mobile Module (VOMM) and Electronic Pricing Tag (EPT) representing the products on the shelves.

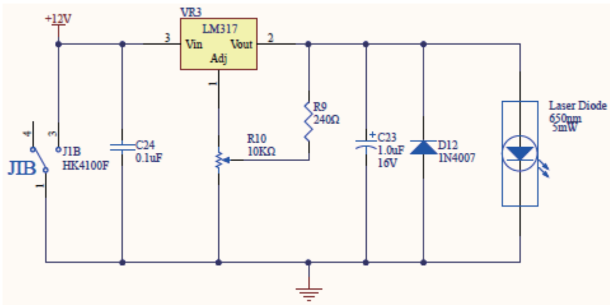
In Figure 8(a), an input DC power source at 12 V is applied to the VOMM. The DC12 V is converted to DC5 V, then down to DC3 V, by two regulators to supply the power for the MCU. The DC12 V power infrared laser emitter module can drive a high current to the infrared laser driving circuit in Figure 8(b) to emit signals. The infrared laser emitter is a strong coherent light source having the advantages of high brightness, pure colour, high energy and directional propagation. In this research, it is proposed to transmit the data signal useful for LoS communication between the emitter and the receiver in the ambient light environment. According to the specifications of infrared LED emitter and infrared LED receiver given in Table 4, they have the transmission and receiving angles of +/-20°, such that the transmission angle should be narrowed down, and the receiver lens should be painted with a coating for ambient light filtering.

The rotatable emitter is mounted on a stand and locked on the shaft of a stepping motor. The driving circuit of the stepping motor is shown in Figure 8(c). All the above functional modules are controlled by the MCU with RAM/ROM memories, with a LCD display, as shown in Figure 8(d). The BLE module in Figure 9 is designed for a 2.4 GHz ISM band with 3 dBm output power for wireless application having the data rate at 1 Mbps with GFSK modulation for a frequency hopping spread spectrum system (FHSS).

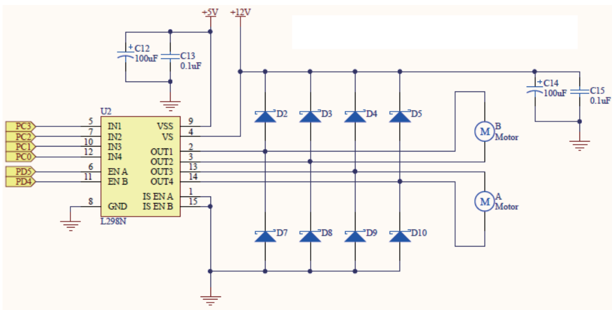
(a)



(b)



(c)



(d)

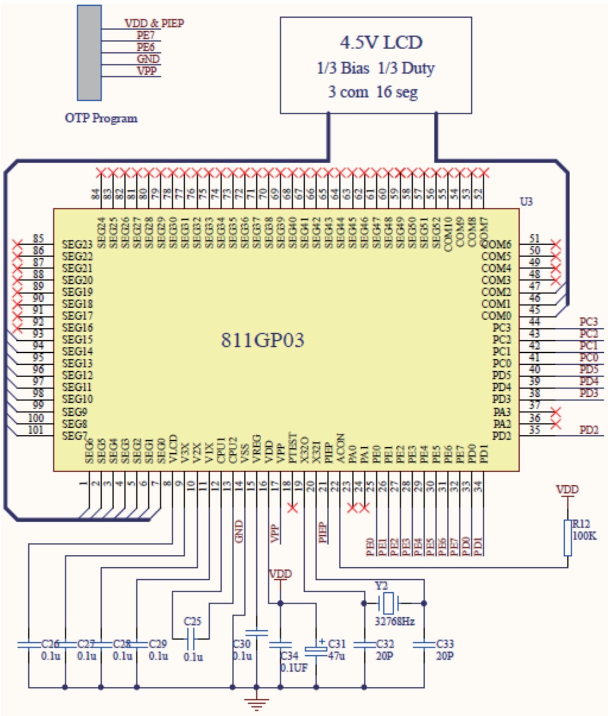


Figure 8. (a) Power source circuit for converting DC12 V input to DC5 V and DC3 V outputs by voltage regulators; (b) signal emission by an infrared laser driving circuit; (c) stepping motor driving circuit; and (d) microprocessor (MCU) with memories and LC display.

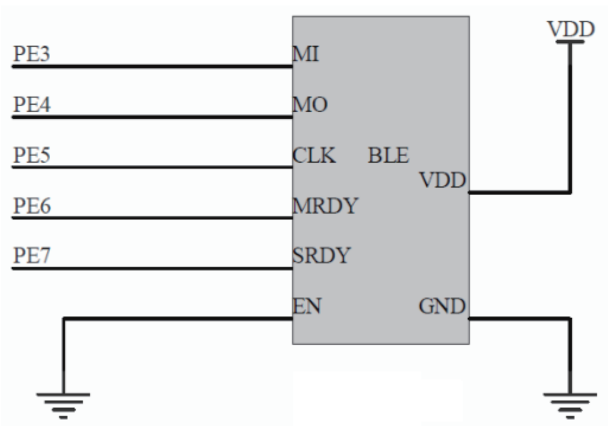


Figure 9. Bluetooth Low Energy (BLE) module.

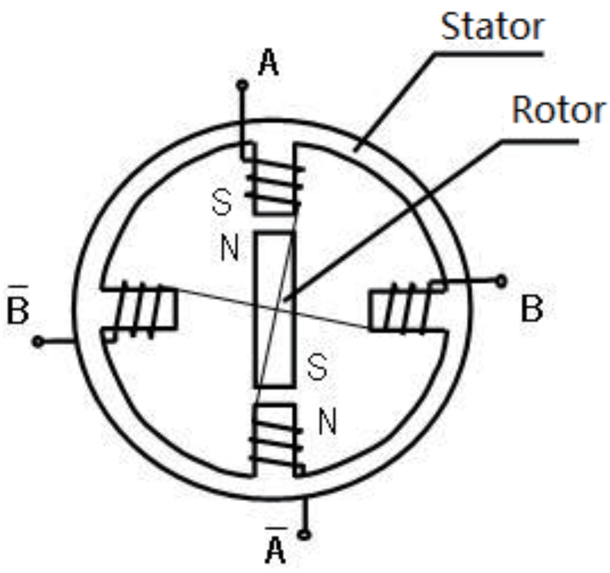
3.2.1 Subdivision control of the stepping motor

The step angle in the stepping motor is related to the number of beats of the motor and the number of teeth of the rotor, θ=360/NS (the formula of a 2-phase motor and N is the number of shots generally determined by the number of lines which is the number of teeth of the rotor). Subdivision control refers to step-by-step control by the step angle.

Figures 8 to 10 mention that the position of each EPT can be segmented by a rotated IR emitter driven by the stepping motor (built-in VOMM), then communicated between each EPT and VOMM while all EPTs are too close and have the same pattern. The camera with image recognition function is for products which match each EPT on the shelf which can prove that the proposed hardware configuration can be workable.

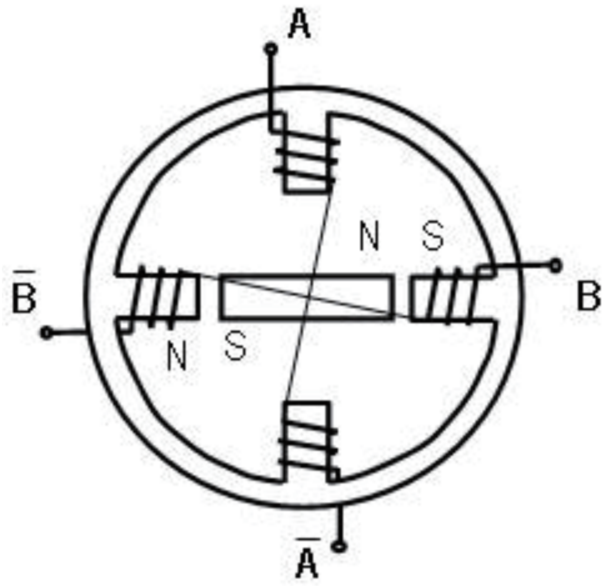
In the application of this research, a two-phase four-wire stepping motor is used as the angle divider for segmentation of the X-Y plane, as illustrated in Figure 13(a). The stepping motor is a part of the low-cost angle tracking device. The subdivision control of the stepping motor is designed to realise a kind of sine pulse width modulation (SPWM) subdivision driving technology, which can control the current waveform of the rotor more effectively (Wang et al., 2017). A full revolution requires 200 steps, while each step turns the shaft only 1.8° for a full step, or 0.9° in the half-stepping mode. In Figures 10 and 11, A / B denotes a forward current, while ~A / ~B denotes a reverse current. One complete step contains four subdivisions by A, ~A, B and ~B, and the turning sequence is A => B => ~A => ~B for the clockwise driving direction. It is assumed that the angle of rotation between each step is 90°, as shown in Figure 10.

(a)



A/~A

(b)

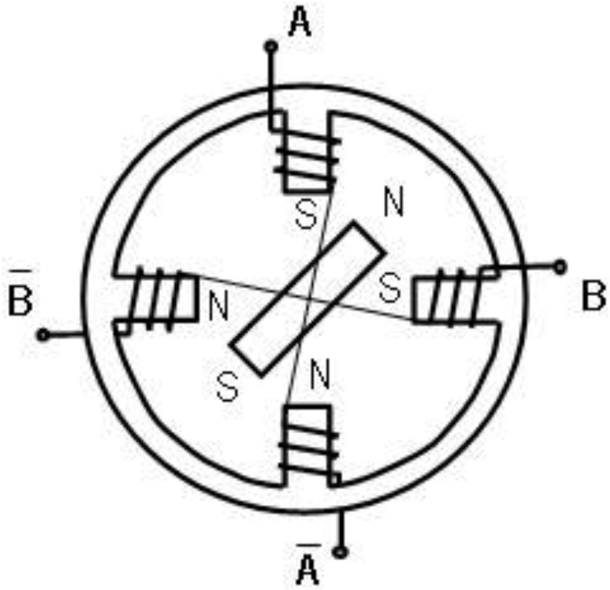


B/~B

Figure 10. (a) and (b) State of one-phase conductive rotor.

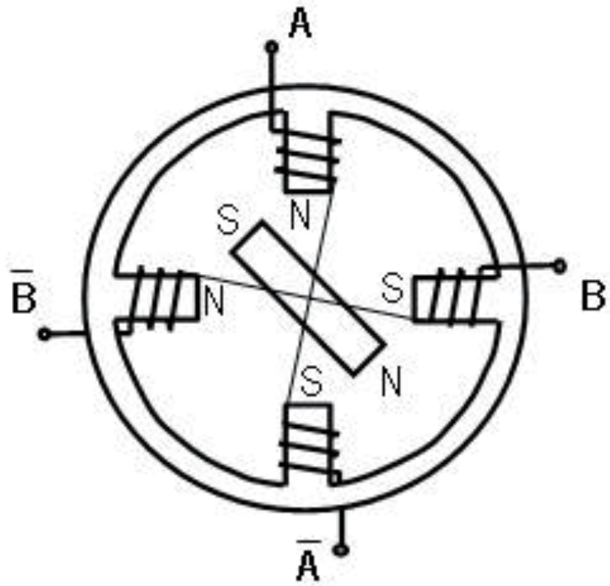
Furthermore, the original four subdivisions drive has only a single-phase current in each state that is an A-phase coil aligned with the rotor, or a B-phase coil aligned with the rotor, as shown in Figure 11. When the pair of rotors are located at the middle of A and B, the driving current needs to energise AB at the same time, such that the two electrodes have the same attraction, then the tooth pair is turned 45° on the centre line of AB. The corresponding timing of eight subdivisions is A ==> AB ==> B ==> B~A ==> ~A ==> ~A~B ==> ~B ==> ~BA.

(a)



AB and ~A~B

(b)



B~A and ~BA

Figure 11. (a) and (b) State of the two-phase conductive rotor.

3.3. The hardware of Electronic Pricing Tags (EPTs)

With reference to Table 2, most of the existing methods have used dozens of wireless sensors (RFID, beacons, ultrasound, optics, etc.) for transmitters, receivers or both, which are installed to enumerate geo-locations for object tracking with positioning within a small area. Higher accuracy of object positioning requires more beacons, transceivers or sensors to narrow the coverage range till the saturation of intersection. Thus, the installation and hardware costs are very high to locate all EPTs for retail shelf management.

In practice, the system does not need to count the exact number of products on a shelf, because the cashier should deduce the quantity of the products in the supermarket during check-out. For a cost-saving approach, it should not be required to detect the ID tagged, special labels for barcodes pasted or printed on thousands of products.

The low cost of non-identified EPTs was designed by the use of a low-power MCU with RAM/ROM memories and LC-display (LCD), as shown in Figure 8(d). The circuit of the signal amplifier for receiving infrared laser signals, as shown in Figure 12, is used to increase the sensitivity of the received signal.

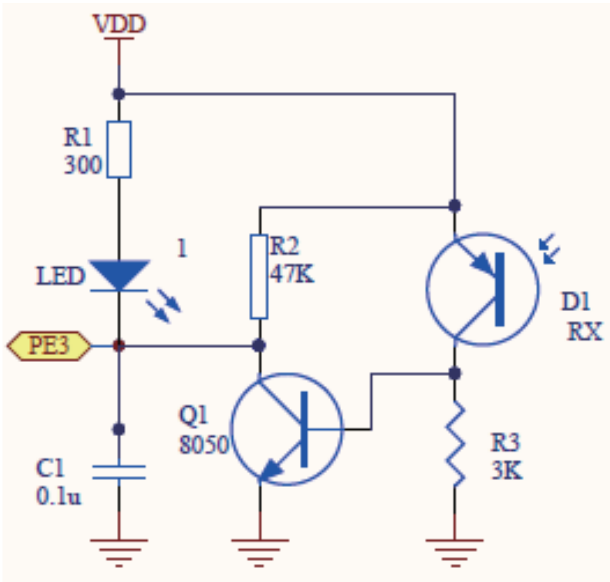


Figure 12. The circuit of signal amplifier for receiving the infrared laser signal.

The infrared receiver was designed with housing, such that the receiving angle will be limited, and also the lens are adapted with a light filter to prevent sunlight or lighting interference. This ensures that only the infrared signal is received. The VOMM can configure each EPT via the communication between the infrared laser emitter and receiver, as shown in Figure 6. The price information or other promotion details are stored in the memory, and are then shown on the LCD after the signal is received.

3.4. Auto-positioning and auto-configuration of Electronic Pricing Tags (EPTs)

For auto-positioning of each EPT on a shelf, the software design of the VOMM for signal transmission provides LoS communication and minimises the AoA of the emitter. The rotatable function of the VOMM can project and segment the partitions in the front portion of the shelf. A method for instantly positioning an EPT by SLAM and configuring an EPT by LoS is illustrated in Figures 13(a) and 13(b). While the infrared receiver built-in EPTs are located in the X-Y plane, the transmission signal from the emitter in the rotatable VOMM can localise and communicate with EPTs. The infrared receiver in an EPT (e.g. represented as the circle in Figure 14) receives this LoS and AoA signal from the emitter in the VOMM in the X-Y plane, where w is the label of infrared receiver in the EPT. The centre (*Xc*, *Yc*) of the EPT is determined based on the coordinates of the infrared receiver (*Xw*, *Yw*), as shown in Figure 15. The distance can be calculated from the X-Y plane of the position of the EPT, thereby obtaining the centre (*Xc*, *Yc*) of the EPT. Figures 13 to 15 illustrate an example of positioning, recognising and configuring an EPT using a VOMM, comprising one camera unit only.

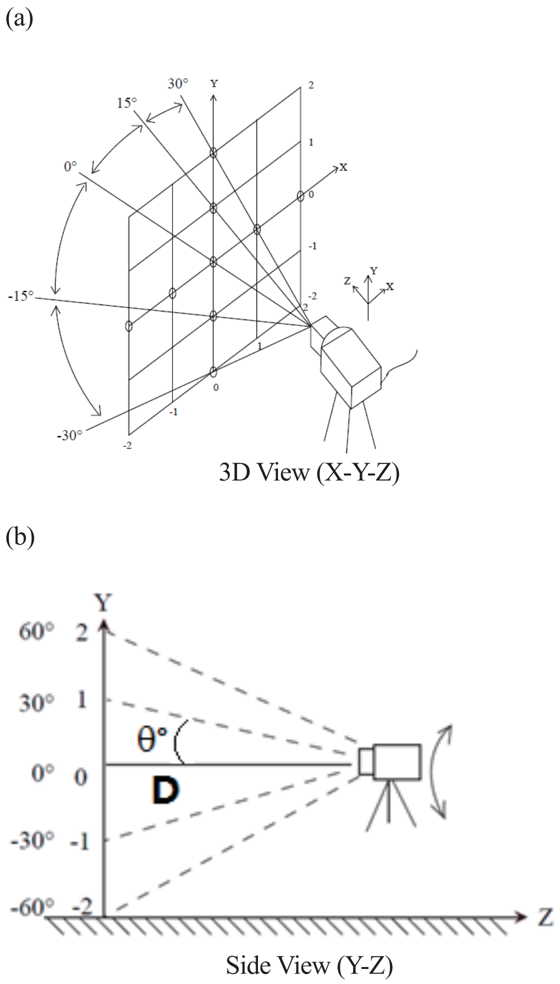


Figure 13. (a) The Simultaneous Localisation and Mapping (SLAM) partitioning of the 3D space (defined by the X, Y, Z axes) by the rotatable VOMM; and (b) the Line-of-sight (LoS) communicating by the infrared laser emitter.

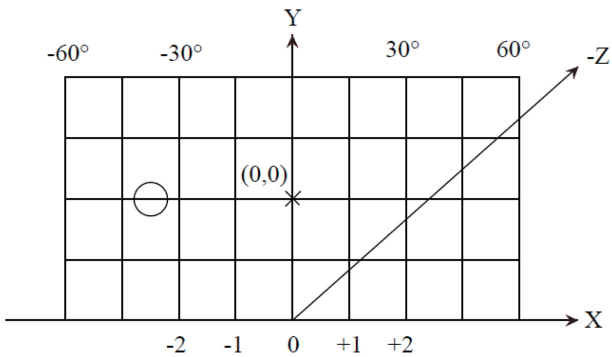


Figure 14. Determination of the number of regions on the X-Y plane by SLAM, and representation of the position of the infrared receiver with the circle centred at (*Xw*, *Yw*) in the EPT.

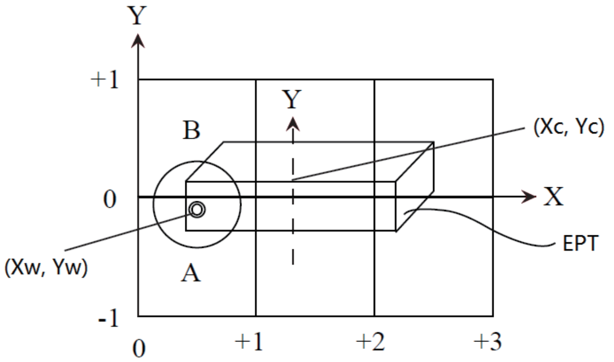


Figure 15. Determination of the distance between the infrared receiver at (*Xw*, *Yw*) and the centre of the EPT at (*Xc*, *Yc*).

The rotatable angle of a stepping motor is normally 1.8° per step. Theoretical calculation of the inter-tag distance (*Xit*) aims to ensure that each EPT is not subjected to any interference from others when it is placed on the shelf (Figure 16).

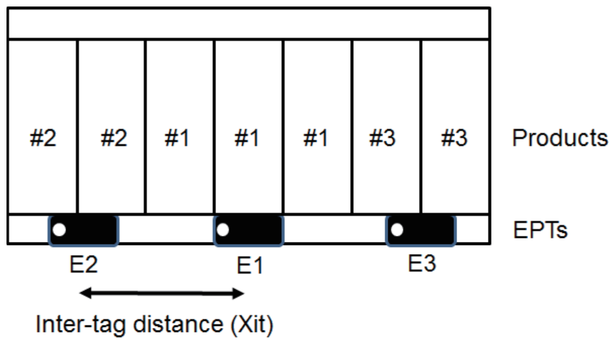


Figure 16. Example of the inter-tag distance (*Xit*) which is measured between the centre coordinate (*Xc*) of EPTs (*E = E1, E2, E3*) of the isomorphism products (P = #1, #2, #3) on the shelf.

While the LoS transmission signal is sent from the IR emitter in the rotatable VOMM and received by the IR receiver in the EPT respectively, the inter-tag distance (*Xit*) is obtained by the below theoretical formula:

$$Xit = C2 - C1, \tag{1}$$

where *C2* and *C1* is the centre (*Xc*) coordinate of *E2* and *E1* respectively;

$$\tan(\theta°) = Xit / D, \tag{2}$$

where θ is an angle of step and *D* is the measurement distance between the VOMM and the X-Y plane shown in Figure 13(d).

For example: If $D = 1$ m and $\theta = 1.8°$, $Xit = \tan(1.8°) \times 1 = 3.15$ cm.

3.5. Auto-mapping between EPTs and all isomorphism products

Object detection methods are used to detect and locate all the products represented by each EPT on a shelf. When a customer wants to locate a particular item using a smartphone, all the required information, including the shelf location, the number of the items on the shelf, etc., will be provided. The auto-mapping is performed by the coordinate of each EPT in the X-Y plane. The VOMM can visually recognise all isomorphism products by the nearest neighbour searching method after all products have been freely placed on the shelf, as shown in Figure 17. In total there are nine shampoos and four EPTs involved in the experiments of visual recognition. The nine shampoo items are labelled from left to right as: A, A*, B, C, D, E, F, G and H, where A* is a repeated product of product A.

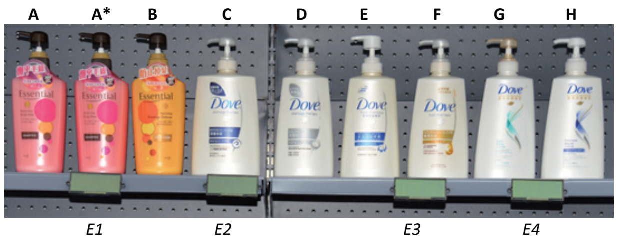


Figure 17. The nearest neighbour search of isomorphism products (A, A* to H) represented by EPTs (E1, E2, E3 and E4).

The visual-recognition algorithm identifies the orientation of isomorphism products on a shelf according to the shelf image captured by the camera in the VOMM. Then, the number of products on each level of the shelf is counted. As the VOMMs are located on the ceiling, feature matching is performed based on a number of 2-D images of the items on the shelf. Then, the position, orientation and size of each object are detected and determined.

Recently, local binary invariant features (Phan et al., 2013) were combined with colour histograms for product recognition. However, in the scenario of a shelf, there are limitations to using colour histograms for the recognition of product variants. It is difficult to differentiate similar variant products, for example, as shown in Figure 17: (i) isomorphism products of the same product line, e.g. repeated products having a uniform colour of products, such as A and A*; (ii) isomorphism products having different colours, according to the different subtype of the products, such as A, A* and B; (iii) products C and D having different sizes of a particular brand; (iv) different series of a product of a particular brand often having slightly different wordings, such as the products D, E and F, and (v) the same series of a products, such as G and H, varying slightly in their patterns. Thus, invariant feature-based methods for product recognition are the main focus of this work.

The first selected method is Scale Invariance Feature

Transform (SIFT) (Lowe, 2004), which is a feature descriptor that is invariant to uniform scaling and rotation, and partially invariant to affine distortion, illumination changes, and partial occlusion. The SIFT key points, which consist of the location, scale, and orientation information about objects, are first extracted from a set of reference images stored in a database. The scale-space extrema detector based on difference of Gaussians, which is an approximation to Laplacian of Gaussian (LoG), is employed for detecting the key points. Local extrema in scales are selected as the potential key points. The keypoint candidates are then further refined using contrast selection. Orientation was then assigned to the key points. A keypoint descriptor is constructed based on a 16x16 neighbourhood, which is then divided into 16 subblocks of 4x4 size. Using 8 bin orientation histograms results in a descriptor with a size of 128.

Three more feature extraction methods used to recognise the products on a shelf are considered in this work. The second selected method is the speeded-up robust features, namely SURF (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. The third selected method is the binary robust independent elementary features, BRIEF (Calonder et al., 2010). It is based on the keypoint location, which is the centre surrounding the extremas for real-time feature detection and matching (Agrawal et al., 2008). The fourth selected method is Oriented FAST and Rotated BRIEF ORB (Rublee et al., 2011), which is an efficient alternative to SIFT and SURF. It is based on a high-speed corner detector, namely the FAST keypoint detector (Rosten and Drummond, 2006; Rosten et al., 2010), which then uses the BRIEF binary descriptor. With the generated features for the key points, the matching method, Fast Library for Approximate Nearest Neighbours FLANN (Suju and Jose, 2017), is used to match the key points. After the matches are found, they are further filtered by using Lowe's ratio test, with a ratio of 0.7. The minimum number of matches was set to be 10.

In Algorithm 1, product P is an element of (A, A*, B, C, D, …., H) in Figure 17, A and A* are the first products labelled by EPT E1 on a particular shelf, and H is labelled by EPT E4, which is the last product on the shelf. The x-coordinates of the product boundaries of identical product items with the same subtype are represented by P_{beg} and P_{end}. For example, A and A* are completely identical adjacent items on the product shelf. The product items on the same shelf are followed by each EPT from left to right on each height level, then matched with the predetermined boundary location of the product P on the shelf. The x-coordinates of boundary positions of items priced the same and the same product item with different subtype are identified by L_{beg} and L_{end}, which are predetermined by the product label of product P on the same height level. For example, one product price label can include three product items A, A* and B which represent two different subtype A/A* and B of the same product priced the same corresponding to an identical price label with boundaries determined by L_{beg} and L_{end}. The correct positions of P_{beg} and P_{end} are determined by the computer vision system when a correct shelf product image is presented to the system. Similarly, the L_{beg} and L_{end} are obtained from the computer vision system.

Algorithm 1. Online verification of sub-graph isomorphism products on a shelf.

For a test image on the shelf and a reference template image of product *P*:

1) Load the centre coordinate (*Xc*, *Yc*) of EPL in Fig. 15 and find the position of each EPT on the same height level of the shelf
2) Perform keypoint detection, feature extraction, and feature matching using FLANN
3) Perform Lowe's ratio test to filter unreliable matched keypoints
   a) If a matched keypoint > threshold
      i) Perform homography to recover the bounding box location after matching product P with the test image and reference template image
      ii) If the bounding box is located within (Pbeg , Pend), product P is located at a correct location; otherwise check if the bounding box is located within (Lbeg , Lend); product P has the same price but different subtype
      iii) Otherwise located product P is of a different brand, or has a different price or different size of a misplaced item
      iv) Remove matched points that lie inside the bounding box and repeat Step 3a to locate the next product image of P (repeat objects of P)
   b) Otherwise product P is a removed item from the shelf
   c) If the next isomorphism product P exists on the left side or right side, go back to Step 2
4) Repeat Step 1 to locate the next EPT at the baseline-position

Each EPT can match the nearest neighbour of isomorphism products on a shelf using visual recognition methods. The VOMM transmits validated information to configure each EPT by LoS communication, so that each EPT can display the information of the products on the shelf, according to the information stored in the memory, such as the types, prices, daily promotion rate, valid time, etc.

3.5.1. Packet format with Manchester coding

For the application of short-range communication in ambient light, the transmission signal is vulnerable to light interference, noise attenuation and signal distortion. The unique characteristics of the Manchester coding signal, and the code containing the synchronisation feature can minimise the use of an I/O port for transmission, eliminate carrier frequency modulation, protect against light interference and reduce the battery power requirement of EPTs.

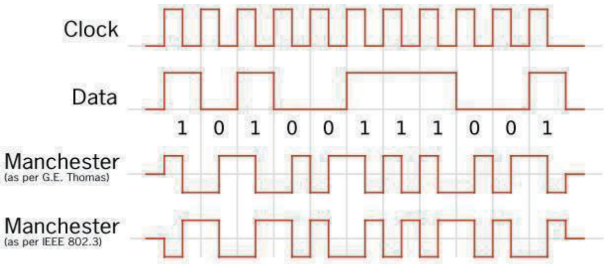


Figure 18. Manchester coding methods.

Manchester coding is a simple block code mapping scheme, which maps “0” and “1” into “01” and “10”, respectively, representing a rising edge and a falling edge in the transmission signal, respectively, as shown in Figure 18. Thus, the rising and falling edges represent the clock synchronisation function which can neglect the delay of signal and act as the carrier during the transmission in ambient light.

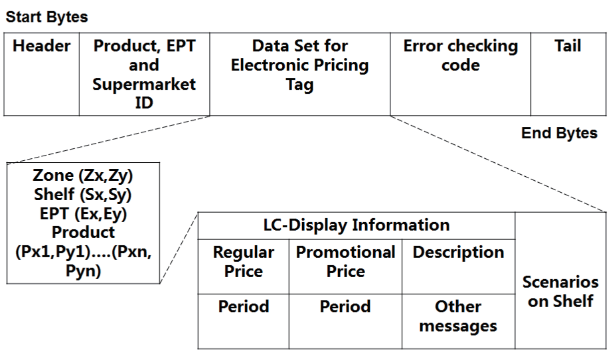


Figure 19. An example of a packet for LoS communication protocol between a VOMM and EPTs.

A packet for the LoS communication protocol, as shown in Figure 19, is defined for the store-front management system in a supermarket. The displayed information for EPTs includes product description, the sale period of regular and promotional prices, and the scenarios, as well as their actual location according to the coordinates of the zone (*Zx, Zy*), shelf (*Sx, Sy*), EPT (*Ex, Ey*), and product (*Px1, Py1*) …. (*Pxn, Pyn*) where n is the number of products on a shelf after auto-positioning of each product on the shelf. A customised message can be displayed by the LCD of an EPT module synchronised with a customer’s smartphone during their navigation.

An application of this system is to allow customers to search for the goods they want. Customers can enter the item they are looking for via a smartphone, and the store-front management system will then provide information about the location, price, availability, etc. of the products. In addition, with the monitoring system, the estimated number of items on a shelf can also be provided. Moreover, the IoT system can automatically set up a link between a VOMM and all EPTs on a shelf. Customised promotions can be shown by the EPTs for any product on a shelf during the navigation of individual customers via their smartphone.

4. Implementation and experiments

4.1. Positioning experiment of EPTs

According to the specification of the stepping motor in Table 4, the subdivision control of the stepping motor is proposed as follows: the control method and its structure bear a striking resemblance as they are the same regardless of what kind of stepping motor is employed.

The driving signal and current output was controlled by C-language software with the external hardware of a digital-to-analogue converter and current amplifier to drive each phase of the stepping motor. The step angle of the stepping motor was selected to be 1.8° for a full step, so the refined rotational angle per step, for X-Y plane segmentation with 32 subdivisions, can achieve AoA of about 0.05625°, where the transmission angle of the infrared laser emitter is ignored. The infrared laser emitter is locked on the shaft of the stepping motor so that the infrared beam can be rotated according to the step angle of the stepping motor. The specification of the infrared laser emitter is shown in Table 4; the laser lens with light filter has a diameter of 5 mm and the housing design covering the lens is about 22.4 mm in diameter, in order to avoid other light interference, as shown in Figure 9.

The measurement of inter-tag distance (*Xit*) defined in Figure 16 is based on the measurement distance (*D*) defined in Figure 13(b) between the VOMM and X-Y plane on the wall, as shown in Figure 20(a), and the ruler for measuring the distance (*D*) between the VOMM on the wall is 1, 1.5 m and 2 m respectively, as shown in Figure 20(b). The inter-tag distance is measured by using the ruler during the rotation of the VOMM by programming control; the spot of the red light beam is generated by the infrared laser emitter in the VOMM, and is marked on top of the ruler on the wall, as shown in Figure 20(c). In total, six testing samples were measured from the start point to the end point on the wall, at a pace of 10 steps, between the test samples, and the best result for the inter-tag distance (*Xit*) was achieved by 1 mm and 2 mm per subdivided step, with a measurement distance (*D*) of 1 m, 1.5 m and 2 m, as shown in Table 3(a). The methods (Blau, 2004; Tobias et al., 2012;

Hasanuzzaman, 2011) for positioning are compared with our proposed method, for single or multiple transmitters to localise the tagged object or receiver by measuring the distance between the transmitter and receiver along the Z-axis, as shown in Figures 13 and 14. The comparison of the inter-tag distance of the proposed positioning method with other methods is shown in Table 3(b). The results show that a novel positioning with configuration method between EPTs and the VOMM for store-front management systems has been developed.

Table 3. (a) Measurements from the start point to the end point per step at a pace of 10 steps between test intervals, and the measured inter-tag distance (*Xit*) with the measurement distance (*D*) between the VOMM and the wall; and (b) the comparison of inter-tag distance by different positioning methods.

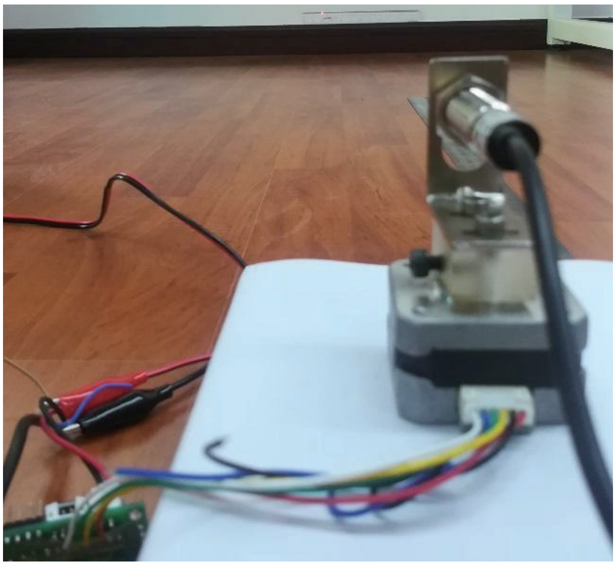
(a)

Measurement distance [*D*] (m)	**Drive 10 steps for each test sample**					**Result of Inter-tag Distance [*Xit*] (cm)**	**Verification of rotatable angle / step**
	Testing samples	**Start point per step (cm)**	**End point per step (cm)**	**10 steps / cm**	**Average of 10 steps (cm)**		
1	#1	0.22	1.29	1.07	1.03	0.10	0.06
	#2	2.49	3.51	1.02			
	#3	6.05	7.08	1.03			
	#4	9.55	10.57	1.02			
	#5	12.45	13.47	1.02			
	#6	16.03	17.06	1.03			
1.5	#1	1.00	2.55	1.55	1.56	0.16	0.06
	#2	7.05	8.60	1.55			
	#3	10.60	12.17	1.57			
	#4	14.99	16.52	1.53			
	#5	19.99	21.52	1.53			
	#6	24.55	26.15	1.60			
2	#1	0.40	2.40	2.00	2.03	0.20	0.06
	#2	5.10	7.15	2.05			
	#3	7.60	9.60	2.00			
	#4	11.30	13.40	2.10			
	#5	14.00	16.00	2.00			
	#6	17.75	19.80	2.05			

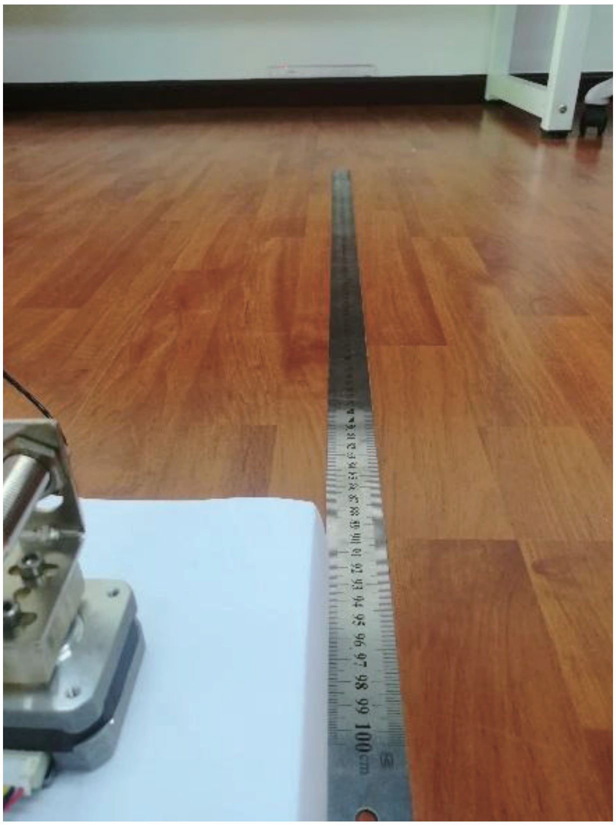
(b)

Transmitter types	**Receiver types**	**Measurement distance (Z-axis)**	**Inter-tag distance (X-axis)**	**References**
Radio frequency	Radio Frequency Identification (RFID) Tag	2 m	80 cm	Boontrai, D. (2009)
Radio frequency + Infrared	Radio-optical Tag	2 m	9 cm	Ashwin A. (2015)
Multiple LEDs	Visual Camera	2 m	0.9 cm	Guan, W. (2018)
Rotatable infrared laser emitter	**Infrared**	**2 m**	**0.2 cm**	**Our proposed method**

(a)



(b)



(c)

Figure 20. (a) Infrared laser emitter module in the rotatable VOMM; (b) the measurement distance (*D*) shown by the ruler between the VOMM and the wall; and (c)the measurement of inter-tag distance (*Xit*) shown by the red spotlight on the wall.

Table 4. Parameters of the experiment platform.

Low-power microprocessor specification	
Manufacturer Part No.:	GPL87103A
Type:	8-bit processor
SRAM	2432-byte
ROM	24K-byte
DPRAM	54 byte
RC oscillator for system	225k/500k/1000k/1800 kHz 32768 Hz
Operation voltage	DC1.8 V~3.6 V
Low standby current	< 1 uA @3.6 V, 25°C
General I/O ports	18 ports
LCD configurations	9 coms x 48 segs (MAX)
Operation temperatures	1.2 V-1.8 V @0~70°C 1.8 V–3.6 V @-20~70°C
Infrared laser emitter specification	
Manufacturer Part No.:	M18 Series
Operating voltage	DC10~30 V
Operating current	20 mA~35 mA
Output response time	3 ms ON and 1.5 ms OFF
Temperature range	-40~70°C
Bluetooth module specification	
Manufacturer Part No.:	GPBTM03x
RC oscillator for system	16 MHz Low power 32 KHz
Voltage converter	DC/DC
Temperature control	Yes
Battery monitor	Yes
Operation voltage	DC1.9 V~3.6 V
Temperature range	-40~85°C
Stepping motor specification	
Manufacturer Part No.:	42BYG228
Drive system:	Unipolar
Step angle:	1.8° full step
	0.9° half-step
Phase/Windings:	4/2
Voltage & current:	12 V at 400 mA
Resistance per phase:	30 ohms
Inductance per phase:	23 mH
Holding torque:	2000 g-cm
Detent torque:	220 g-cm max
Weight:	0.24 kg (0.5 lbs.)
Max continuous power:	5 W
Infrared receiver specification	
Manufacturer Part No.:	AX180PTC
Operation voltage	1.2 V
Temperature range	-40~80°C
Viewing angle	20°+/-5°

4.2. Visual recognition experiment

The images captured are in colour JPEG format, with an 8 megapixel camera, providing images of resolution 3,456 x 2,304. The hardware platform used was an ARM9 microprocessor (MCU), and the application software systems used were OpenCV and Python, for the development of novel visual recognition algorithms, on a Windows operating system.

This section presents the essential module of the computer vision system for product recognition. The whole computer vision system consists of several modules: fast projection module, fast colour space histogram processing module and high-accuracy keypoint matching algorithms for identification of products on a shelf. The automated keypoint generation and matching algorithm is shown to be invariant to size and front-back location of a shelf in this paper. Further work on a large number of products with different rotation angles has shown the approach to give very high accuracy for frontal product placement and good accuracy even for side views of products on the shelves.

4.2.1. Results of local feature matching of products on the shelves

For feature matching, the affine homography method (Calonder, 2010) was used to locate the reference template in the testing image. As per the example shown in Figure 17, there are eight template images, corresponding to the eight different products captured in the image taken against a white background. The eight different products are located from left to right on the shelf, denoted as A, A*, B, C, D, E, F, G and H (* is a repeated object) in Figure 17.

The search and detection of the first product on a shelf is carried out above each EPT, from left to right, as shown in Figure 21, and based on a graph-based image of the shelf as shown in Figure 22. Searching for an object to the left or to the right side will be stopped when either the last object or a different object is found. The object searching methodology is presented as Algorithm 1.

As the four visual recognition methods considered are based on greyscale, the products A, A* and B are recognised as the same type, which is represented by EPT #1. A result of finding the nearest neighbour of product A and recognising the repeated product A* is that the searching time can be reduced by about a half. Meanwhile, the computation time can be significantly reduced when the same type of products are placed together on a shelf. Product C and product D have different sizes, but belong to the same brand, while only product C is represented by EPT E2. Even though the products D, E and F have different wordings for different subtype, they are collectively represented by EPT E3. After searching the items described by EPT E3, products G and H are found to have different patterns, but belong to the same brand. However, no more EPTs are found on the same level of the shelf, as shown by the dotted line in Figure 21. An error message will then be sent to the store-front management system by the VOMM.

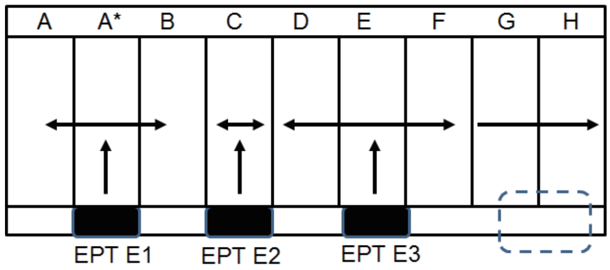


Figure 21. The sequence of product searching and detection on each level of a shelf.

Figure 22. Matching of product items on a shelf with the EPTs below it, and identification of all the EPTs and their respective coordinates (start, end) determined by auto-positioning in section 3.4: EPT *E1* (1696, 2136), EPT *E2*(2660, 3032), EPT *E3* (3825, 4262).

Figure 23 shows the matching of the sample products on a shelf. Keypoints are automatically generated by the four computer vision algorithms to represent the local features of the images working similarly like the fingerprint of a thumb. However, the keypoints from images are more noisy and prone to mismatch where usually a hypothesis of how many matching keypoints there are is needed to correctly identify an image. The matching results based on SIFT, SURF, BRIEF, and ORB are shown in Figures 23(a) to (d), respectively. Table 5 shows the keypoint statistics of the SIFT matching with the testing image. Although most of the objects are correctly identified, the bounding box of product G determined by SIFT is quite different from a rectangle. Referring to the number of features that are matched, template G has the smallest number of matched keypoints, which is 24 only. The results based on SURF are tabulated in Table 6. We can see that the number of keypoints generated are generally two times more than that by SIFT. From the results shown in Table 7, the number of features generated by the BRIEF method is the smallest. For the keypoint statistics in Table 8, the ORB method generates the largest number of keypoints. To produce correct results, the ORB method needs to increase the upper bound to over 5,000 keypoints; otherwise a smaller number of keypoints would lead to incorrect matching of some of the products.

The number of matched keypoints of A actually includes two shampoos, which are labelled as A and A*. After locating the bounding box of homography A, the keypoints inside the bounding box are removed. So only around half of the keypoints are left and then computed with other homography products again.

Analysing the keypoint statistics of the store-front management of product items on the shelves from Table 5 to Table 8, the BRIEF keypoint extraction algorithm has the smallest number of template keypoints as well as the smallest number of test image keypoints, while all products in the test image are correctly matched with their respective templates. In terms of computation requirements, the number of keypoints from the BRIEF method for the test image is only 32% of those of SIFT and 16% of those from SURF. The number of keypoints from the BRIEF method for template H is only 29% of those from SIFT and 11% of those from SURF. Furthermore, the computation time and storage space is much reduced because each template image is reduced to less than two hundred keypoints. The keypoint matching processing time is proportional to the number of extracted keypoints in both the test image and the image template.

(a)

(b)

(c)

(d)

Figure 23. (a) SIFT method matching results with test images marked by white frames; (b) SURF method matching results with test images marked by white frames; (c) BRIEF method matching results with test images marked by white frames; and (d) ORB method matching results with test images marked by white frames.

The results of feature matching are shown in the following table for the SIFT, SURF, BRIEF and ORB methods:

Table 5. Keypoint statistics for the SIFT method.

Shampoo marks	**Template keypoints**	**Test image keypoints**	**Pass ratio test / Repeat object test***
A	737	8,102	190
A*	737	8,102	87*
B	614	8,102	293
C	705	8,102	296
D	395	8,102	129
E	405	8,102	155
F	315	8,102	106
G	128	8,102	24
H	182	8,102	72

Table 6. Keypoint statistics for the SURF method.

Shampoo marks	**Template keypoints**	**Test image keypoints**	**Pass ratio test / Repeat object test***
A	1,065	16,016	149
A*	1,065	16,016	62*
B	999	16,016	432
C	1,006	16,016	305
D	817	16,016	207
E	780	16,016	240
F	728	16,016	168
G	383	16,016	72
H	500	16,016	175

Table 7. Keypoint statistics for the BRIEF method.

Shampoo marks	**Template keypoints**	**Test image keypoints**	**Pass ratio test / Repeat object test***
A	174	2,576	31
A*	174	2,576	17*
B	111	2,576	44
C	196	2,576	79
D	114	2,576	34
E	128	2,576	55
F	84	2,576	22
G	61	2,576	13
H	53	2,576	24

Table 8. Keypoint statistics for the ORB method.

Shampoo marks	**Template keypoints**	**Test image keypoints**	**Pass ratio test / Repeat object test***
A	2,551	36,767	105
A*	2,551	36,767	60*
B	2,442	36,767	274
C	2,513	36,767	244
D	1,408	36,767	116
E	1,599	36,767	184
F	1,163	36,767	108
G	567	36,767	36
H	804	36,767	107

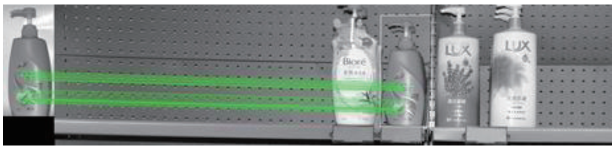
4.2.2. Matching results of multi-layer products placed on a shelf

The SURF method is able to recognise product locations when the products are placed at the front of a shelf. Once a product in front is removed, inner products can be seen and will become at the front on the shelf. Hence, three rows of products placed on a shelf are considered, which are defined as the outermost row, the middle row, and the innermost row, as shown in Figure 24.

According to the results of the feature matching experiments, the three rows of the same product can be located correctly. The size of the bounding box and the number of matched keypoints between the template and the test image are shown in Table 9. We can see that the size of the bounding box becomes smaller for products placed at the rear of the shelf.

(a)

(b)



(c)

Figure 24. (a) Recognised product in the outermost layer; (b) recognised product in the middle layer; and (c) recognised product in the innermost layer.

Table 9. Bounding box and location statistics of product matching of product placement on shelf location of outermost, middle and innermost positions on a shelf.

Layers	Box length	Box height	Template keypoints	Test image keypoints	Pass ratio test
Front	419.5	878.2	434	8007	74
Middle	416.4	893.0	434	8852	88
Back	390.0	831.9	434	9238	83

Figure 25 shows the matching results of missing products using the SURF method when the original designated product is removed from the shelf. However, a product with the same brand logo but a different subtype is found in the shelf location and the logo brand name was matched to the template product. In this particular case, 22 feature keypoints were correctly matched. To handle this situation, the following two rules can be adopted. First, check the x-axis position of the matched point to recognise whether they are in the position of the brand name. Second, compute the class conditional probability P (product is designated product | number of keypoints matched) and the class conditional probability of products with different subtype P (Product is Different subtype | number of matched points) and using the Naive Bayes classifier to determine the product that maximises the class conditional probability. In the first experiment shown with one template and one set of test images, all products in test images are correctly identified. In the second experiment, each shelf image contains 38 products and three images are taken under different lighting conditions. Three-fold cross validation is used where one image is used for training templates and two images are used for testing. The total number of training product images is 38. The total number of test product images is 76, repeated three times. The average accuracies of six different keypoint-based matching methods are from 94% to 97%.

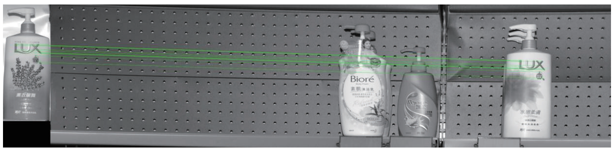
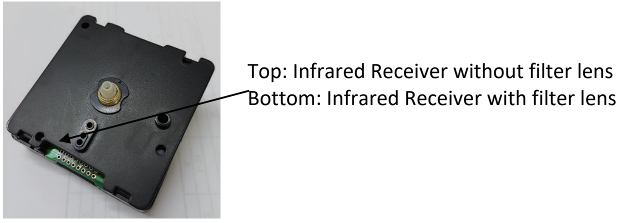


Figure 25. Matching of removed products using the SURF method.

4.3. Auto-configuration experiment of EPT

A simulated module with built-in IR receivers with or without a filter lens for testing is shown in Figure 26(a). A receiver without a filter lens may sometimes be affected by ambient light or sunlight. The module is operated by DC3 V power and the specification of the IR Receiver is shown in Table 4. The filter lens of the IR receiver has a diameter of 3 mm. The housing design can minimise the receiving angle and avoid other light interference as shown in Figure 26(a). The Cathode Ray Oscilloscope (CRO) demonstrates the results when a signal is received from the infrared laser emitter. While the spot of the red light beam does not reach the infrared receiver, the CRO without any signal received exhibits a high voltage, as shown in Figure 26(b). On the other hand, the CRO with a sudden signal received is shown in Figure 26(c), when the spot of the red light beam reaches the infrared receiver of the simulated module. The Manchester coding signal containing the EPL data transmission using infrared laser beam demonstrated by CRO is shown in Figure 27.

(a)



(b)

(c)

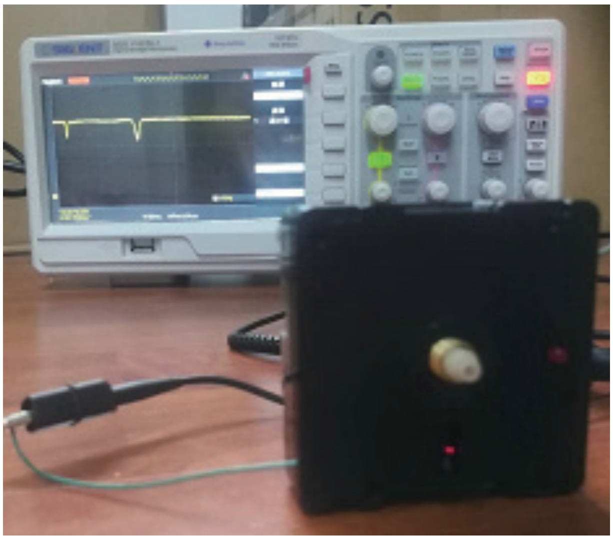


Figure 26. (a) Simulated module with a plastic light guide covering both infrared receivers; and (b) and (c) the testing results are shown by CRO: the emitter signal is not received as shown in (b) and the signal is received as shown in (c).

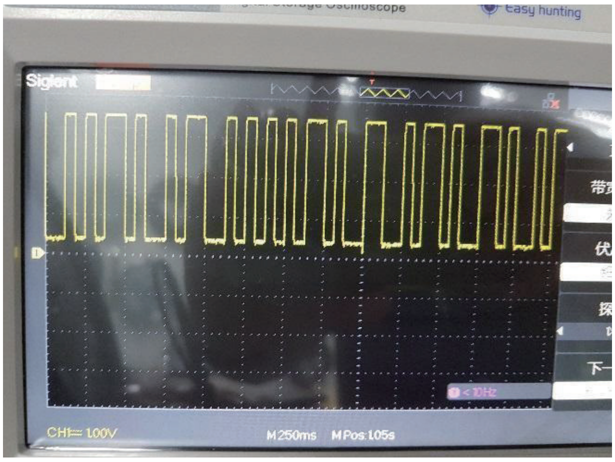


Figure 27. Manchester Coding Signal Transmission using an infrared laser beam.

5. Conclusion

A low-cost store-front management system has been proposed, which connects the inventory and pricing management systems for an unmanned supermarket. The novelty of the visual-optical approach is that it can position and configure all EPTs automatically by using VOMMs. The matching between EPTs and products on the shelf is performed using the affine homography method, with several visual recognition methods based on the SIFT, SURF, BRIEF and ORB features. The automatic pricing system can be extended to each non-tagged product via EPTs on shelves, thereby allowing thousands of products to be placed freely anywhere on shelves. Thus, the man-hour setup cost can be reduced significantly, and the operators may immediately handle and take action in response to abnormal conditions for real-time store-front management. This may also minimise any mistakes of EPTs and product placement on shelves.

Furthermore, the size of the bounding box for a product can obviously distinguish between an outermost and an innermost item, as mentioned in Table 9. The quantity of products can be accurately counted on the shelves. It is not necessary for operators immediately to fill up space on a shelf which is not empty, so this can increase the usage rate of the shelves.

5.1. Further investigation of retail shelf management

For future research and investigation, the novel visual-recognition methodology can improve the runtime and recognition accuracy when dozens of items are on a shelf. In this paper, a total of six scenarios are identified, which include 1) correct item, 2) removed item, 3) identification of misplaced item, 4) shifted item, 5) inverted item, and 6) empty space for an out-of-stock item. Some recent literature (Marder et al., 2015; Azad et al., 2009; Saran et al., 2015) has employed visual recognition methods, and their results were based on 1) correct items, 2) removed items, and 3) misplaced items without identification as shown in Figure 28. Then, the quantity of an out-of-stock product is roughly counted. Future research on visual recognition can minimise the computation time and achieve higher accuracy for more different types and a larger quantity of products on the shelves.

Figure 28. Investigate the computation time and recognition accuracy of 39 spaces of product item location on the shelves simultaneously.

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

[2] Ashok A, Xu C, Vu T, Gruteser M, Howard R, Zhang Y, Mandayam N, Yuan W and Dana C (2015). *Low-Power Radio-Optical Beacons for In-View Recognition*. In: Proceedings of 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall). IEEE, pp. 1 – 5.

[3] Azad P, Asfour T and Dillmann R (2009). *Combining Harris interest points and the SIFT descriptor for fast scale-invariant object recognition*. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 4275-4280.

[4] 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.

[5] Beul M, Droesche D, Nieuwenhuisen M, Quenzel J, Houben S and Behnke S (2018). Fast Autonomous Flight in Warehouses for Inventory Applications. *IEEE Robotics and Automation Letters*, 3(4), pp. 3121 - 3128.

[6] Blau J (2004). Supermarket's futuristic outlet. *IEEE Spectrum*, 41(4), pp. 21 - 25.

[7] Boontrai D, Jingwangsa T and Cherntanomwong P (2009). *Indoor Localization Technique using Passive RFID Tags*. In: Proceedings of IEEE International Conference on Communications and Information Technology. IEEE, pp. 922 – 926.

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

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

[10] Chou T and Wykes C (1997). An Integrated Vision/Ultrasonic Sensor for 3D Target Recognition and Measurement. In: Proceedings of 1997 Sixth International Conference on Image Processing and Its Applications. IEEE, pp. 189-193.

[11] Evans J, Shober R, Wilkus S and Wright G (1996)., A low-cost radio for an electronic price label system. *Bell Labs Technical Journal*, 1(2), pp. 203 – 215.

[12] Fortin-Simard D, Bouchard K, Gaboury S, Bouchard B and Bouzouane A (2012). *Accurate passive RFID localization system for smart homes*. In: Proceedings of 2012 IEEE 3rd International Conference on Networked embedded Systems for Every Application (NESEA). IEEE, pp. 1 – 8.

[13] Fowers S and Lee D (2012). An Effective Color Addition to Feature Detection and Description for Book Spine Image Matching. *International Scholarly Research Notices*, 2012, pp. 1-15.

[14] Gimpilevich Y and Savochkin D (2013). *RFID Indoor Positioning System Based on Read Rate Measurement Information*. In: Proceedings of 2013 IX Internatioal Conference on Antenna Theory and Techniques. IEEE, pp. 546 - 548.

[15] Guan W, Chen X, Huang M, Liu Z, Wu Y and Chen Y (2018). *High-Speed Robust Dynamic Positioning and Tracking Method Based on Visual Visible Light Communication Using Optical Flow Detection and Bayesian Forecast*. IEEE Photonics Journal, 10(3), pp. 1-12.

[16] Guo H and Li J (2014). *Research on the Application of Intelligent Campus Supermarket System -- Based on the Internet of Things (IOT) Technology*. In: Proceedings of 2014 Seventh International Symposium on Computational Intelligence and Design. IEEE, pp. 390-394.

[17] Harris A, Khanna V, Tuncay G and Kravets R (2016). Smart LaBLEs: *Proximity, Autoconfiguration, and a Constant Supply of Gatorade(TM)*. In: Proceedings of 2016 IEEE/ACM Symposium on Edge Computing (SEC). IEEE, pp. 142-154.

[18] Hasanuzzaman F, Tian Y and Liu Q (2011). *Identifying Medicine Bottles by Incorporating RFID and Video Analysis*. In: Proceedings of 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW). IEEE, pp. 528 – 529.

[19] Hong W, Samadani R and Baker M (2013). Line-of-Sight Signaling and Positioning for Mobile Devices. In: Proceedings of 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). IEEE, pp. 1 – 6.

[20] Jämsä J, Luimula M, Pieskä S, Brax V, Saukko O and Verronen P (2010). *Indoor Positioning with Laser Scanned Models in the Metal Industry*. In: Proceedings of 2010 Ubiquitous Positioning Indoor Navigation and Location Based Service. IEEE, pp. 1 – 9.

[21] Kejriwal N, Garg S and Kumar S. (2015). *Product counting using images with application to robot-based retail stock assessment*. In: Proceedings of 2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA). IEEE, pp. 1 – 6.

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

[23] Kusari A, Pan Z and Glennie C (2014). *Real-time Indoor Mapping by Fusion of Structured Light Sensors*. In: Proceedings of 2014 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS). IEEE, pp. 213 – 219.

[24] Lowe D (2004). Distinctive image features from scale-invariant keypoints. *J. Computer Vision*, 60(2), pp. 91-110.

[25] Lu X, Liu H and Liu F (2014). *A Novel Algorithm for Enhancing Accuracy of Indoor and Position Estimation*. In: Proceedings of the 11th World Congress on Intelligent Control and Automation. IEEE, pp. 5528 - 5533.

[26] Ma Z, Poslad S, Hu S and Zhang X (2017). A fast path matching algorithm for indoor positioning systems using magnetic field measurements. In: Proceedings of 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, pp. 1-5.

[27] 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), pp. 1-11.

[28] Melek C, Sonmez E and Albayrak S (2017). *A survey of product recognition in shelf images*. In: Proceedings of 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, pp. 145-150.

[29] Murata S, Yara C, Kaneta K, Ioroi S and Tanaka H (2014). Accurate Indoor Positioning System Using Near-Ultrasonic Sound from a Smartphone. In: Proceedings of 2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies. IEEE: pp. 13-18.

[30] Phan D, Oh C, Kim S, Na I and Lee C (2013). *Object Recognition by Combining Binary Local Invariant Features and Color Histogram*. In: Proceedings of 2013 2nd IAPR Asian Conference on Pattern Recognition. IEEE, pp. 466 – 470.

[31] Rosten E and Drummond T (2006). *Machine learning for high-speed corner detection*. In: Proceedings of European conference on computer vision. Springer, pp. 430-443.
[32] 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.
[33] Rublee E, Rabaud V, Konolige K and Bradski G (2011). *ORB: An efficient alternative to SIFT or SURF*. In: Proceedings of Computer Vision (ICCV). IEEE, pp. 2564-2571.
[34] Rykowski J (2017). *Multi-dimensional identification of things, places and humans*. In: Proceedings of 2017 IEEE International Conference on RFID Technology & Application (RFID-TA). IEEE, pp. 152-157.
[35] Salomon R, Schneider M and Wehden D (2006). Low Cost Optical Indoor Localization System for Mobile Objects without Image Processing. In: Proceedings of 2006 IEEE Conference on Emerging Technologies and Factory Automation. IEEE, pp. 629-632.
[36] Saran A, Hassan E and Maurya A (2015). Robust visual analysis for planogram compliance problem. In: Proceedings of 14th IAPR International Conference on Machine Vision Applications (MVA). IEEE, pp. 576-579.
[37] Savochkin A, Mickhayluck Y, Iskiv V, Schekaturin A, Lukyanchikov A, Savochkin D and Levin E (2014). *Passive RFID System for 2D Indoor Positioning*. In: Proceedings of 2014 20th International Conference on Microwaves, Radar and Wireless Communications (MIKON). IEEE, pp. 1 – 3.
[38] Suju D and Jose H (2017). *FLANN: Fast approximate nearest neighbor search algorithm for elucidating human-wildlife conflicts in forest areas*. In: Proceedings of 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN). IEEE, pp. 1-6.
[39] Takahashi Y, Y. Li, Jian M, Jun W, Maeda Y, Takeda M, Nakamura R, Miyoshi, Takeuchi H, Yamashita Y, Sano H and Masuda A (2013). *Mobile robot self localization based on multi-antenna-RFID reader and ICtag textile*. In: Proceedings of 2013 IEEE Workshop on Advanced Robotics and its Social Impacts. IEEE, pp. 106 – 112.
[40] Tao Q, Zhong C, Lin H and Zhang Z (2018). Symbol Detection of Ambient Backscatter Systems With Manchester Coding. *IEEE Transactions on Wireless Communications*, 17(6), pp. 4028-4038.
[41] Tobias T, Janson M, Zetik R and Thielecke J (2012). Infrastructureless indoor mapping using a mobile antenna array. In: Proceedings of 2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE, pp. 36-39.
[42] Tonioni A and Di Stefano L (2017). *Product Recognition in Store Shelves as a Sub-Graph Isomorphism Problem*. In: Battiato S et al. (eds). Image Analysis and Processing. Springer, pp. 682-693.
[43] Vargheese R and Dahir H (2014). An IoT/IoE enabled architecture framework for precision on shelf availability: Enhancing proactive shopper experience. In: Proceedings of 2014 IEEE International Conference on Big Data (Big Data). IEEE, pp. 21-26.
[44] Wang C and Cheng L (2011). *RFID & vision based indoor positioning and identification system*. In: Proceedings of 2011 IEEE 3rd International Conference on Communication Software and Networks. IEEE, pp. 506-510.
[45] Wang H, Zhang Z, Chen X and Wang Y (2017). *Stepper motor SPWM subdivision control circuit design based on FPGA*. In: Proceedings of 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). IEEE, pp. 889-893.
[46] Yang X, He D, Huang W, Ororbia A, Zhou, Kifer D and Giles C (2017). *Smart Library: Identifying Books on Library Shelves Using Supervised Deep Learning for Scene Text Reading*. In: Proceedings of 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL). IEEE, pp. 1 – 4.