Vol.2022-MBL-102 No.11
Vol.2022-UBI-73 No.11
2022/3/7

小売店舗向け行動分析支援技術：商品選び取りに関するRegion of Interestの自動抽出

小林　由枝[1,a)]　茂木　厚憲[1]　平井　由樹雄[1]　張　茜丹[2]　潭　志明[2]　鈴木　源太[1]

概要：カメラ映像から人の行動を分析する技術開発が盛んに行われており，特に小売店舗では人の購買行動解析をマーケティングへ応用する動きがある．しかし，行動のみをキーにして解析を行うと購買行動ではない行動でも購買行動として検出されることがあり，解析したい行動が頻繁に生じる領域（Region of Interest）を指定して解析することで，誤検知を抑制している．この領域指定は手動で設定する必要があり，他の店舗で使用する際にはレイアウトに合わせて再設定する必要があり手間となる．本論文では，購買行動である商品の選び取りの際に人の体の向きが通路の移動方向と比べてばらつきが生じやすいという傾向を利用して，購買行動が生じやすい領域を自動で抽出する技術を提案する．実験では，実際の小売店舗の監視カメラ映像を用いて提案手法の有効性を示した．

1. はじめに

小売店舗には防犯目的のために監視カメラが多数設置されているが，その映像を人の行動を分析する技術を応用してマーケティングに活用する事が期待されている．マーケティングに応用するには映像から「人の属性情報（性別，年齢など）」，「人の動線情報」，「棚前での行動分析」を取得する必要がある．その中でも特に棚前での行動分析では，顧客と棚上の商品とのインタラクションを詳細に分析することで，顧客がどのようにして商品を選び購入に至るのか（または購入に至らないのか）の情報が得られ，売り場改善などに活用することができる．

顧客が商品棚に手を伸ばし，商品の選び取りを行う動作を検出するには，図?に示すように画像中の商品棚と設定された領域に人が手を伸ばす動作を検出する必要がある．人が手を伸ばす動作は基本となる動作を用いて事前に定義されたルールから検出する手法 [1] があるが，商品棚のエリアは手動または画像解析により定義する必要がある．

本論文では，商品棚に囲まれた通路領域では顧客が商品棚をキョロキョロ見ながら歩く行動を取りやすいという性質と画像解析により得られる画像中の物体の情報を利用して，商品棚の領域を検出する手法を提案する．商品棚に囲まれた通路領域では顧客がキョロキョロしながら歩くため，通路を顧客が移動する方向に対して顧客の体の向きのバラつきが生じやすい．体の向きのバラつきが生じやすい領域周辺には商品棚があるため，画像解析で得られた物品のうちこの領域に近い物品は商品棚として分類される可能性が高い．商品棚の可能性が高い領域を商品棚として抽出することで，商品棚のエリアを自動的に定義することができる．本論文では商品棚に囲まれた通路領域を抽出する手法を提案し，実験により有効性の検証を行う．

本論文は以下のような構成となっている．2 章では関連研究について述べる．3 章では提案手法の詳細について述べる．4 章では実験について，5 章でまとめと今後の展望について述べる．

2. 関連研究

本章では，画像をピクセル単位で物体領域を識別する手法や人の行動分析についての紹介を行う．画像解析による物品の検出手法には Semantic Segmentation というピクセルごとの物品の識別を DeepLearning を用いて推定する手法がある．画像を入力すると物体認識結果を表すマスク画像の出力を得られるように学習したネットワークである．Semantic Segmentation の手法には画像の入力サイズを気にしなくて良いようなネットワーク構造のものや特徴量の周辺のコンテキストを利用するものなど様々な手法 [2], [3], [4] が提案されている．Semantic Segmentation により商品棚領域を推定しようとすると，商品棚に含まれている多種多様な物品を識別してしまい，商品棚全体とし

[1] 富士通株式会社
Fujitsu Limited, 4-1-1, Kamikodanaka, Nakahara-ku, Kawasaki, Kanagawa 211–8588, Japan
[2] 富士通研究開発中心有限公司
Fujitsu R and D Center Co., Ltd., 8 Jianguomen Outer St, Chaoyang, Beijing 100020, China
[a)] k.yoshie@fujitsu.com

図 1 手伸ばし行動検知と棚のリーチ分析の例．左図は人が商品棚に手を伸ばしているのを行動分析により検知した結果を示しており，右図は手伸ばし行動が検出された商品棚のヒートマップを示している．

て識別することが難しいという課題がある．

そこで，人の購買行動を検出することでその付近にある商品棚領域を検出することが考えられる．そのためには，人の行動分析が必要になる．行動分析は，DeepLearning がネットワーク構造の改良により動画にも適用されるようになることで，手法としての広がりを見せてきた．動画を何フレームか連続で学習可能な LSTM を用い動画に対して単一のラベルを割り振り学習させることで，特定の動作を検出する手法 [6] が提案されている．また，動画の内容を説明する文を自動生成する手法（キャプショニング）[7], [8] では，フレームの前後のつながりから動画の解析を行う．これらの手法で購買行動を検出しようとすると商品の選び取り行動の様々なバリエーションの動画を用意する必要があり，学習コストも高い．

前述の動画を直接用いた分析手法は，動画を直接入力して分析可能であるという利点があるが，分析させたい動作が増えると学習データも膨大になり，学習コストが非常にかかるという課題がある．そこで，学習データを削減するために動画を直接用いるのではなく，人物の骨格情報を用いる手法がある．骨格情報の取得には Kinect [11] のような 3 次元センサを用いる手法や Openpose [12] のような動画データから DeepLearning により骨格情報を取得する手法がある．

骨格情報を用いた行動分析では骨格情報の時系列データと動作内容のラベルをセットに学習させ，分析を行う手法 [9] が提案されている．しかし，この手法では学習に使用したラベルでの分析となり，解析したい行動が増えると学習データが増えてしまう課題が残る．そこで，行動分析自体は「歩く」,「走る」,「座る」などの基本的な行動を分析するが，解析したい対象に応じて人手でルールを設定し，設定されたルールと同じ行動が検出することで行動分析をする手法（Actlyzer）[1] が提案されている．Actlyzer では商品棚領域は手動で設定しているため，店舗ごとにレイアウトの変更があると手動で領域の修正が必要になり手間がかかる．また，センサから取得された人の動作から周辺にある人とインタラクションをしている物体の推定を行う手法 [10] の提案もされているが，購買行動分析ではセンサを顧客一人一人につけることは難しい．

3. 提案手法

本章では，行動分析支援技術についての説明と本論文で提案する商品棚領域付近の通路を検出する手法についての説明をする．

3.1 行動分析支援技術

小売店舗での購買行動分析をするには，従来商品棚領域を設定しその領域に人が手を伸ばしたという動作から購買行動検出を行っていた．しかし，商品棚領域の設定を手動で実施していたため，レイアウトごとに設定する必要があり，多店舗に導入する際には手間がかかっていた．そこで，商品棚領域を自動で設定する事ができると，多店舗への導入コストを下げることが出来る．

Semantic Segmentation のようなピクセル単位での物体認識を用いて商品棚領域を抽出することは前述したように難しい．そこで，Semantic Segmentation の認識結果と商品棚領域付近の通路領域では購買行動をとる人はキョロキョロしながら歩く動作（図**?**）をする傾向にあることを利用して商品棚領域の抽出を行う．Semantic Segmentation を小売店舗に適用すると，通路領域は検出可能であるが，商品棚領域は様々な物体を認識してしまい商品棚として認識することは難しい．しかし，購買行動をとる人が多い通路領域に隣接している認識結果は商品棚であると仮定して，Semantic Segmentation 結果を再ラベリングすることで，Semantic Segmentation だけでの認識結果を修正することが可能である．

3.2 提案手法：商品棚領域付近の通路検出

本章では商品棚領域付近の通路領域を抽出する部分と

図 2　購買行動が生じる通路（緑色）とそれ以外の通路を分離した結果．赤矢印は購買行動を生じている人の体の向き，青矢印は通路の移動方向を示す．

Semantic Segmentation 結果を再ラベリングする部分のうち，商品棚領域付近の通路領域を抽出する手法について述べる．商品棚領域付近の通路領域は監視カメラに対して，移動方向が正対していると仮定して分析をしている．監視カメラは防犯を目的として設置されているケースが多く，死角をなくすため観察したい通路に対し映りやすいよう平行または垂直に近い方向にカメラを設置しているためである．

提案手法は以下のアルゴリズムとなっている．図?にはフローを示している．

(1) 時系列の骨格情報データを取得 [13]
(2) 画像中の同一人物の画像の水平/垂直方向に対する一定距離以上の移動軌跡を取得
(3) 画像の水平な移動方向の通路領域を抽出
(4) 画像の垂直方向の移動軌跡のヒストグラムをクラスタリングし，垂直方向の通路領域を抽出
(5) 2 で求めた移動軌跡から各クラスに対する通路の移動方向ベクトルを算出
(6) 5 で求めたベクトルと 1 の骨格情報から得られる人の体の向きのなす角を計算
(7) 6 で計算された角度のうち閾値以上の角度を多く含む領域を囲う多角形の座標を算出

4. 実験

実験には実際の小売店舗に設置されている 8 台の監視カメラ映像を使用した．撮影時刻は午前 11 時から正午までの 1 時間で，一般的には小売店舗の混雑時間帯である．顧客は 1 つの動画あたりのべ 300 人以上が通過している．図 4，5，6，7 に抽出された商品棚に囲まれた通路領域と Semantic Segmentation 結果を示している．

図の商品棚に囲まれた通路領域を抽出した結果から提案手法で抽出したい領域がおおよそ特定できていることが分かった．しかし，棚の際や画像中の通路領域があまり多く映り込んでいないところなどは，提案手法での抽出ができていない．これは，分析に用いた映像中の人が商品棚の際近くであまり近寄っていない領域が生じているためである．また，通路領域があまり多く映り込んでいないところでは，そもそも人が多く観測されないので，提案手法の (7) においてヒストグラムの閾値を取る際に除かれているためだと考えられる．したがって，提案手法ではカメラと正対に近いような配置となっている商品棚に囲まれた通路領域の抽出については上手く動作すると考えられる．実験で検出されなかった商品棚に囲まれた通路領域については，Semantic Segmentation 結果とマージすることで商品棚の際の領域を抽出することができる．また，通路領域があまり多く映り込んでいないところも他のカメラ映像を使う事でカバーできる．

今回の実験では Semantic Segmentation 結果との合わせ込みまでは行っていないが，抽出された通路領域や商品棚領域との比較を行うために Semantic Segmentation 結果も示している．Semantic Segmentation 結果より通路領域の抽出は行えているが，商品棚領域では商品自体を細かく領域分割している結果となり，商品棚全体としての検出が困難であることが分かる．そのため，提案手法で得られた商品棚に近い領域の通路を用いることで，その近辺で抽出された通路ではない領域は商品棚であると再ラベリングできる．今回の実験では，商品棚領域を抽出することが出来る事が示せた．

5. まとめと今後の展望

本論文では，人の移動軌跡や体の向きのバラつきを利用して商品棚に囲まれた通路領域の抽出を行う手法の提案を行った．実験では，実店舗に設置された監視カメラ映像を用いて，商品棚に囲まれた通路領域の抽出と Semantic Segmentation による領域分割を行った．これにより，提案手法を用いることで商品棚に囲まれた通路領域の抽出が可能であることが分かった．また，提案手法の結果と Semantic Segmentation 結果とをマージすることで，商品棚に近い領域の通路を用いることで，その近辺で抽出された通路ではない領域は商品棚であると再ラベリングの可能性も示すことが出来た．今後は実店舗の監視カメラ映像を用いての再ラベリングの検証や三次元的な奥行をカメラ映像から取得し奥行情報を利用した商品棚領域の認識などにつなげていけると考えている．

参考文献

[1] 杉村由花，内田大輔，鈴木源太，遠藤利生：“映像から人の様々な行動を認識する「行動分析技術 Actlyzer」”，人工知能学会第 34 回全国大会論文集，4Rin1-57 (2020)
[2] J. Long, E. Shelhamer and T. Darrell, ”Fully convolutional networks for semantic segmentation,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015 pp. 3431-3440.
[3] V. Badrinarayanan, A. Kendall and R. Cipolla, ”SegNet:

図 3 Semantic Segmentation への入力画像と出力画像の例

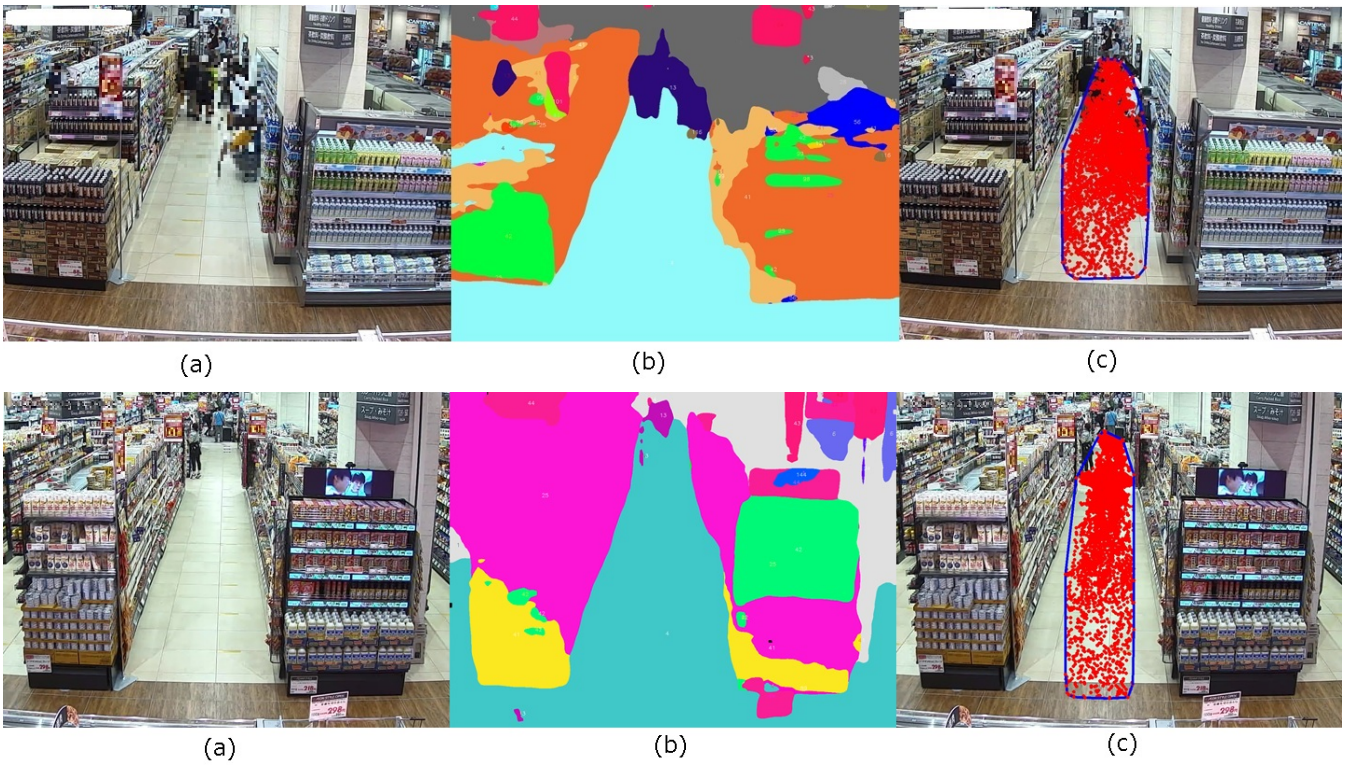

図 4 商品棚に囲まれた通路領域を抽出した結果☒．(a) 入力画像 (b)Semantic Segmentation 結果 (c) 商品棚に囲まれた通路領域を抽出した結果の画像．青色が抽出した領域，赤点は領域内の人の移動軌跡を示す．

A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017

[4] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-Assisted Intervention(MICCAI), 2015.

[5] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[6] Carreira et. al, "Quo vadis, action recognition? a new model and the kinetics dataset", CVPR 2017.

[7] Vinyals, O., Toshev, A., Bengio, S., Erhan, D. Show and tell: A neural image caption generator. In CVPR, 2015.

[8] J. Lu, J. Yang, D. Batra, and D. Parikh. Neural baby talk. In CVPR, 2018.

[9] S. Yan, et. al."Spatial Temporal Graph Convolu-

Vol.2022-MBL-102 No.11
Vol.2022-UBI-73 No.11
2022/3/7

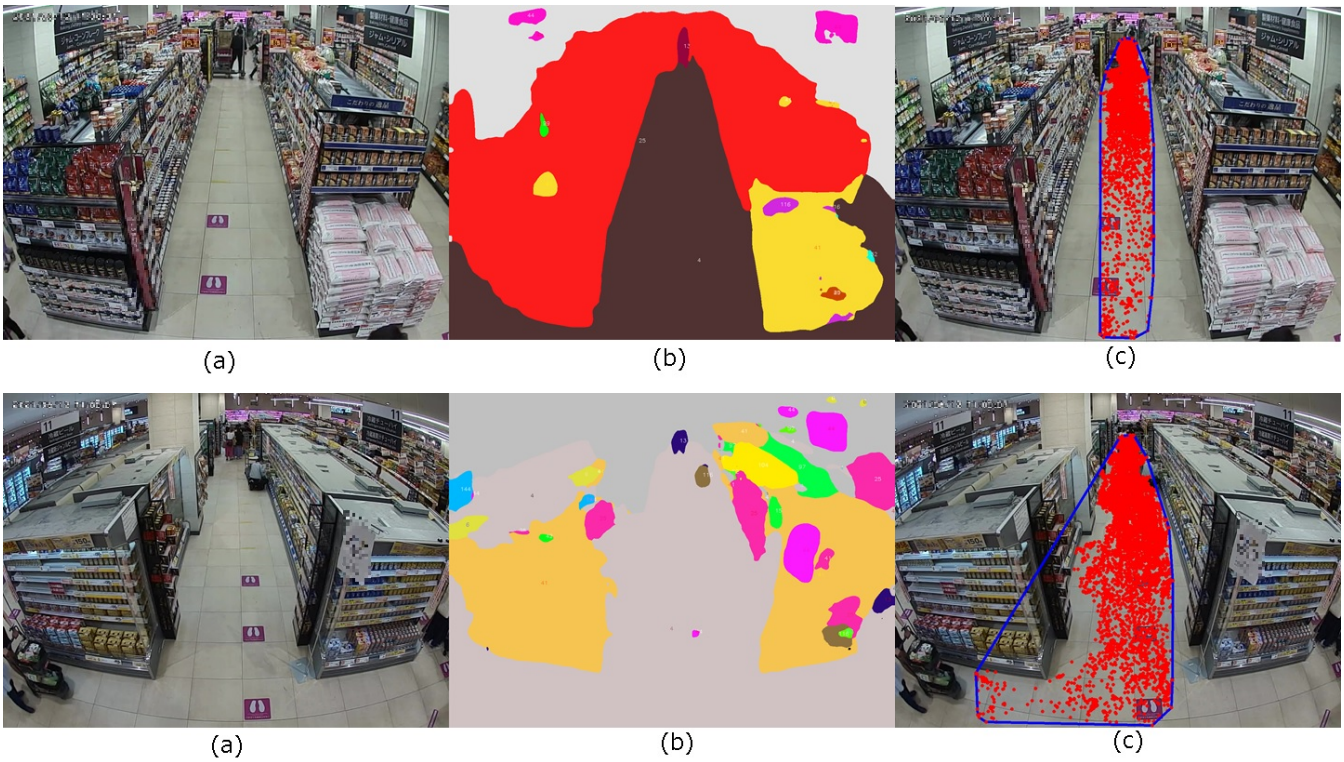

図 5 商品棚に囲まれた通路領域を抽出した結果☒．(a) 入力画像 (b)Semantic Segmentation 結果 (c) 商品棚に囲まれた通路領域を抽出した結果の画像．青色が抽出した領域，赤点は領域内の人の移動軌跡を示す．

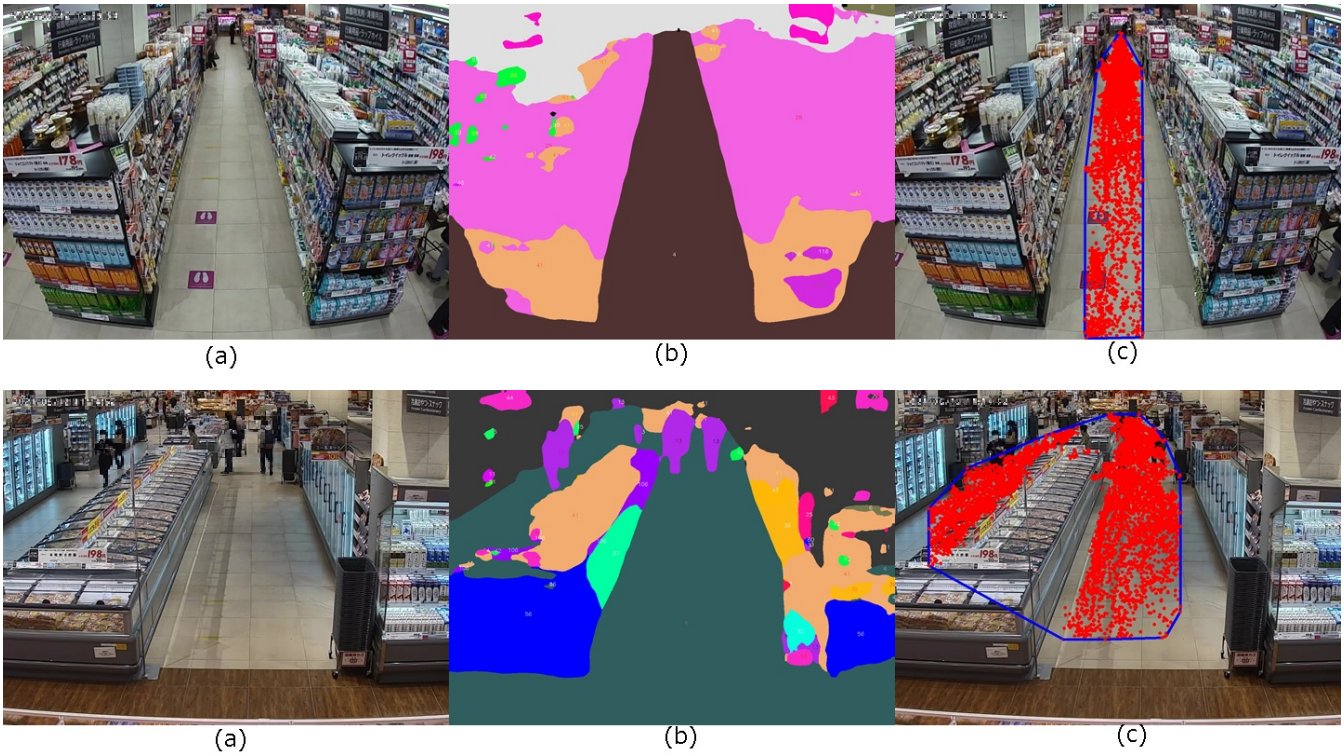

図 6 商品棚に囲まれた通路領域を抽出した結果☒．(a) 入力画像 (b)Semantic Segmentation 結果 (c) 商品棚に囲まれた通路領域を抽出した結果の画像．青色が抽出した領域，赤点は領域内の人の移動軌跡を示す．

tional Networks for Skeleton-Based Action Recognition”, AAAI, 2018.

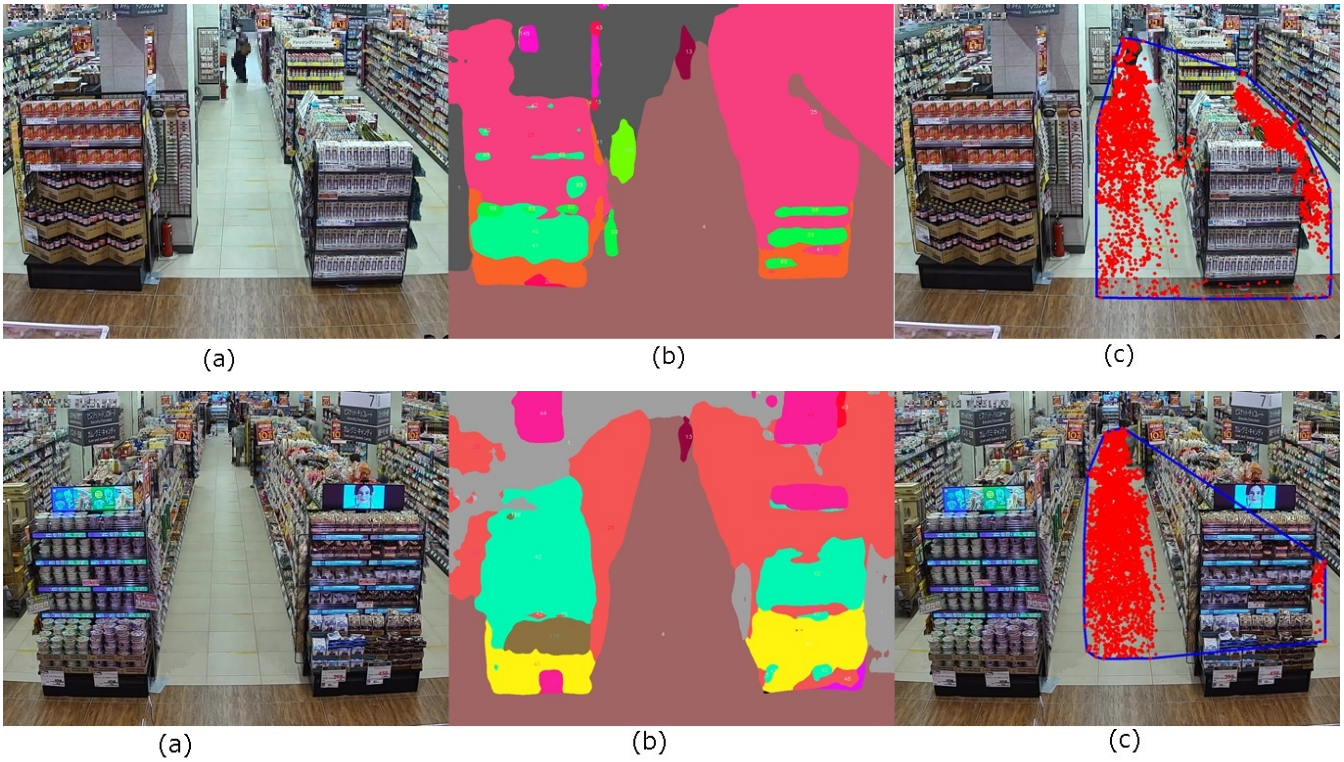

図 7 商品棚に囲まれた通路領域を抽出した結果☒．(a) 入力画像 (b)Semantic Segmentation 結果 (c) 商品棚に囲まれた通路領域を抽出した結果の画像．青色が抽出した領域，赤点は領域内の人の移動軌跡を示す．

[10] Yinyu Nie and Angela Dai and Xiaoguang Han and Matthias Nie?ner, "Pose2Room: Understanding 3D Scenes from Human Activities," arXiv, 2021.
[11] J. Shotton et al., "Real-time human pose recognition in parts from single depth images," CVPR 2011
[12] Z. Cao, G. Hidalgo, T. Simon, S. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields" in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 01, pp. 172-186, 2021.
[13] Huigang Zhang, Liuan Wang, Jun Sun, Nobutaka Imamura, Yusaku Fujii, and Hiromichi Kobashi, "CPNAS: Cascaded Pyramid Network via Neural Architecture Search for Multi-Person PoseEstimation." CVPR 2020 Workshop on Neural Architecture Search and Beyond for Representation Learning.