Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

Kejriwal, Nishant, Garg, Sourav, & Kumar, Swagat
(2015)
Product counting using images with application to robot-based retail stock assessment.
In *Proceedings of the 2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA 2015).*
Institute of Electrical and Electronics Engineers Inc., United States of America.

This file was downloaded from: https://eprints.qut.edu.au/204527/

© IEEE 2015

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

https://doi.org/10.1109/TePRA.2015.7219676

Product Counting using Images with application to Robot-based Retail Stock Assessment

Nishant Kejriwal, Sourav Garg and Swagat Kumar
Innovation Lab, Tata Consultancy Services, New Delhi, India
Email: { nishant.kejriwal, sourav.garg, swagat.kumar}@tcs.com

***Abstract*—In this paper, we propose a novel method for obtaining product count directly from images recorded using a monocular camera mounted on a mobile robot. This has application in robot-based retail stock assessment problem where a mobile robot is used for monitoring the stock levels on the shelves of a retail store. The products are recognized by carrying out a nearest-neighbor search in the template feature space using a k-d tree. Unlike current approaches which only provide approximate stock level, we propose a method which can compute the exact number of discrete products visible in a given image. The product count is obtained by fitting bounding box around each product and removing them sequentially from the image. A second stage of grid-based search is carried out in the neighborhood of each detected product to detect new products which were missed out in the previous step. This detection is based on a confidence measure that includes various information such as histogram matching and spatial location. The efficacy of the proposed approach is demonstrated through experiments on different datasets obtained using robot camera as well as mobile phone camera. These results show that the robot-based retail stock assessment may become a viable alternative to the currently prevailing manual mode of carrying out these surveys.**

***Index Terms*—Retail Robotics, stock assessment, product counting, OOS, object recognition, service robotics**

I. Introduction

In this paper, we look into the problem of carrying out stock monitoring and assessment in retail stores using mobile robots [1] [2] [3]. The robot uses on-board cameras to capture video that contains the images of the shelves on either side of the robot. These images are processed, either on-board or on a remote server, to generate statistics of the products on the shelf and detect various situations like out-of-stock (OOS), misplaced items etc. An illustration of robot-based retail stock assessment system is shown in Figure 1. The robot may carry a pair of cameras that can move up and down on a shaft or may carry multiple cameras placed at different heights. Use of robots may not only reduce the cost of such surveys, but also increase the accuracy of data collected by avoiding human related factors.

The robot has to identify various products, know their location based on a given planogram and detect incidents like out-of-stock situations and misplaced items. A number of methods have been proposed to solve this problem. For instance, Zimmerman [3] decodes a product barcode from the shelf image. It then retrieves the product image from a database and segments the shelf image to match with the retrieved image. If no match is found, out-of-stock flag is set. On the other hand, Gokturk [4] uses camera and multiple lighting sources to compute occupancy in a enclosed compartment using triangulation methods. It also suggests using depth sensors or stereo-vision system for occupancy measurement. There are other patents such as [5], [6] which talk about generic systems that can identify products, generate planogram, detect out-of-stock situations and provide percentage occupancy of products.

In this paper, we look into the problem of obtaining accurate product count directly from images recorded using on-board camera. We do not use depth sensors, stereo-vision system or any other range measuring device like IR or laser for obtaining the product count. We are interested in counting the number of products which are visible in a given image. The method involves two steps - in the first step, the product category or label is identified and in the second step, the product count is estimated.

The product is identified using interest point features like SURF [7]. A k-d tree is created in the feature space comprising of SURF descriptors from all the product templates. For each query image, a nearest neighborhood search is carried out in the descriptor space to identify the matching product templates. We provide two methods for obtaining the product count. The first method involves computing *feature repeatability* for each product which is counting the maximum number of times a particular feature is repeated in a given image. This factor is more or less proportional to the number of products present in the image. The second method consists of obtaining the bounding box for each identified product by using homography coupled with RANSAC [8] and removing them sequentially. A second stage of search based on histogram matching is employed to detect those products which were left out in the previous step. This search is performed by creating a 3×3 grid around each detected product. More will be discussed in the later sections of this paper. This second method provides not only product count but also product arrangement in a given shelf.

The main contributions made in this paper are as follows: (1) We provide two novel methods for obtaining accurate product count from images. (2) We have provided performance evaluation over different test cases and carried out experiment with actual robots to demonstrate the utility of the proposed approach. This is in contrast to other works such as [1] [6], where authors have reported systems with similar capabilities but, do not provide either the method description or performance evaluation. To our knowledge, such results for retail stock assessment is not yet reported in the literature.

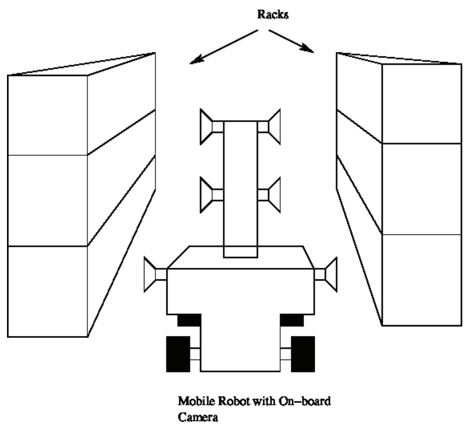

Fig. 1. Retail stock assessment using robots

The rest of this document is organized as follows. We provide a brief literature survey of related work in the next section. The proposed method for identifying and counting products is provided in Section III. The details of actual experiment and analysis of various results are provided in Section IV. The summary and future directions are provided in Section V.

II. Related Works

The problems faced by large retail stores today is well documented in the literature [9] [10] [11]. Some of these challenges include, frequent out-of-stock situations, product misplacement, organized retail crime including theft, lower profit margins due to stiff competition. High manpower costs makes it difficult to deploy more people required for efficiently managing the stores. This has prompted researchers to look for technologies which can be utilized to improve the current store management practices. Use of RFID based smart shelves [12] [13] is one such example. Using mobile robots for retail monitoring is a new concept which has been pioneered by Priya Narasimhan's group at CMU with their AndyVision project [1]. An implementation of robot-based retail monitoring system is demonstrated by Kumar *et al.* [2]. This work does not provide any algorithm for recognizing products and obtaining product count.

Apart from this, there are a number of patents that focus on methods where the images acquired and then processed to detect various stock levels. For instance, Birch and Kasper [14] a camera takes images of the product location, and finds the difference in the current and last image to find out if a product has been restocked or removed. Similarly, in [15] wireless cameras (sensor nodes) are placed at product locations. They send images to a central server which analyzes these images to detect the change in the stock level. This is done by comparing the images with their previously taken image. The patents [16] [5] focus on methods which generate planogram through image processing. The images could be processed over a remote server and it may be compared with a target planogram to detect misplaced items.

The patent [17] presents a method where the images acquired through a static or moving camera is used to assess the stock level. An object is identified with an "optically identifiable characteristic" which is unique to the location of the object. The patent [4] talks about a robot which moves around in an environment with a camera mounted on it. Other cameras are placed inside the room and triangulation is done to find depth of the obstacles as well as objects inside the room. An image of empty room is taken as a reference and the occupancy is detected by comparing the data. In [18], a smart bookshelf is proposed which uses a pair of cameras to identify if the books are added or removed through background subtraction. The patent by Groenevelt *et al.* [5] talks about a system which captures images and processes them over a remote server to extract planogram. It claims to detect partial and full stock depletion apart from detecting different orientation of a given product. In [3], a mobile robot is used performing inventory of products using images. It decodes barcode of a product from the shelf image and retrieves the product template image from the database and tries to match with at least one of the products in the shelf. If no match is found, the out-of-stock flag is set. In [19], author is identifying products using image analysis. User queries the image and process returns the candidate images based on similarity and product features. The patent by Limer *et al.* [20] describes a system that can count discrete units (such as tablets) using a camera and an illuminated stage. The light provides discrimination between a background field and a quantity of imageable units. The patent by Hofman [6] describes an image based system which can identify products, provide its count, detect OOS situations and provide arrangement of products. They make use of OCR to identify text in the logos and multiple features such as SURF and colour to identify products. However, they don't provide the details of the approach for product counting and do not provide any performance evaluation for their method.

Based on this study, one can surmise that robot-based retail stock assessment is a comparatively new problem which has received a good amount of attention in the research community in the past couple of years. It is fraught with several challenges and there is a need to develop reliable algorithms which can make it a viable alternative to currently employed manual mode of operation. The current work aims at moving closer to this realization by proposing an image-based product counting algorithm. The details of the algorithm is explained next in this paper.

III. The Methods

The method for the product counting is shown in the form of a flow chart in Figure 2. It primarily involves two steps. The first step aims at recognizing the product through template matching and while the second steps aims to obtaining the product count. These methods are described next in this section.

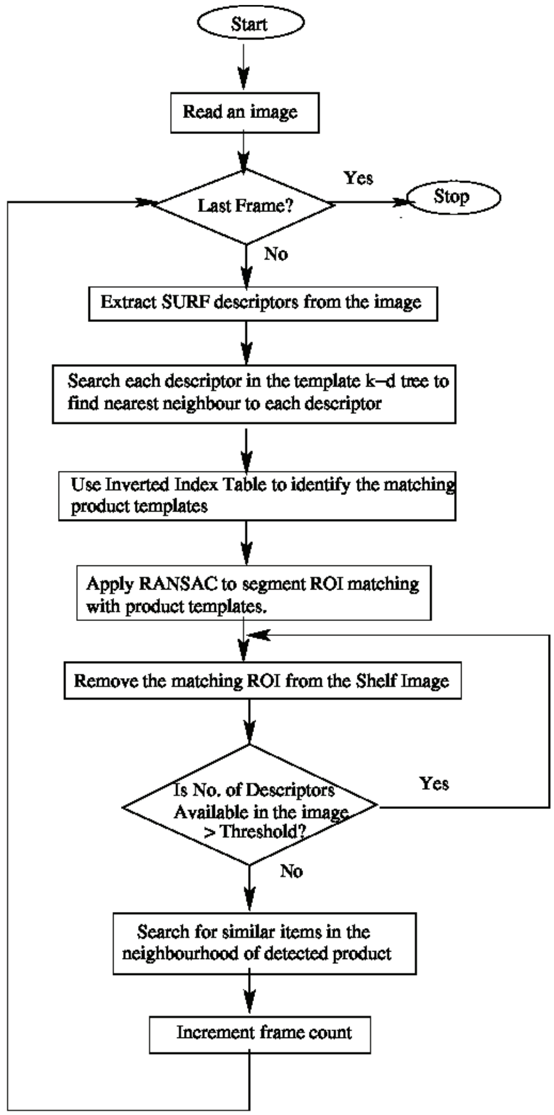

Fig. 2. Flowchart for counting products using SURF

A. Product Recognition using k-d tree

A k-d tree is created in the SURF descriptor space for all product templates. This step is carried out off-line. For a given query image, the matching descriptors are obtained using nearest neighborhood search. All the neighbors in the tree which satisfies an user-defined threshold on the distance ratio [21] are considered to be the valid matches. The product template associated with each neighboring node in the tree is obtained using an inverse index table. A product is declared to be found if it contains a number of matching descriptors above a given threshold.

B. Methods for Product counting

In this paper, we focus on counting the products which are lying on the front row of the shelves and are visible in the camera images. We do not use stereo-camera system or range sensors to compute the shelf occupancy as has been done by other researchers like [4]. Obtaining discrete product count is useful for high valued items and bigger products which are easy to detect using image processing algorithms. This product count is also sufficient for detecting out of stock (OOS) situations and misplaced items provide a planogram is available.

1) Product count using feature repeatability: In this method, we compute the maximum repeatability of a given product feature. Repeatability of a descriptor is the number of times this particular feature is repeated in a given set. It is based on the observation that the same SURF descriptors will get repeated if there are multiple products of the same type. Counting the number of times these features are repeated can provide a clue about the actual number of products present in the query frame. This approach is fast and easily implementable. The method is also robust to rotation or scaling effects. This method, however, relies on finding at least one descriptor for all the products. It is also prone to noise and hence one has to fine tune the distance threshold to remove wrong observations.

2) Product counting using SURF and Color: In the second method, we try to fit a rectangular bounding box around each product and then count the number of such boxes found to know the product count. The bounding box for each product is obtained by using SURF correspondence and homography. This is more robust when sufficient number of descriptors are available for a given product. When descriptors are available, we do a local search based on colour histograms around the detected products. The method is described in the flowchart shown in Figure 2. The steps involved are further illustrated in Figure 3. The first step involves extracting SURF descriptors from a query image. At the end of step 2, the products are recognized as explained in Section III-A. For each recognized product, we use SURF correspondence along with RANSAC to detect each product and remove them sequentially as shown in Step 3. If sufficient number of descriptors are not available, it might not be possible to detect a product as shown the right hand side image of the step 3.

In the next step, we do a grid search in the neighborhood of each of the detected product to find additional products not detected in the previous step. The cells which overlap with other detected products are removed from further consideration. This constitutes the step 4 of the method.

The remaining cells are matched with the centre product through colour histogram matching and the non-matching cells are further removed from consideration. The cell that satisfies the histogram matching threshold with the centre cell is assigned the label of the centre cell. This constitutes the step 5 of the method. In step 6, we use SURF correspondence and homography to fit the rectangular bounding box around the newly detected ROI. The number of ROIs thus detected provides the number of products present in the image.

The second method is computationally more complex, but provides better improvement in few cases where not enough SURF features are available.

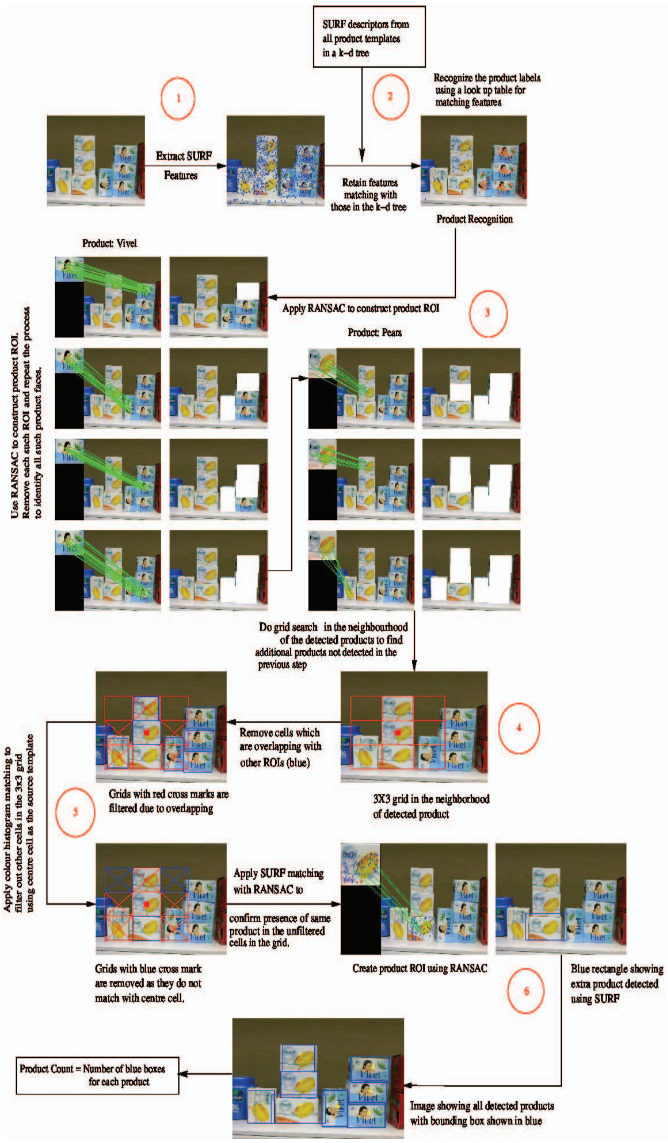

Fig. 3. Explaining the method of product counting through pictures.

IV. Experiment Results

Our robotic system consists of a Turtlebot 2 robot with an on-board USB camera facing the rack on either side of the aisle. The entire software is implemented using ROS [22] software framework. The image processing is carried out using OpenCV [23]. The images are collected at a speed of 15 frames per second. The robot moves at a speed of 0.1 m/s to avoid blurring of images. The accuracy of product recognition using SURF is 100%. In other words, we are able to identify a product if it is present on the rack and is easily identifiable under ambient illumination. The product counting is carried out using two methods - one based on descriptor repeatability and other making use of colour along SURF descriptors. The dataset D1 and D2 are collected using camera mounted on a mobile robot while the datasets D3 and D4 are recorded using a mobile phone. So, the later videos have other effects like scaling and change in view angles. In dataset D2, the products of same category may have different orientations. The precision-recall performance of our algorithm on these datasets is provided in Figure 4. This figure shows the best precision and recall obtained by varying various user defined parameters. Some of the instances of product detection and counting is shown in Figure 5. The first two rows show the images where the products of one kind have same orientation. The third row shows the case where products have different depth levels. The fourth row shows the case where products of one kind may have different orientations. The last row shows the case where the images have been taken from a mobile phone.

The summary of performance evaluation for our approach on four different datasets is provided in Table I. As seen in this table, the second method provides lesser recall compared to the first for first 3 datasets. However, for last dataset, the second method provides better recall than the first. The second method is important due to three reasons - first, it fits a bounding box around each product and hence makes it possible distinguish one product from the other. Second, if bounding box could be found for some products, localized search using colour histogram could be carried out in the neighborhood regions to locate additional products which were missed out in the first step. Third, this helps in obtaining the arrangement of products in a given shelf, which is not obtainable in the first method which provides product count based on feature repeatability.

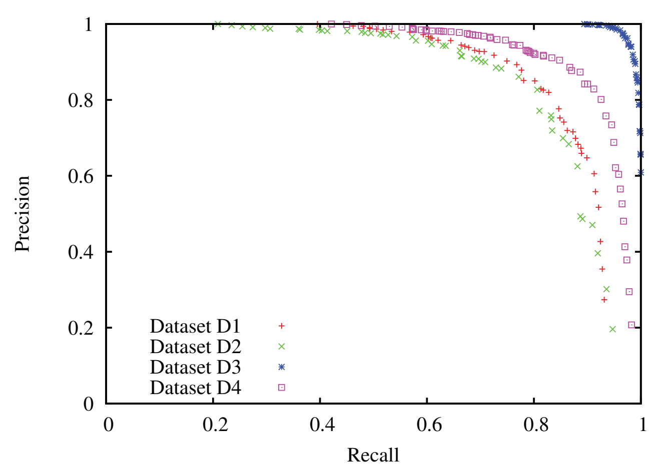

Fig. 4. Precision-Recall curve for product counting method I on different datasets. This is best performance obtained by varying the user-defined parameters in the algorithm.

V. Conclusion

Carrying out stock assessment using robots is still fraught with several challenges. One of the challenge is to reliably detect the stock level on the shelves. Low cost of visual sensors have encouraged people to generate several meaningful statistics by processing images. In this work, we show that it is possible to obtain very high level of precision in obtaining product count using features like SURF and colour histogram.

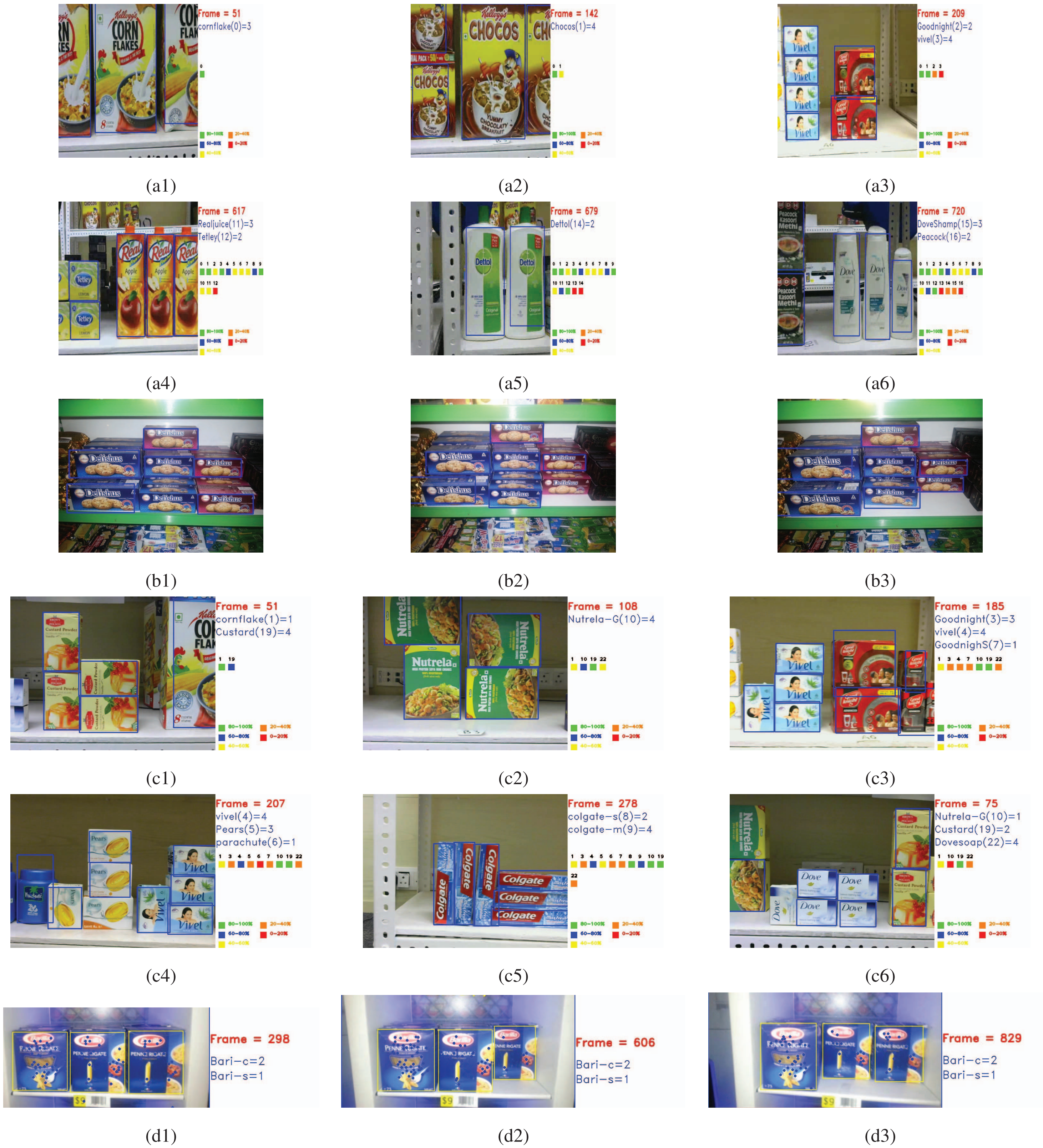

Fig. 5. Product count obtained from various Shelf Images. (a1)-(a6): Images have been recorded from a webcam mounted on a mobile robot running at a speed of 0.5 m/s. (b1-b3) Images have been taken from a hand held camera. These views have different zoom level and orientation. (c1)-(c6): Products with rotated front face. Some of the cases where the number of features is low, the product may not get detected.(d1-d3): Images are taken using mobile phone camera.

TABLE I
PERFORMANCE SUMMARY

Dataset	No. of images	No. of prod-ucts	Avg. No. of descriptors / frame	Maximum Recall with Precision $\geq 90\%$	
				Method I	Method II
D1	861	22	1021	74.95	72.47
D2	315	23	891	70.17	56.00
D3	370	24	1268	98.86	97.36
D4	1135	27	669	84.79	85.59

The experimental results show that the proposed methods are reliable and efficient for observing the stock levels on the shelf and thus, takes us closer towards the realization of a complete robot-based solution for retail stock monitoring. The work presented in this paper is among very few which provide detail implementation of the algorithm with performance evaluation results on experimental datasets.

REFERENCES

[1] K. Mankodiya, R. Gandhi, and P. Narasimhan, "Challenges and opportunities for embedded computing in retail environments," in *Sensor Systems and Software*. Springer, 2012, pp. 121–136.

[2] S. Kumar, G. Sharma, K. Nishant, *et al.*, "Retail monitoring and stock assessment using mobile robots," in *Proc. of IEEE Int. Conf. on Technologies for Practical Robot Applications (TePRA)*, Woburn, MA, USA, 2014.

[3] T. Zimmerman, "System and method for performing inventory using a mobile inventory robot," Mar. 27 2008, uS Patent App. 11/534,162. [Online]. Available: https://www.google.com/patents/US20080077511

[4] S. B. Gokturk and A. Rafii, "Occupancy detection and measurement system and method," Oct. 2 2003, uS Patent App. 10/678,998.

[5] R. Groenevelt, A. Opalach, A. Fano, and F. Linaker, "Detection of stock out conditions based on image processing," Jan. 14 2014, uS Patent 8,630,924. [Online]. Available: https://www.google.com/patents/US8630924

[6] Y. Hofman and M. Rotenberg, "System and method for identifying retail products and determining retail product arrangements," June 20 2012, uS Patent App. 13/528,189.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding, Elsevier*, vol. 110, pp. 346–359, December 2008.

[8] L. Juan and O. Gwun, "A comparison of sift, pca-sift and surf," *International Journal of Image Processing (IJIP)*, vol. 3, no. 4, pp. 143–152, 2009.

[9] T. W. Gruen and D. S. Corsten, *A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry*. Grocery Manufacturers of America, 2007.

[10] H. Che, J. Chen, and Y. Chen, "Investigating effects of out-of-stock on consumer sku choice," 2011.

[11] D. Corsten and T. Gruen, "Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks," *International Journal of Retail & Distribution Management*, vol. 31, no. 12, pp. 605–617, 2003.

[12] NeWave, "RFID-based smart shelf system." [Online]. Available: http://newavesensors.com/

[13] M. Kärkkäinen, "Increasing efficiency in the supply chain for short shelf life goods using RFID tagging," *International Journal of Retail & Distribution Management*, vol. 31, no. 10, pp. 529–536, 2003.

[14] T. BIRCH and M. Kasper, "Method and apparatus for managing product placement on store shelf," Oct. 10 2013, wO Patent App. PCT/US2012/032,420. [Online]. Available: https://www.google.com/patents/WO2013151553A1?cl=en

[15] B. Goldberg and G. Messinger, "Inventory or asset management system," June 12 2008, uS Patent App. 11/846,764. [Online]. Available: https://www.google.com/patents/US20080140478

[16] A. Opalach, A. Fano, F. Linaker, and R. Groenevelt, "Planogram extraction based on image processing," Mar. 5 2009, uS Patent App. 11/849,165. [Online]. Available: https://www.google.com/patents/US20090059270

[17] R. Cato and T. Zimmerman, "Using cameras to monitor actual inventory," May 14 2009, uS Patent App. 11/937,095. [Online]. Available: https://www.google.com/patents/US20090121017

[18] D. Crasto, A. Kale, and C. Jaynes, "The smart bookshelf: A study of camera projector scene augmentation of an everyday environment," in *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, vol. 1. IEEE, 2005, pp. 218–225.

[19] T. Grigsby and S. Mishra, "Product identification using image analysis and user interaction," June 5 2012, uS Patent 8,194,985. [Online]. Available: https://www.google.com/patents/US8194985

[20] D. Limer, D. Lang, C. Burt, P. Gouin, N. Tarr, and R. Dischinger, "Machine vision counting system apparatus and method," Oct. 6 2009, uS Patent 7,599,516. [Online]. Available: https://www.google.com/patents/US7599516

[21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 149–167, 2004.

[22] ROS, "Robot operating system," 2014. [Online]. Available: http://www.ros.org

[23] OpenCV, "Open source computer vision," 2014. [Online]. Available: http://www.opencv.org