6ZD-07

A Detection Method of Customer's Products Selection Behavior Based on Top-view Images in Retail Environments

Jiahao Wen[†1] Muhammad Alfian Amrizal[†2] Toru Abe[†1,†3] Takuo Suganuma[†1,†3]

[†1]Graduate School of Information Sciences, Tohoku University
[†2]Research Institute of Electrical Communication, Tohoku University
[†3]Cyberscience Center, Tohoku University

1 Introduction

Effective ways of data collection for supporting a good marketing plan are essential to retail. In retail environments, customer's behavior when selecting products provides valuable information toward improving marketing effectiveness. Compared to traditional retail, smart retail collects customer's behavior effectively by ubiquitous cameras in retail stores. However, existing methods fail to recognize some important customer's behaviors such as product reinstatement behavior, i.e., the condition when a customer takes a product off the shelf, but end up not purchasing it and reinstate it back to the shelf or change it with another product. Recognition of such behavior is important to further improve marketing effectiveness.

In this research, we propose a method to recognize customer's behavior with top-view images from a normal RGB camera. The proposed method records customer's hands' positions and products in hands on a sequence. This sequence is analyzed to output a history of customer's primitive behavior including product reinstatement. It reveals customer's product selection process in detail, which is helpful for making further marketing plans.

2 Related Work

Surveillance camera is the typical camera used in retail environments. However, due to the high probability of occlusions, it's difficult to detect the interaction between customer's hand and a product using this camera [1]. Therefore, many researchers recently proposed the use of top-view cameras to detect such interactions [2], [3]. Top-view camera is promising because it avoids lots of occulsions which results in hands being exposed to the camera clearly during the hand-product interaction. To obtain top-view images, an RGB-Depth camera is installed on the ceiling [2], [3]. In these researches, recognition for reinstated products is realized but reinstatement and exchange aren't distinguished and the output interaction map only shows interactions' position without the order of interactions. In [4], experimental results reached a better accuracy with lots of behaviors' recognition, but reinstatement behavior is missing. Also, it provides output with interactions' order, but it is still useless since its lack of some basic interactions such as reinstatement.

To sum up, the existing methods mainly missed the recognition of some basic customer's behavior when selecting products, such as reinstatement. Compared to existing methods, our proposed method recognize those behaviors with the normal camera instead of depth camera. We recognize some basic selection behaviors, distinguish reinstatement and exchange which isn't done in [2],[3] and provide the order of interactions for each individual customer.

3 Proposal

Fig.1 shows our camera installation and top-view image, a camera is installed on the top of the shopping shelf to take top-view images. In our proposed method, received top-view images are sent to the detection part firstly as inputs. Hands and picked products will be figured out in this part. Hands' position and picked products' id of each frame are recorded on a state sequence. The next classification part receives this state sequence and identifies behavior by matching some certain behavior's pattern in the sequence. To avoid some unnecessary problems in this research, there are some assumptions for the environment.

(1) Each person walks into the frame is a different person. Thus identify each person in this store is avoided in this research.
(2) Each customer is alone. One won't interact with another one.

3.1 Detection Part

As we are recognizing interactions, this part receives images from top-view camera and de-

tects human and products in each frame. A faster RCNN [5] is applied to detect human's region and extract hands.

Since the camera is motionless, products' position on the shelf and the shelf area of all frames are preset. The shelf area is as shown in Fig.1(b) by red rectangle. With the data of products and shelf area, we are able to know if one's hand is inside or outside the shelf and if there is any product in the hand. Products are extracted from the region around hands also by faster RCNN [5] pretrained by images of products.

For frame t, the state sequence $S^{(n)}$ of customer n is updated/created to record his data. This $S^{(n)}$ stores hands' position h and products in hands p. Their details are shown as below:

- (S) It stands for customer's state. For customer n, $S^{(n)} = \{s_1^{(n)}, s_2^{(n)}, s_3^{(n)}, \ldots, s_t^{(n)}, \ldots, s_T^{(n)}\}$. "$T$" is the number of the latest frame. For frame t, $s_t^{(n)} = (h_t^{(n)}, p_t^{(n)})$.
- (h) It stands for hand's staying area. For frame t, $h_t^{(n)} \in \{IN, OUT\}$. IN is recorded when hand is inside the shelf area. Otherwise, record OUT.
- (p) It stands for products in hands. For frame t, $p_t^{(n)} \in \{null, 1, 2, \ldots, P_{max}\}$. "$null$" means nothing in hands. The other numbers are products' ID. P_{max} is ID's max number.

In this part, we detect hands and products in hands, then record results on $S^{(n)}$. This updated $S^{(n)}$ is sent as outputs to the next part.

3.2 Classification Part

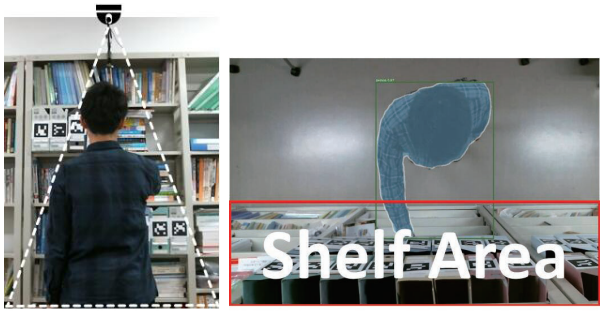
This part receives $S^{(n)}$ from detection part. Each behavior has a certain pattern and we classify those behaviors by matching the corresponding pattern in $S^{(n)}$. For instance, here we got a part of $S^{(n)} = \{(IN, null), (IN, 1), (OUT, 1)\}$.

This is a behavior's pattern of picking product. $(IN, null)$ means hands into shelf area with nothing. Then $(IN, 1)$ means product 1 is picked up in shelf area. Finally $(OUT, 1)$ means hands leave from shelf area with product 1.

Briefly, this part classifies behavior by matching certain patterns. And rewrite $S^{(n)}$ with classified behaviors. The final $S^{(n)}$ is a behavior sequence which reveals customer's shopping process. Therefore, this research is able to recognize the customer's behavior which are missing in existing methods, such as reinstatement.

4 Experiment

Since the experiment is unfinished, there is only introductions about finished parts. In Fig.1(a), a shelf with books is used to simulate retail environment and a camera is installed on the top of the shelf. Top-view image from the camera is shown as Fig.1(b). With fast RCNN [5], human's region is successfully detected and shown as the green rectangle in Fig.1(b). Hand's sequence H updates when the detected body cross the red area, Products' images, position on shelf and id are also prepared.



(a) Camera (b) Top-view Iamge

Fig. 1: Camera Installation and Top-view Image

5 Conclusion

In this research, we propose a method to detect customer's behavior when selecting products by a normal camera and output shopping process of each customer. It provides possible solutions to incomplete recognition for reinstatement and output's lacking the order of interactions which are disadvantages in existing methods. As the experiment is unfinished, it is added to our future work to evaluate its performance.

References

[1] Standard Cognition: Autonomous Checkout, Real Time System v0.21., available from ⟨https://standard.ai/blog/the-next-autonomous-revolution/⟩ (accessed 2017-08-17).

[2] Frontoni, E., Raspa, P., Mancini, A., Zingaretti, P. and Placidi, V.: Customers' Activity Recognition in Intelligent Retail Environments, *ICIAP*, Vol. 8158, pp. 509–516 (2013).

[3] Liciotti, D., Contigiani, M., Frontoni, E., Mancini, A., Zingaretti, P. and Placidi, V.: Shopper Analytics: A Customer Activity Recognition System Using a Distributed RGB-D Camera Network, *VAAM*, Vol. 8811, pp. 146–157 (2014).

[4] Yamamoto, J., Inoue, K. and Yoshioka, M.: Investigation of customer behavior analysis based on top-view depth camera, *WACVW*, pp. 67–74 (2017).

[5] Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *NIPS*, Vol. 1, pp. 91–99 (2015).