sensors

MDPI

Article

Robust Shelf Monitoring Using Supervised Learning for Improving On-Shelf Availability in Retail Stores [†]

Kyota Higa * and Kota Iwamoto

Data Science Research Laboratories, NEC Corporation, Kawasaki, Kanagawa 211-8666, Japan; k-iwamoto@ay.jp.nec.com

* Correspondence: k-higa@ah.jp.nec.com

[†] This paper is an extended version of our paper published in "Higa, K.; Iwamoto, K. Robust Estimation of Product Amount on Store Shelves from a Surveillance Camera for Improving On-Shelf Availability. In Proceedings of the 2018 IEEE International Conference on Imaging Systems and Techniques (IST 2018), Kraków, Poland, 16–18 October 2018".

Received: 28 April 2019; Accepted: 14 June 2019; Published: 17 June 2019

Abstract: This paper proposes a method to robustly monitor shelves in retail stores using supervised learning for improving on-shelf availability. To ensure high on-shelf availability, which is a key factor for improving profits in retail stores, we focus on understanding changes in products regarding increases/decreases in product amounts on the shelves. Our method first detects changed regions of products in an image by using background subtraction followed by moving object removal. It then classifies the detected change regions into several classes representing the actual changes on the shelves, such as "product taken (decrease)" and "product replenished/returned (increase)", by supervised learning using convolutional neural networks. It finally updates the shelf condition representing the presence/absence of products using classification results and computes the product amount visible in the image as on-shelf availability using the updated shelf condition. Three experiments were conducted using two videos captured from a surveillance camera on the ceiling in a real store. Results of the first and second experiments show the effectiveness of the product change classification in our method. Results of the third experiment show that our method achieves a success rate of 89.6% for on-shelf availability when an error margin is within one product. With high accuracy, store clerks can maintain high on-shelf availability, enabling retail stores to increase profits.

Keywords: image processing; change detection; change classification; background subtraction; convolutional neural network; on-shelf availability; product amount; surveillance camera; retail

1. Introduction

To improve profits, retail stores, such as supermarkets and convenience stores, should aim to reduce lost sales opportunities. One of the criteria for measuring loss of sales opportunities is on-shelf availability, which is generally defined as the availability of products for sale to shoppers, in the place they expect them and at the time they want to buy them [1,2]. When shoppers see a desired product to be out of stock (i.e., it has no on-shelf availability), the shoppers purchase the same or similar product at a competing retailer in the worst case. Therefore, on-shelf availability is a key factor for improving profits in retail stores. For further improving profits, products need to be stocked in places visible to the shoppers, that is, not at the back but on the front of shelves. In this paper, we define high on-shelf availability as a condition where a sufficient amount of products is stocked on the front of the shelves visible to the shoppers.

For maintaining high on-shelf availability, store clerks must often look around a store and replenish products or rearrange the shelves. However, this is labor-intensive. For effectively maintaining high on-shelf availability, a method is needed to automatically detect the amount of products on the front of the shelves, which frequently changes and to alert the store clerks to replenish products or rearrange the shelves when the product amount on the shelves is low.

We need to know the number of products on the shelves for accurately understanding the product amount. A check-out system shows the number of remaining products in the inventory. However, even if products are in the inventory, they might not reach the shelves due to inefficient shelf management. In other words, they are not on the shelves but in the store backroom. Hence, we cannot accurately understand the product amount on the shelves by using current check-out systems.

On the other hand, radio frequency identification (RFID) tags help us to understand the product amount on the shelves since they accurately detect the products in real-time. We can maintain high on-shelf availability since we can replenish products before they go out of stock. However, the RFID-based approach faces the problems of the exponential cost of sensor installation and time-consuming work of attaching tags to each product and removing them again at a billing counter (i.e., it is a high-cost solution).

For a low-cost solution, we apply an image processing technology to an image of the shelves. Figure 1 shows an overview of our target application to monitor the shelves in retail stores using a video captured from a surveillance camera on the ceiling for maintaining high on-shelf availability. It estimates the product amount on the shelves visible in the image and alerts the store clerks to replenish products or rearrange the shelves when the product amount on the shelves is low.

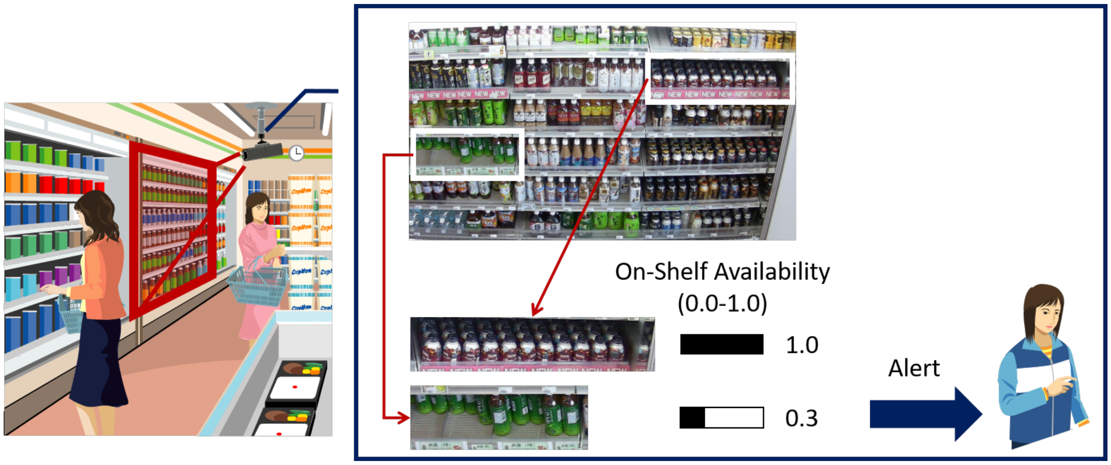


Figure 1. Overview of our target application to monitor the shelves in retail stores using a surveillance camera for maintaining high on-shelf availability.

Methods [3,4] have recently been proposed to monitor the shelves by using the images for retail stores. Kejriwal et al. [3] proposed a method to directly obtain product count from images using mobile robots for stock monitoring and assessment in retail stores. This method identifies products on the shelves by feature matching with SURF (Speeded-Up Robust Features) [5] using product templates. Hence, it requires a large database of high-quality product templates, which is labor-intensive to maintain since products frequently change due to new and seasonal products arriving. Rosado et al. [4] proposed a method to detect out-of-stock regions in grocery retail shelves without using a large database. This method detects out-of-stock regions using points of interest (keypoints) detected by FAST (Features from Accelerated Segment Test) [6]. However, it cannot detect keypoints on products located on the front of the shelves when shoppers stand in front of the shelves, which leads to inaccurate estimation of the product amount on the shelves. Therefore, it cannot robustly monitor the shelves for maintaining high on-shelf availability.

This paper proposes a method to monitor the shelves robust to shopper occlusion and without a database of product templates for high on-shelf availability (this paper is an extended version of a previous paper [7]).

2. Related Work

Several methods have been proposed to identify products on shelves from an image or to enhance the accuracy of the reconstruction of large-scale objects from images. Liu et al. [8,9] proposed a method to check planogram (product layout) compliance by using recurring patterns. Product layout information is extracted from one single input image by unsupervised recurring pattern detection and matched via graph matching with the expected product layout. Saran et al. [10] proposed a visual analysis based framework using SURF for automated planogram compliance check in retail stores. These methods [8–10] focus on recognizing a planogram from an image. Cali and Ambu [11] proposed a method to improve the quality of the 3D digital surface reconstruction of large-scale objects from images. It focuses on reconstructing the surface of objects.

Kejriwal et al. [3] proposed a method to directly obtain product count from images captured using a monocular camera mounted on a mobile robot for stock monitoring and assessment in retail stores. Moorthy et al. [12] proposed a method to detect and count the front-facing products on the shelves as well as identify out-of-stock regions. These methods [3,12] identify products on the shelves by extracting SURF features around keypoints and by comparing them with those of product templates. Satapathy et al. [13] proposed a method to detect products on the shelves by using exhaustive template matching for raising alerts to replenish products. Bao et al. [14] proposed a method to accurately identify and locate a large number of objects in an image by using a scalable and compact binary local descriptor, named the BRIGHT (Binary ResIzable Gradient HisTogram) descriptor [15]. Zhang et al. [16] proposed a dual-layer density estimation-based architecture to detect multiple object instances from an image for robot inventory management. It identifies multiple object instances by feature matching with SIFT (Scale-Invariant Feature Transform) [17]. Scovanner et al. [18] proposed a 3-dimensional SIFT descriptor for video or 3D imagery. Scale and rotation invariant local descriptors, such as SIFT, SURF and BRIGHT, have enabled identification of real-world objects in images with robustness against differing camera viewpoints, occlusions and lighting conditions. To achieve high performance, these methods [3,12–14,16,18] require a large database of high-quality product templates that must be updated every time products on the shelves change. However, the large database is time-consuming to maintain since the products on the shelves frequently change due to new and seasonal products arriving.

On the other hand, a method has been proposed to detect out-of-stock regions without a large database of high-quality product templates. Rosado et al. [4] proposed a supervised learning approach for out-of-stock detection from high-resolution panoramic images of grocery retail shelves. This method detects out-of-stock regions on the basis of the density of keypoints detected by FAST since those regions are more homogeneous in terms of color and texture than product regions, that is, a sufficient number of keypoints cannot be detected from out-of-stock regions. Detecting keypoints from the images needs no database of product templates. However, we cannot accurately understand the product amount on the front of the shelves by using this method since it detects keypoints on products located in the second row of the shelves even when shoppers take products from the front of the shelves. In addition, when shoppers stand in front of the shelves, it cannot detect keypoints on products located on the front of the shelves, which also leads to inaccurate estimation of the product amount on the shelves. Therefore, it cannot robustly monitor the shelves for maintaining high on-shelf availability.

Thus, there is still no method to easily and robustly monitor the shelves from the image for high on-shelf availability.

3. Proposed Method

3.1. Overview

The proposed method robustly monitors the shelves by detecting and classifying changes in products regarding increases/decreases in product amount, such as "product taken (decrease)" and "product replenished/returned (increase)", using images from a surveillance camera and by updating the shelf condition representing presence/absence of products on the front of the shelves from the initial shelf condition given in advance. Background subtraction (or foreground detection) [19–21], which is a widely used approach for detecting changes in images as foregrounds, cannot distinguish between "product taken" and "product replenished/returned" since it detects regions that do not match a reference image, often called "background image" or "background model". Hence, background subtraction alone is not sufficient for robustly monitoring the shelves for maintaining high on-shelf availability. To solve this problem, we need to classify the changes in the image into the actual changes in products on the shelves, such as "product taken" and "product replenished/returned". The process flow of our method consists of three blocks as shown in Figure 2. It first detects change regions of the shelves in the image by background subtraction on the basis of statistical information of pixels. To reduce false positives of change regions such as shoppers, it tracks detected foregrounds between consecutive images and removes them on the basis of their moving distance. It then classifies the detected change regions into four classes representing the actual changes on the shelves by supervised learning using convolutional neural networks (CNNs) [22]. It finally updates the shelf condition representing presence/absence of products on the front of the shelves using classification results and computes the product amount visible in the image as on-shelf availability.

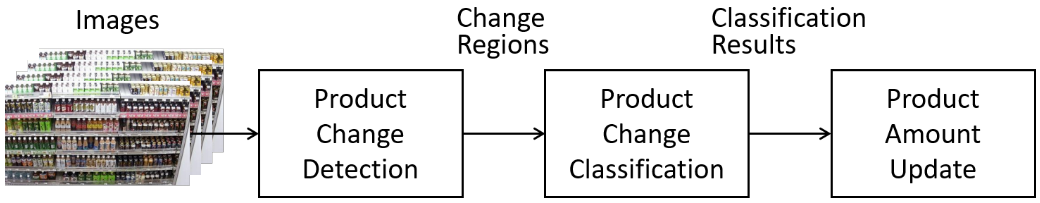


Figure 2. Process flow of proposed method.

3.2. Detection of Product Changes

We accurately detect only the change regions of products on the shelves by background subtraction followed by moving object removal. Our method detects the change regions in images as foregrounds by using the background subtraction [19]. However, background subtraction detects not only the change regions of products but also false positives such as moving objects (i.e., shoppers and store clerks). Therefore, we track the foregrounds between consecutive images and remove them on the basis of their moving distance. The foregrounds with a moving distance larger than T_d pixels are removed as moving objects, whereas the static foregrounds detected for more than T_s seconds are selected as the change regions of products.

The Hungarian method [23], which is a combinatorial optimization algorithm that solves the assignment problem, is used for tracking the foregrounds in the consecutive images. In particular, the foregrounds are tracked by minimizing the summation of assignment cost among N foregrounds detected in a current image and M foregrounds detected in a previous image. The cost matrix of the Hungarian method used in our method is as follows:

$$C_{i,j} = \begin{bmatrix} c_{1,1} & \cdots & c_{1,N} \\ \vdots & \ddots & \vdots \\ c_{M,1} & \cdots & c_{M,N} \\ c_{def} & & c_{max} \\ & \ddots & \\ c_{max} & & c_{def} \end{bmatrix}. \quad (1)$$

Here, the parameters c_{def} and c_{max} are a default value and a value sufficiently larger than the default value, respectively.

The assignment cost $c_{i,j}$ consists of the similarity of the color histograms in the foregrounds, area ratio of the foregrounds and aspect ratio of bounding rectangles of the foregrounds.

The similarity cost $c^{color}_{i,j}$ is computed as,

$$c^{color}_{i,j} = 1.0 - \frac{\Sigma_k (H^j_t(k) - \bar{H}^j_t)(H^i_{t-1}(k) - \bar{H}^i_{t-1})}{\sqrt{\Sigma_k (H^j_t(k) - \bar{H}^j_t)^2 \Sigma_k (H^i_{t-1}(k) - \bar{H}^i_{t-1})^2}}, \quad (2)$$

$$\bar{H}^n_m = \frac{1}{K} \Sigma_k H^n_m(k), \quad (3)$$

where H^j_t and H^i_{t-1} are color histograms of the jth and ith foregrounds in the current and previous images, respectively. K is the number of bins. Note that the total area of a histogram has been normalized to 1.0.

The area ratio cost $c^{area}_{i,j}$ is computed as,

$$c^{area}_{i,j} = 1.0 - \begin{cases} S^j_t / S^i_{t-1} & \text{if } S^j_t < S^i_{t-1} \\ S^i_{t-1} / S^j_t & \text{else} \end{cases}, \quad (4)$$

where S^j_t and S^i_{t-1} are areas of the jth and ith foregrounds in the current and previous images, respectively.

The aspect ratio cost $c^{aspect}_{i,j}$ is computed as,

$$c^{aspect}_{i,j} = 1.0 - \begin{cases} A^j_t / A^i_{t-1} & \text{if } A^j_t < A^i_{t-1} \\ A^i_{t-1} / A^j_t & \text{else} \end{cases}, \quad (5)$$

where A^j_t and A^i_{t-1} are aspect ratio of the bounding rectangle for the jth and ith foregrounds in the current and previous images, respectively.

Finally, the assignment cost $c_{i,j}$ is computed as,

$$c_{i,j} = w_1 \times c^{color}_{i,j} + w_2 \times c^{area}_{i,j} + w_3 \times c^{aspect}_{i,j}, \quad (6)$$

where w means weight.

Figure 3 shows an example of detecting and tracking the foregrounds. Left and right images show the previous and current images, respectively. Top and bottom images show an input image and the foregrounds, respectively. Note that the person in the top-left image is retouched for privacy protection.

Here, three small foregrounds and one large foreground are detected in the previous image and three small foregrounds are detected in the current image. Our method accurately detects only the changes in products even when the shoppers stand in front of the shelves by tracking the foregrounds between the consecutive images on the basis of three similarities by using the Hungarian method.

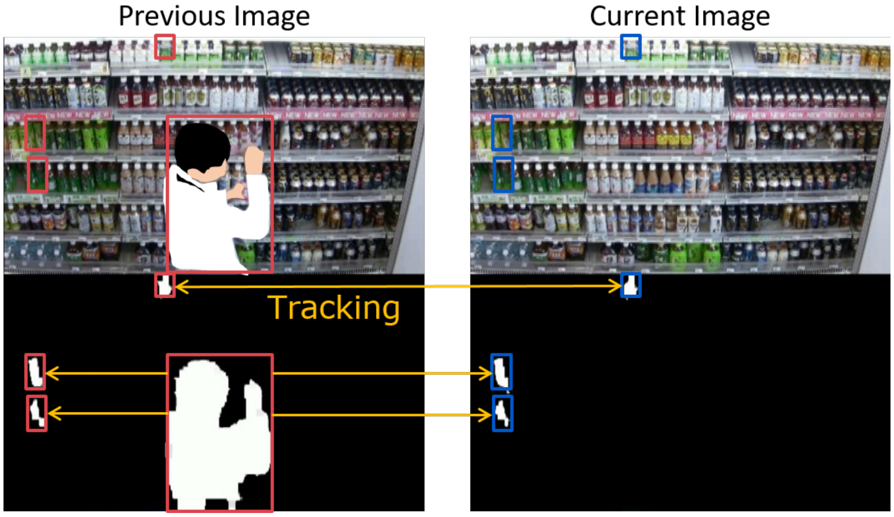


Figure 3. Example of detecting and tracking foregrounds.

3.3. Classification of Product Changes

The detected change regions are classified into four classes representing the actual changes on the shelves by supervised learning using CNNs for accurately understanding increases/decreases in the product amount.

In this paper, we try two networks for classifying the detected change regions: a CIFAR-10-based network shown in Figure 4 and a CaffeNet-based network (http://caffe.berkeleyvision.org/) shown in Figure 5. The input and final fully-connected layers in both networks are modified compared with those of the original networks. These networks are learned using training samples collected from videos taken in real stores. Input data of our network are six-channel images constructed from images before and after the change as shown in Figures 4 and 5. Zhao et al. [24] proposed a method to classify difference images by deep neural networks. However, it cannot distinguish the actual changes on the shelves such as "product taken" and "product replenished/returned" since they are essentially symmetrical operations (i.e., the difference images of those change regions are almost identical) as shown in Figure 6. Therefore, we use the six-channel images to which no extra operations are applied as input data of our network. Note that the images before and after the change are resized to 32 × 32 pixels for the CIFAR-10-based network or 227 × 227 pixels for the CaffeNet-based network before constructing the six-channel image.

Our network classifies the detected change regions into four classes shown in Table 1. Examples of images for each class are shown in Figure 7. Classes 1 and 2 show "product taken" and "product replenished/returned", respectively. Class 3 shows that the position or direction of a product slightly changed since a shopper touched it. Class 4 shows false positives unrelated to human movement, such as illumination changes. Classes 1 and 2 show that the product amount on the shelves decreases or increases, respectively, whereas classes 3 and 4 show that the product amount on the shelves does not change.

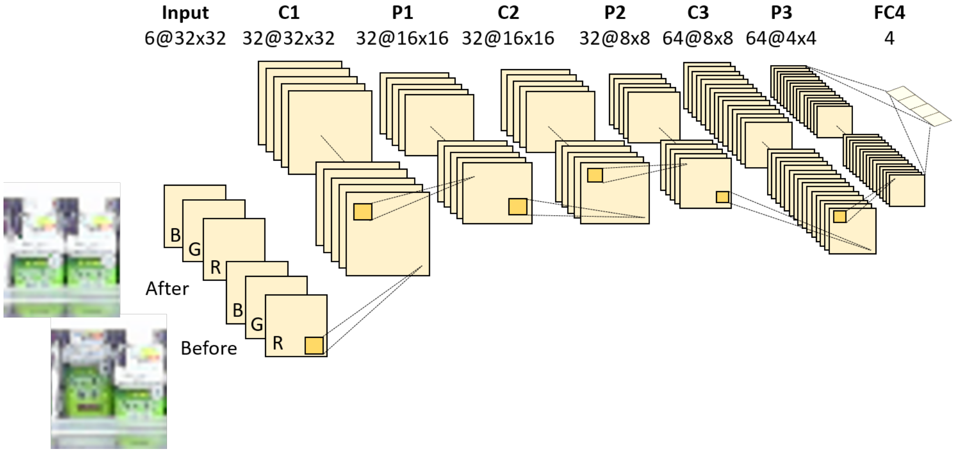


Figure 4. Network architecture based on CIFAR-10 (Canadian Institute For Advanced Research). It consists of three convolutional layers, three pooling layers and one fully-connected layer.

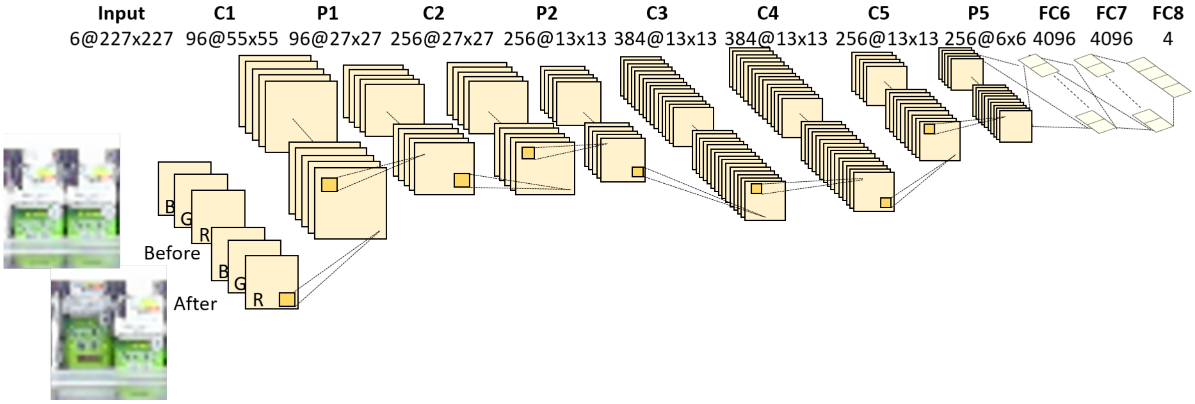


Figure 5. Network architecture based on CaffeNet. It consists of five convolutional layers, three pooling layers and three fully-connected layers.

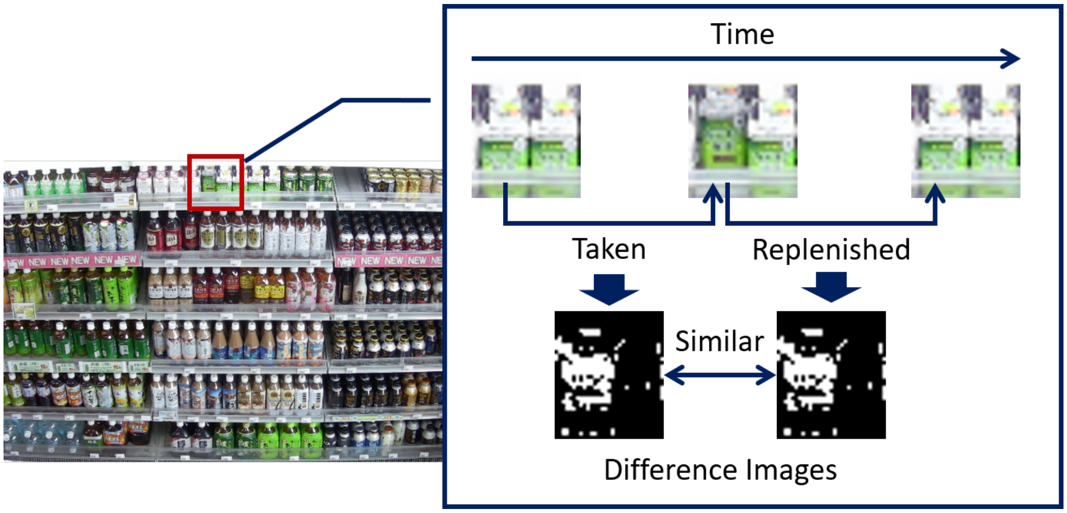


Figure 6. Examples of difference images for change regions of "product taken" and "product replenished".

Table 1. Output of our network in product change classification.

Class	Description	Product Amount
1	A product was taken.	decrease
2	A product was replenished by store clerks or returned by shoppers.	increase
3	A position or direction of a product slightly changed since a shopper touched it.	no change
4	It was a false positive unrelated to human movement, such as an illumination change.	no change

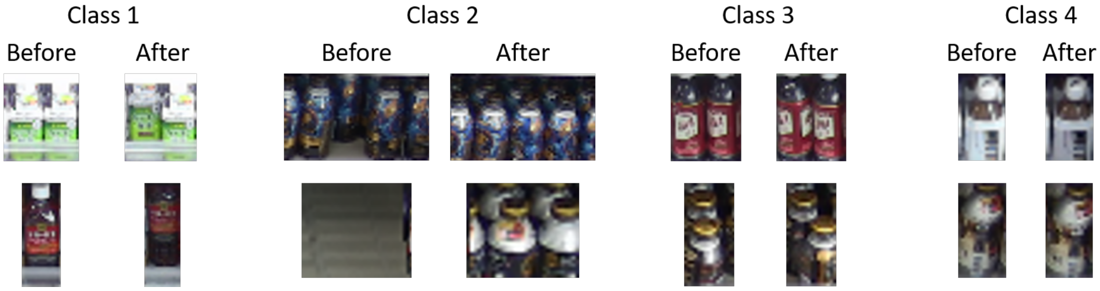


Figure 7. Examples of images for each class in product change classification.

3.4. Update of Product Amount

Our method updates the shelf condition representing presence/absence of products on the front of the shelves using the classification results and computes the product amount visible in the image as on-shelf availability using the updated shelf condition and predefined monitoring areas.

Figure 8 shows examples of the shelf condition represented as a binary image. Left and right images are the shelf condition before and after an update, respectively. The white and black regions of the shelf condition show “presence” and “absence” of products, respectively. When the classification results are “product taken (class 1)”, the shelf condition corresponding to the change regions is set to “absence (black)” as shown in Figure 8a. When the classification results are “product replenished/returned (class 2)”, the shelf condition corresponding to the change regions is set to “presence (white)” as shown in Figure 8b. The shelf condition is not changed in the case of other classification results (classes 3 and 4).

Figure 9 shows examples of the shelf condition updated by our method. The shelf condition at 0 min shows the initial shelf condition given in advance. As shown in Figure 9, our method updates the shelf condition, which frequently changes on the basis of “product taken” and “product replenished/returned” detected from the image.

After updating the shelf condition, our method computes on-shelf availability (product amount) visible in the image between the range of 0.0 and 1.0 using the predefined monitoring areas shown in Figure 10 as follows,

$$OSA(n) = S_p(n)/S_m(n), \tag{7}$$

where $OSA(n)$ is on-shelf availability for the nth shelf. $S_p(n)$ and $S_m(n)$ are the product areas of the nth shelf (i.e., the number of white pixels in the shelf condition corresponding to the nth shelf) and nth monitoring area (i.e., the number of pixels in the nth monitoring area), respectively. Figure 11 shows an overview of computing on-shelf availability for the nth shelf. When some products are taken/replenished simultaneously as shown in the lower example of class 2 in Figure 7, on-shelf availability increases/decreases on the basis of the summation of changed areas.

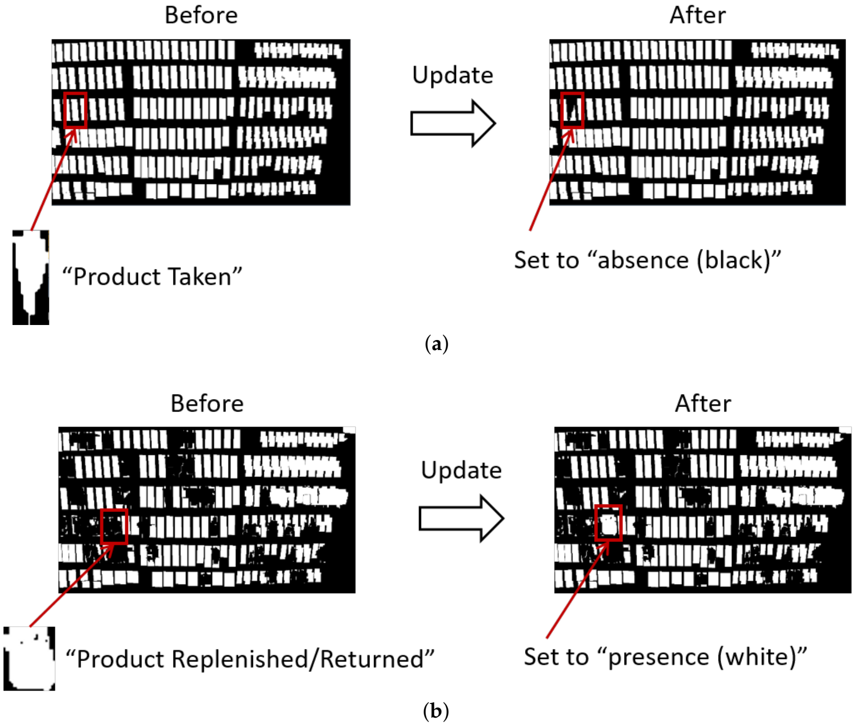


Figure 8. Examples of shelf condition represented as binary image. (**a**) "Product taken". (**b**) "Product replenished/returned".

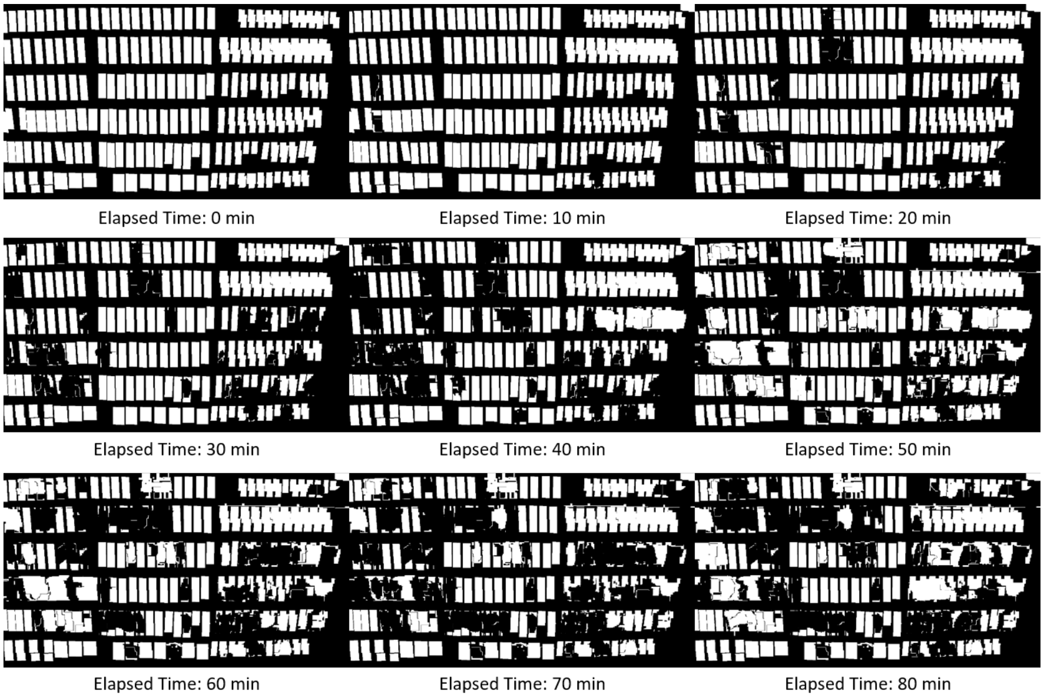


Figure 9. Examples of shelf condition updated by our method.

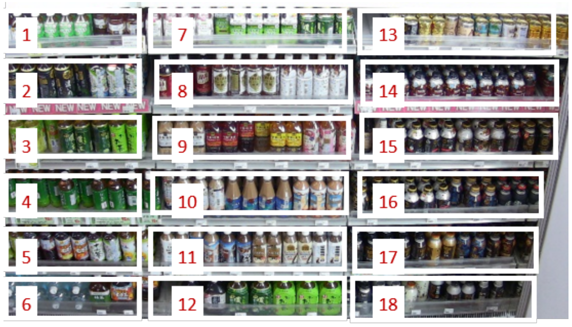


Figure 10. Predefined monitoring areas with shelf number.

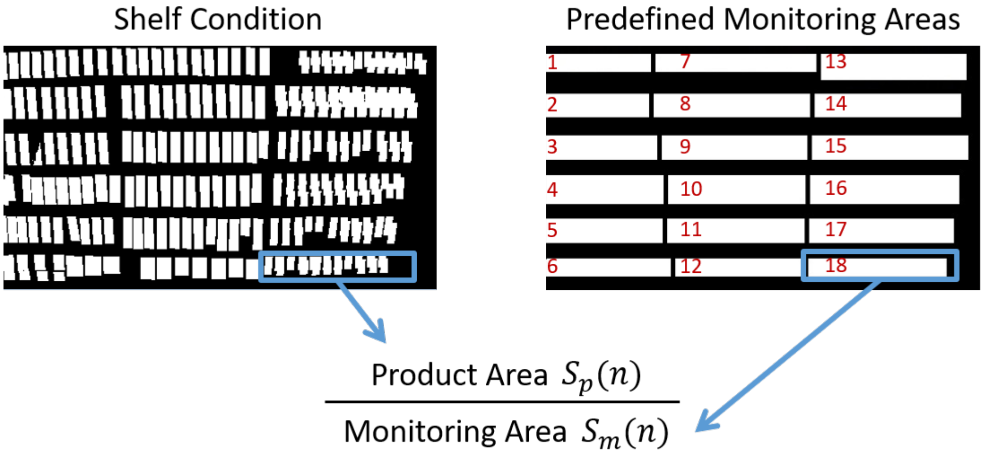


Figure 11. Overview of computing on-shelf availability for *n*th shelf.

4. Experiments

4.1. Experimental Conditions

We conducted three experiments on two videos captured from a fixed camera on the ceiling in a real store for evaluating the accuracy of our method. Tables 2 and 3 show details of the videos used in the experiments and parameters in the product change detection, respectively. The videos of shelves with bottled beverages shown in Figure 10 were used in the experiments and showed more than 200 shoppers in total. As shown in Table 2, the resolution and frame rate of the videos are low, which makes it difficult to track moving objects between consecutive images. However, such conditions are suitable for retailers since they decrease costs of data storage to save videos. The networks for the product change classification shown in Figures 4 and 5 were learned using training samples extracted from other videos taken in the real store with the same camera conditions. The number of training samples for each class is shown in Table 4. The initial shelf condition shown in Figure 8 and the monitoring areas shown in Figure 10 were set manually. For every minute, we computed on-shelf availability (i.e., product amount on shelves) using Equation (7) and then counted the number of successful frames whose difference between the estimated on-shelf availability $OSA(n)$ and ground truth is less than an error margin. The success rate for each shelf is computed as,

$$SuccessRate = S_n/(S_n + U_n), \tag{8}$$

where S_n and U_n are the numbers of successful and unsuccessful frames for the *n*th shelf, respectively.

Table 2. Details of videos used in experiments.

Video Length	**Video 1**	185 min
	Video 2	95 min
Resolution		480 × 270 pixels (1/16 of Full HD)
Frame Rate		1 fps

Table 3. Details of parameters in product change detection.

T_d ***in Section 3.2***	10 pixels
T_s ***in Section 3.2***	30 s
c_{def} ***in Equation (1)***	1.0
K ***in Equation (3)***	32
w ***in Equation (6)***	1.0

Table 4. Number of training samples in product change classification.

Class	Training Samples	Class	Training Samples
1	16,038	***3***	1448
2	16,038	***4***	3486

4.2. Experiment 1: Comparison of Networks for Product Change Classification

We evaluated the estimation accuracy of the CIFAR-10-based network and the CaffeNet-based network for selecting the network for the product change classification. Figure 12 shows the average success rate for on-shelf availability for all the shelves at various error margins. The vertical and horizontal axes are the success rate and the error margin, respectively. The blue and red lines show the success rates of the CIFAR-10-based and CaffeNet-based networks, respectively. As shown in Figure 12, the two networks barely differ in terms of the success rates for on-shelf availability, which means that the network for the product change classification can be selected on the basis of hardware resources such as memory size or system requirements such as processing time. We select the CaffeNet-based network for the product change classification since its success rate is slightly higher than that of the CIFAR-10-based network although it needs much memory (i.e., the model size is large).

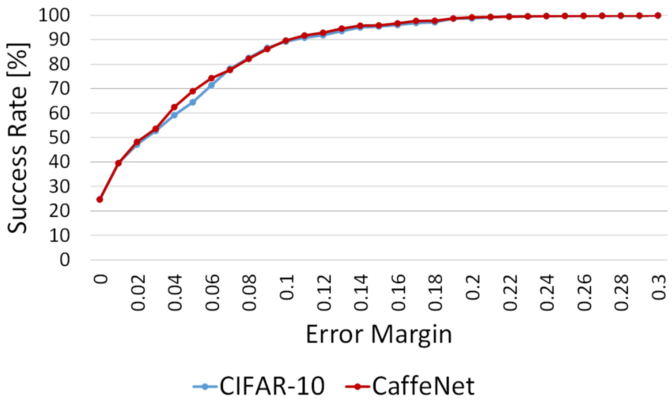


Figure 12. Average success rate for on-shelf availability for all the shelves at various error margins in experiment 1.

4.3. Experiment 2: Evaluation of Effectiveness for Product Change Classification

For evaluating the effectiveness of the product change classification in our method, we compared the estimation accuracy of our method with that of a method accumulating the product change regions detected by background subtraction [19] (i.e., the product change detection only). The comparison method accumulates all the detected change regions on the shelves as "product taken". Figure 13 shows the average success rate for on-shelf availability for all the shelves at various error margins in this experiment. The red and blue lines show the success rates of the proposed and comparison methods, respectively. Our method achieves the success rate of 89.6% for on-shelf availability at the error margin of 0.10, which is equivalent to an error of approximately ±1 product or less since approximately 10 products are on each shelf as shown in Figure 10. Our method improves the success rate by 46.7 points compared with that of the comparison method (42.9%) at the error margin of 0.10, which shows the effectiveness of the product change classification by supervised learning using CNNs in our method.

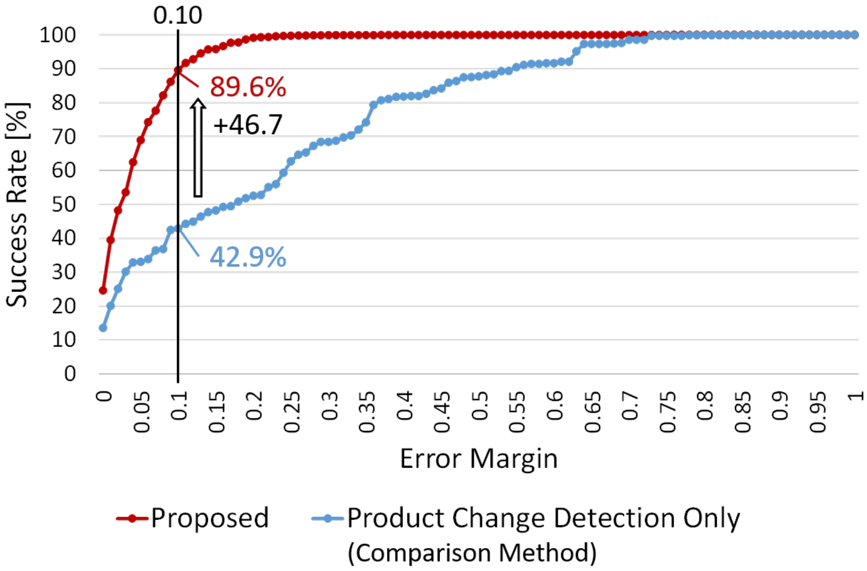


Figure 13. Average success rate for on-shelf availability for all the shelves at various error margins in experiment 2.

4.4. Experiment 3: Comparison of Proposed and Conventional Methods

We compared the estimation accuracy of the proposed method with that of the conventional method [4], which is the state-of-the-art technology of out-of-stock detection. Figure 14 shows the average success rate for on-shelf availability for all the shelves at various error margins in this experiment. The red and blue lines show the success rates of the proposed and conventional methods, respectively. Our method (89.6%) improves the success rate by 25.1 points compared with that of the conventional method (64.5%) at the error margin of 0.10.

Table 5 shows the success rate for each shelf in our method at the error margin of 0.10. High (over 90%) and low (under 80%) success rates are written in red in bold and blue with an underline, respectively. As shown in Table 5, our method achieves high success rates for 10 out of 18 shelves at the error margin of 0.10.

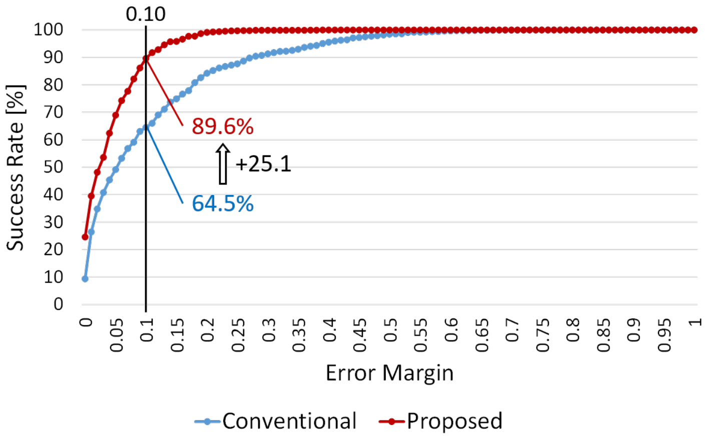
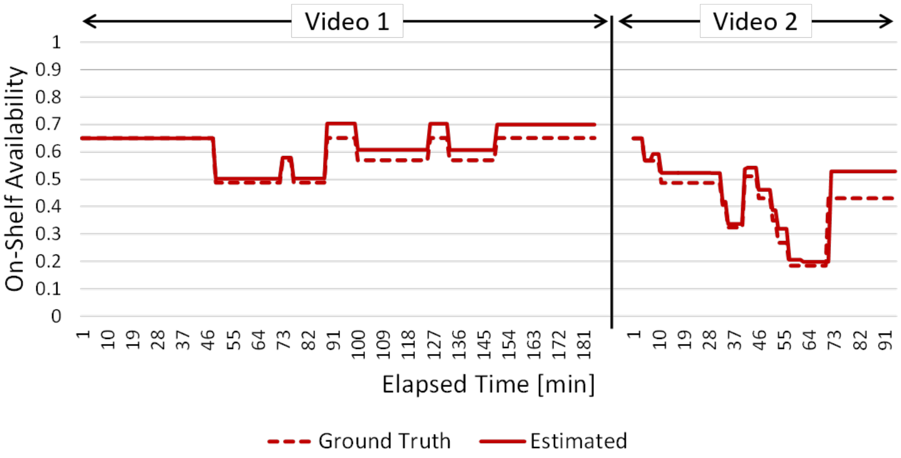


Figure 14. Average success rate for on-shelf availability for all the shelves at various error margins in experiment 3.

Table 5. Success rate for each shelf at error margin of 0.10.

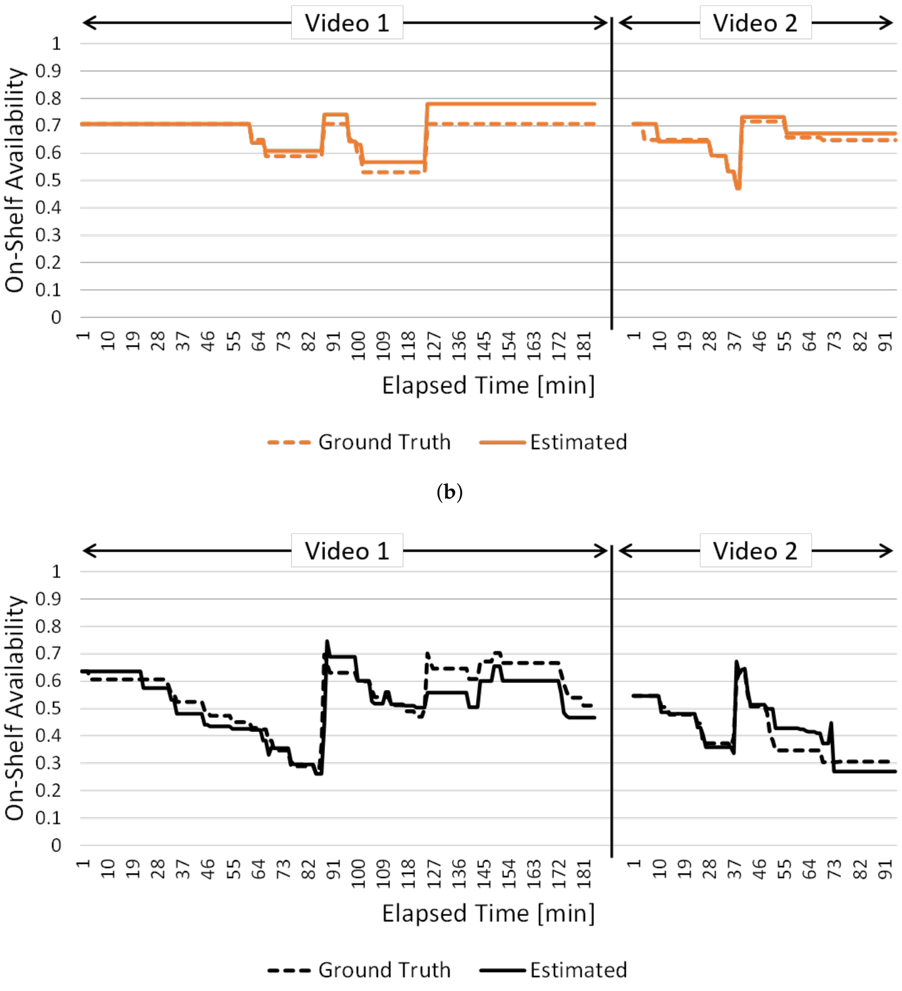
Shelf Number	Success Rate	Shelf Number	Success Rate	Shelf Number	Success Rate
1	77.7%	***7***	**100%**	***13***	**99.3%**
2	80.6%	***8***	83.3%	***14***	**100%**
3	**99.6%**	***9***	**97.1%**	***15***	51.4%
4	87.4%	***10***	**98.2%**	***16***	**90.3%**
5	80.6%	***11***	88.8%	***17***	**96.0%**
6	**99.6%**	***12***	**99.3%**	***18***	83.5%

Figures 15 and 16 show the changes in on-shelf availability every minute. The results for shelves #3, #7 and #17 are shown in Figure 15 as examples of high success rates. The results for shelves #1 and #15 are shown in Figure 16 as examples of low success rates. The vertical and horizontal axes are on-shelf availability and elapsed time of the videos, respectively. Dotted and solid lines show ground truth and the estimated on-shelf availability, respectively.



(**a**)

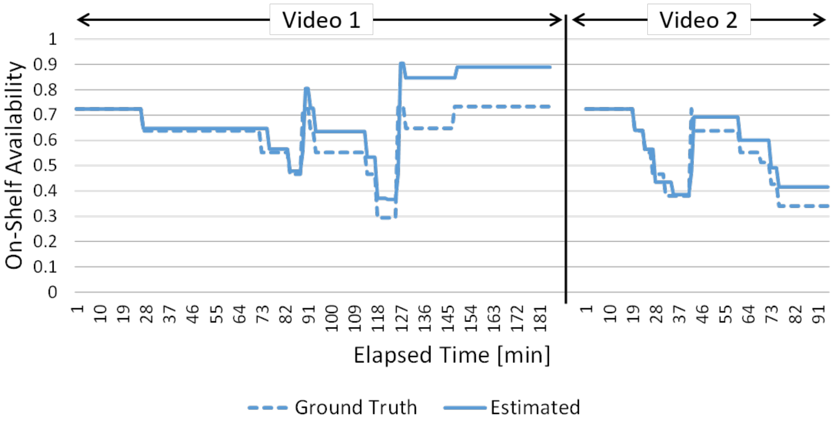
Figure 15. *Cont.*



(**b**)

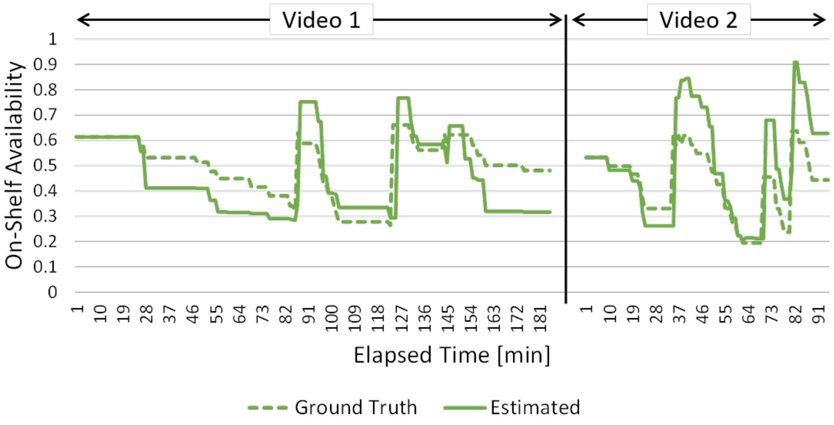
(**c**)

Figure 15. Changes in on-shelf availability every minute in cases of high success rates. (**a**) Results for shelf #3. (**b**) Results for shelf #7. (**c**) Results for shelf #17.



(**a**)

Figure 16. *Cont.*



(**b**)

Figure 16. Changes in on-shelf availability every minute in cases of low success rates. (**a**) Results for shelf #1. (**b**) Results for shelf #15.

4.5. Discussion

Figure 15 shows that our method accurately tracks the changes in on-shelf availability since the increases/decreases in on-shelf availability estimated by our method (solid lines) are very similar to those of the ground truth (dotted lines).

Figure 16 shows some points to be improved. In the results for shelf #1 shown in Figure 16a, our method fails to estimate on-shelf availability at 127 min in video 1 due to a classification error; specifically, on-shelf availability estimated by our method increases more than the ground truth. Since our method tracks relative changes in on-shelf availability, the classification error affects estimation of on-shelf availability after 127 min. The estimation accuracy can be improved by introducing a scheme to regularly reset the accumulated error. In the results for shelf #15 shown in Figure 16b, our method fails to estimate on-shelf availability several times. For example, at 27 and 161 min in video 1, two kinds of change regions (“product taken” and “illumination change”) were detected as one change region as shown in Figure 17 since those regions adjoin and then the detected change region was classified into “product taken”. Hence, our method estimated that on-shelf availability had decreased significantly at 27 and 161 min in video 1. This problem can be expected to be solved by dividing the detected change region in detail by using an image segmentation method and by classifying the divided regions.

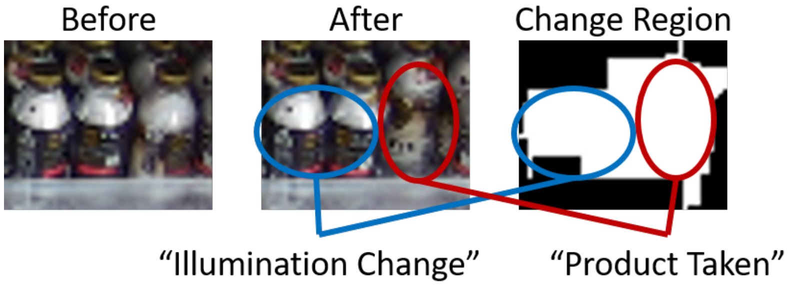


Figure 17. Example in which different kinds of change regions are detected as one change region since those regions adjoin.

Figures 18 and 19 show some results for the updated shelf condition in videos 1 and 2. The first and second rows show the captured images and the display condition (i.e., presence/absence of products), respectively. Some regions in the images are marked as examples of “product taken” and “product replenished”. This shows that our method can track the changes in products on the shelves regarding increases/decreases in the product amount. These results suggest that our method monitors the shelves from the images robust to shopper occlusion and without the database of product templates for maintaining high on-shelf availability.

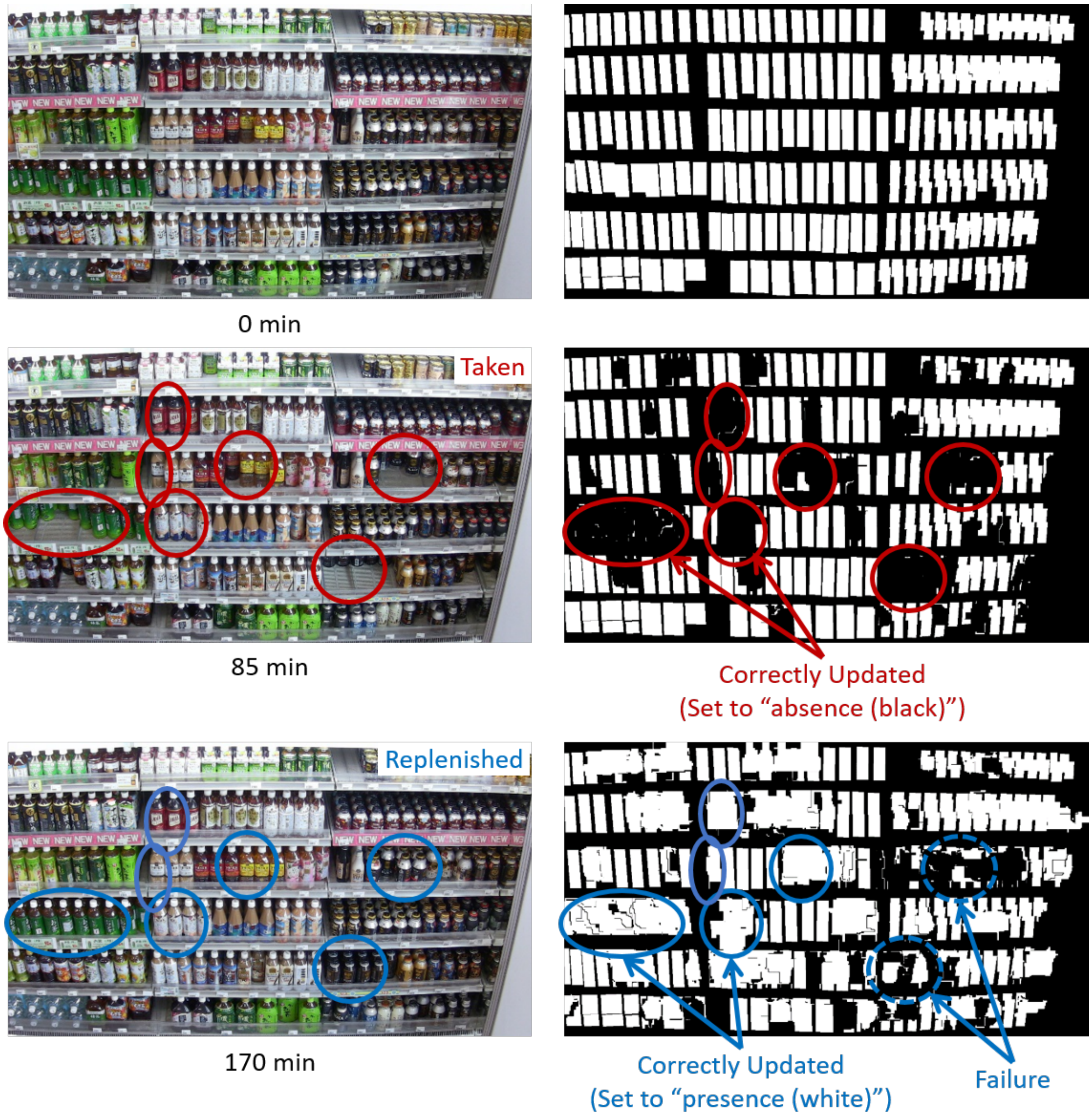


Figure 18. Some results for updated shelf condition in video 1.

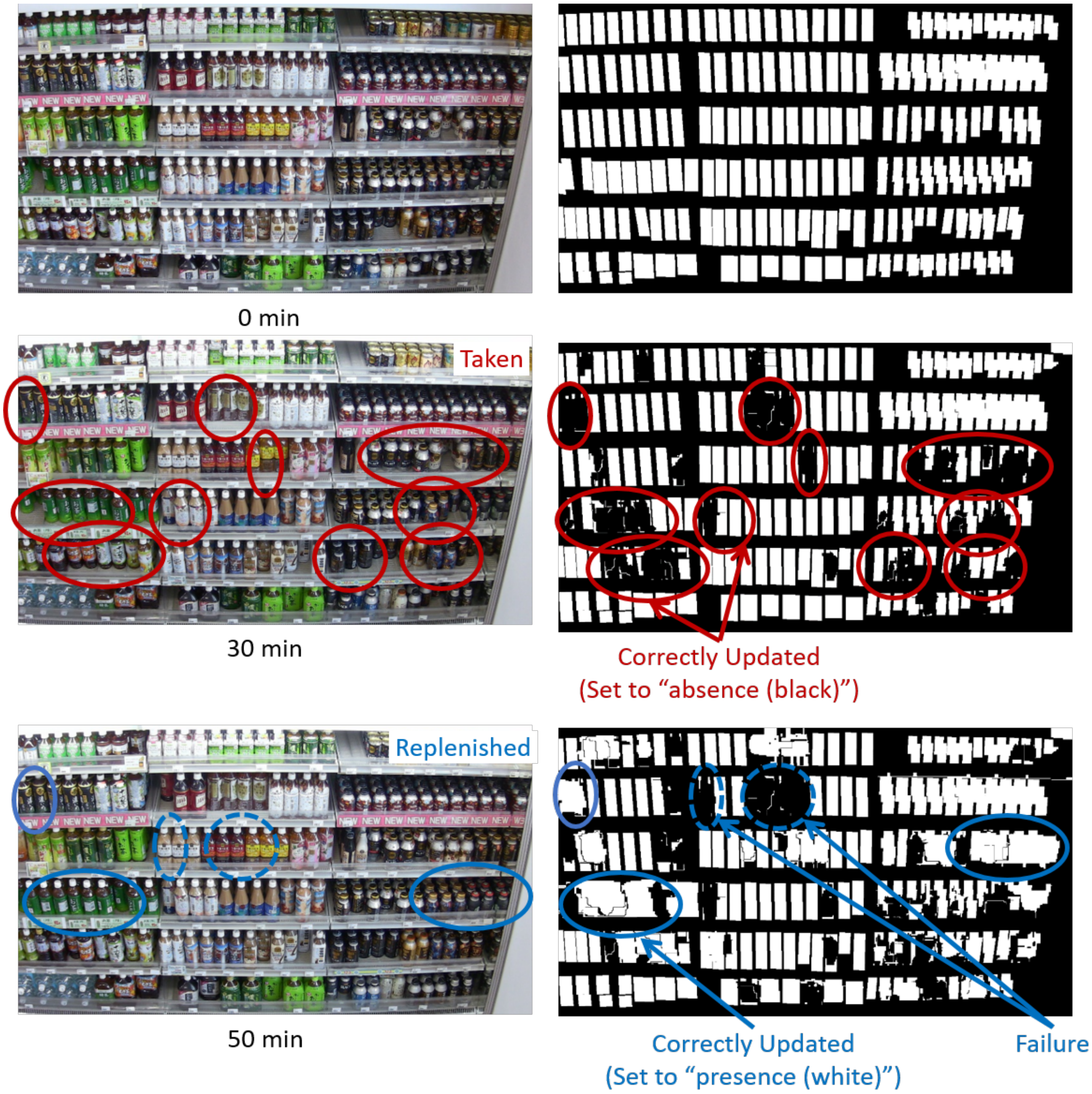
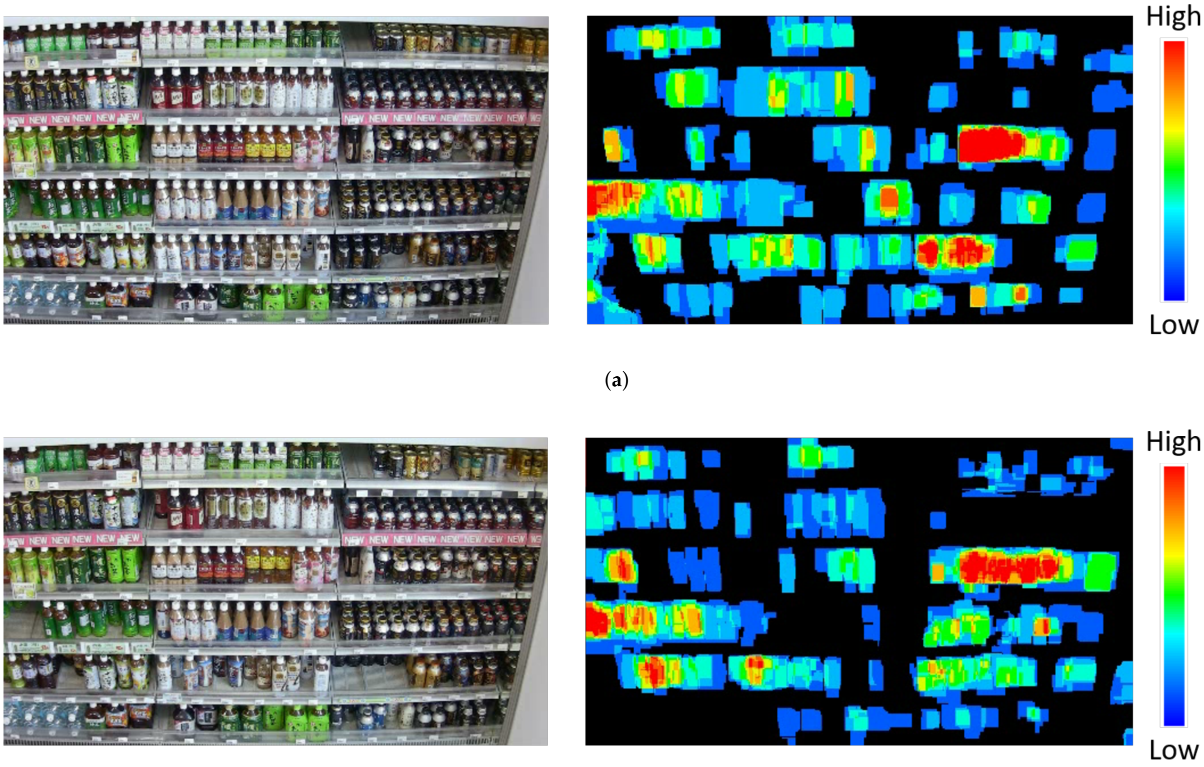


Figure 19. Some results for updated shelf condition in video 2.

Figure 20 shows examples of another application using our method. Left and right images are a captured image and a heatmap, respectively. Each heatmap shows the accumulation of the product changes detected from the images. In other words, it shows which shelves shoppers frequently accessed. By visualizing such information, retailers can plan a suitable planogram (product layout) for increasing their profits (e.g., a planogram to reduce low-accessed shelves and a planogram to increase the number of products displayed on high-accessed shelves).



(**a**)

(**b**)

Figure 20. Examples of another application using our method. (**a**) Result for video 1. (**b**) Result for video 2.

5. Conclusions

This paper proposed a method to robustly monitor shelves in retail stores using supervised learning for improving on-shelf availability. To ensure high on-shelf availability, which is a key factor for improving profits in retail stores, we focused on understanding changes in products regarding increases/decreases in product amount on the shelves. Our method first accurately detects only change regions of products in an image by using background subtraction followed by moving object removal for removing detected change regions unrelated to the changes of products (i.e., shoppers and store clerks). To reduce the moving objects, we track detected foregrounds between consecutive images and remove them on the basis of their moving distance. The detected change regions are then classified into four classes representing the actual changes on the shelves, such as "product taken (decrease)" and "product replenished/returned (increase)", by supervised learning using convolutional neural networks for accurately understanding increases/decreases in the product amount. Finally, the shelf condition representing presence/absence of products on the front of the shelves is accurately updated using classification results and the product amount visible in the image is computed as on-shelf availability using the updated shelf condition and predefined monitoring areas. Three experiments were conducted using two videos captured from a surveillance camera on the ceiling in a real store. The results of the first and second experiments show the effectiveness of the product change classification using a CaffeNet-based network in our method. The results of the third experiment show that the proposed method achieves a success rate of 89.6% for on-shelf availability when an error margin is within one product, which is 25.1 points higher than that of a conventional method (64.5%). With high accuracy, store clerks can maintain high on-shelf availability, enabling retail stores to increase their profits.

Author Contributions: Conceptualization, K.H.; Data curation, K.H.; Formal analysis, K.H.; Investigation, K.H.; Methodology, K.H.; Project administration, K.I.; Resources, K.H.; Software, K.H.; Supervision, K.H.; Validation, K.H.; Visualization, K.H.; Writing—original draft, K.H.; Writing—review & editing, K.I.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mitchell, A. *Improving on-Shelf Availability*; White Paper; SymphonyIRIGroup: Chicago, IL, USA, 2011.
2. Moorthy, R.; Behera, S.; Verma, S. On-shelf availability in retailing. *Int. J. Comput. Appl.* **2015**, *116*, 47–51. [CrossRef]
3. Kejriwal, N.; Garg, S.; Kumar, S. Product counting using images with application to robot-based retail stock assessment. In Proceedings of the 7th Annual IEEE International Conference on Technologies for Practical Robot Applications, Woburn, MA, USA, 11–12 May 2015; pp. 1–6.
4. Rosado, L.; Gonçalves, J.; Costa, J.; Ribeiro, D.; Soares, F. Supervised learning for out-of-stock detection in panoramas of retail shelves. In Proceedings of the IEEE International Conference on Imaging Systems and Techniques, Chania, Greece, 4–6 October 2016; pp. 406–411.
5. Bay, H.; Ess, A.; Tuytelaars. T.; Gool. L.V. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
6. Rosten, E.; Poter, R.; Drummond, T. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119. [CrossRef] [PubMed]
7. Higa, K.; Iwamoto, K. Robust estimation of product amount on store shelves from a surveillance camera for improving on-shelf availability. In Proceedings of the IEEE International Conference on Imaging Systems and Techniques, Kraków, Poland, 16–18 October 2018; pp. 164–169.
8. Liu, S.; Tian, H. Planogram compliance checking using recurring patterns. In Proceedings of the IEEE International Symposium on Multimedia, Miami, FL, USA, 14–16 December 2015; pp. 27–32.
9. Liu, S.; Li, W. ; Davis, S. ; Ritz, C.; Tian, H. Planogram compliance checking based on detection of recurring patterns. *IEEE Multimed.* **2016**, *23*, 54–63. [CrossRef]
10. Saran, A.; Hassa, E.; Maurya, A.K. Robust visual analysis for planogram compliance problem. In Proceedings of the 14th IAPR International Conference on Machine Vision Applications, Tokyo, Japan, 18–22 May 2015; pp. 576–579.
11. Calì, M.; Ambu, R. Advanced 3D photogrammetric surface reconstruction of extensive objects by UAV camera image acquisition. *Sensors* **2018**, *18*, 2815. [CrossRef] [PubMed]
12. Moorthy, R.; Behera, S.; Verma S.; Bhargave S.; Ramanathan P. Applying image processing for detecting on-shelf availability and product positioning in retail stores. In Proceedings of the 3rd International Symposium on Women in Computing and Informatics, Kochi, India, 10–13 August 2015; pp. 451–457.
13. Satapathy, R.; Prahlad, S.; Kaulgud, V. Smart shelfie—Internet of shelves: For higher on-shelf availability. In Proceedings of the IEEE Region 10 Symposium, Ahmedabad, India, 13–15 May 2015; pp. 70–73.
14. Bao, R.; Higa, K.; Iwamoto, K. Local descriptor based multiple object instance identification using scale and rotation invariant implicit shape model. In Proceedings of the International Workshop on Robust Local Descriptors for Computer Vision, Singapore, 1–5 November 2014; pp. w6–w7.
15. Iwamoto, K.; Mase, R.; Nomura, T. BRIGHT: A scalable and compact binary descriptor for low-latency and high accuracy object identification. In Proceedings of the 20th IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 2915–2919.
16. Zhang, Q.; Qu, D.; Xu, F.; Jia, K.; Sun, X. Dual-Layer Density Estimation for Multiple Object Instance Detection. *J. Sens.* **2016**, *2016*, 1–12. [CrossRef]
17. Lowe, D.G.; Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

18. Scovanner, P.; Ali, S.; Shah, M. A 3-dimensional sift descriptor and its application to action recognition. In Proceedings of the 15th ACM International Conference on Multimedia, Augsburg, Germany, 25–29 September 2007; pp. 357–360.
19. Zivkovic, Z. Improved adaptive Gaussian mixture model for background subtraction. In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 23–26 August 2004; Volume 2, pp. 28–31.
20. KaewTraKulPong, P.; Bowden, R. An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-Based Surveillance Systems*; Remagnino, P., Jones, G.A., Paragios, N., Regazzoni, C.S., Eds.; Springer: Boston, MA, USA, 2002; pp. 135–144.
21. Godbehere, A.B.; Matsukawa, A.; Goldberg, K. Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation. In Proceedings of the American Control Conference, Montréal, QC, Canada, 27–29 June 2012; pp. 4305–4312.
22. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
23. Munkres, J. Algorithms for the assignment and transportation problems. *J. Soc. Ind. Appl. Math.* **1957**, *5*, 32–38. [CrossRef]
24. Zhao, J.; Gong, M.; Liu, J.; Jiao, L. Deep learning to classify difference image for image change detection. In Proceedings of the International Joint Conference on Neural Networks, Beijing, China, 6–11 July 2014; pp. 411–417.