CH-001

デプスセンサによる把持判定に基づく商品棚前動作認識システム
In-Store Customer-Product Interaction Recognition System based on Grasped Object Detection in Top-view Depth Images

白石　壮馬† 　井下　哲夫† 　岩元　浩太† 　西脇　大輔†
Soma SHIRAISHI Tetsuo INOSHITA Kota IWAMOTO Daisuke NISHIWAKI

1. はじめに

コンビニエンスストア・スーパーなどの小売店舗では，POS 端末から得られる売上情報だけでなく，購買客の実際の行動を把握・解析した情報を活用することで，店舗レイアウト選定・マーケティング効果を高めたいというニーズがある．例えば，動線解析を活用すると，店舗内の大局的な人の流れや滞留をデータ化でき，店舗全体のレイアウトやエリアごとの興味度を評価できる．一方で，商品棚前での人物個々の購買動作は，棚レベル，商品レベルでの興味を反映しており，棚割検討やキャンペーン商品選定などのより細かい解析・活用を行う上で重要な情報である

商品棚前での購買動作に対する解析はこれまでも提案されている．棚レベルでの解析例として，[1][2][3]では商品棚前で「棚を見ている」「手で持った商品を見ている」「手を伸ばす」などの動作をデプスセンサやカラーカメラを用いて解析している．さらに詳細な解析として，[3]では，手の軌跡の情報から「商品を取得した」「返却した」「接触したが取得に至らなかった」という商品に対する動作推定まで行っている．このような商品に対する動作は，商品レベルでの購買客興味情報を集積するために非常に重要な情報であると考えられる．

商品に対する動作取得にさらに特化したシステムとして[4]では Top-View（上から下を見おろす画角）の RGB-D センサを用い，商品を棚の「どの位置」から，「取得した」か，「返却した」か，もしくは「接触したが取得に至らなかった」かを識別し，購買客の興味情報を集積するシステムを提案している．接触位置の情報と棚割を対応付けることで，動作対象となった商品まで特定できる点で，商品レベルの解析という目的に対して有効なシステムである．

しかし，このシステムには技術的な課題も存在する．[4]では上記 3 種類の「棚前動作」を，購買客の棚への手伸ばし前後の画像での色変化とサイズ変化に基づき識別する．しかし，手伸ばし前後の，腕の向きや洋服袖による色変化のバラつき，および商品の違いによるサイズのばらつきによって，正しい識別結果が得られない場合がある．また，実用面では，現状，監視カメラのように PC レスで動作する RGB-D センサがほぼ無いため，比較的選択肢の多い，デプスセンサのみで動作するシステムへの要求がある．

本稿では，Top-view のデプスセンサのみを用いた棚前動作認識システムを提案する．新しい棚前動作認識手法は課題であった色・サイズ変動に対して頑健で，デプス画像のみを用いた場合にも認識精度を向上できる．以降，第２章で提案システムを説明する．次に第３章で，実験について記述し，第４章で実店舗での解析例を紹介する．最後に第５章でまとめを述べる．

† NEC データサイエンス研究所

図 1 提案システムの模式図
(a)センサ設置, (b)デプス画像，(c) 棚ヒートマップ

2. 提案システム

2.1 システム概要

図 1 に提案システムの模式図を示す．提案システムは Top-View のデプスセンサを用いて商品の「取得」「返却」「接触のみ」の３つの棚前動作を認識するシステムである．Top-View 配置により，購買客の商品に対する動作をオクルージョンなく捉えることができる．また，デプスセンサの距離情報で，棚のどの位置に対して動作が行われたかを正確に取得できる．得られた結果は図 1(c)のように棚上のヒートマップとして可視化される．また接触位置に基づいて商品ごとに購買客の興味を解析することもできる．

提案システムの棚前動作認識は，“人物検出・追跡”，“接触区間検出”，“動作識別”から構成される（図 2）．次節以降，各ステップについて詳細を述べる．

2.2 人物検出・追跡

本ステップでは，画像中に現れた人物に固有の ID を割り当て，画像外に出るまで同定し続ける．これをデプス画像での頭部検出および追跡によって行う．

具体的には，まず，デプス画像の背景差分により前景を取得する．次に，Non Minimum Suppression により，カメラの向きに凸型の領域を頭部候補として取得する．これらの頭部候補には頭部以外にも，肩や腕，店舗内に置かれたコンテナ，引き出し等が含まれる可能性がある．したがって，最後に， 各頭部候補に対して 2 クラス識別を適用し，頭部以外を除外する．識別には CNN 識別器を用いる．

追跡にはアピアランスベースの KCF Tracker[5]を用いることで，頭部への部分的なオクルージョンやしゃがみなどにより頭部の検出漏れが発生しても追跡できる．

2.3 接触区間検知

本ステップでは 2.2 で頭部検出した人物に対し，手の位置を求めるとともに，棚に手を伸ばしている時間的区間の検出を行う．この区間に基づいて，次ステップにて動作識

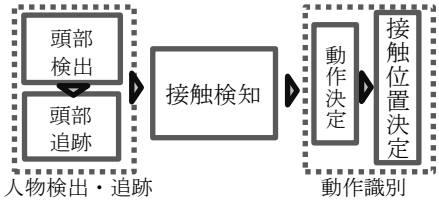

図 2　提案システムのフロー

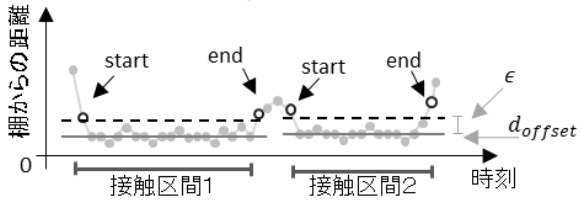

図 3　接触区間検知

別と棚上の接触位置決定が行われるため，重要なステップである．

2.2 で得られた前景から，手領域の候補を抽出する．まず，前景各点に対し，３次元座標を用いて k-means クラスタリングを行い，複数の小領域へと分割する．これら小領域を手領域候補とする．さらに，後述する接触判定を行い，接触したと判定された小領域のみを手領域として扱う．

従来手法[4]では，棚の前面を 3 次元平面で近似し，平面から手までの距離が閾値以下となる場合に接触としていた．ただし，実用上，設定された 3 次元平面と実際の棚とのずれや，ポップのせり出し，商品の飛び出しによってオフセット誤差が発生する．すなわち，手を伸ばす位置や商品の配置によって接触時の距離が異なることになる．オフセットの最大値に合わせて閾値設定することで上記問題に対応した場合，連続する細かい接触を一つの区間にまとめてしまう可能性があり，適切に接触区間検知ができない．

本システムでは，接触区間において棚から手までの距離の変動がほぼ一定となること（図 3）に着目して区間検出を行う． 具体的には，まず距離が時間方向で局所的に最小となる時刻を検出する．次に，距離値誤差ϵを仮定し，上記時刻周辺の値に対して，外れ値に強い Tukey 損失によるフィッティングを行う．こうして得られた値d_{offset}は，接触区間のみの距離の平均値と考えることができる．したがって，距離が$d_{\mathrm{offset}}+\epsilon$以下となる区間を抽出することで，１つの接触区間を検出する．なお，最小の点が複数時刻得られた場合，各最小の時刻に上記処理を適用し，複数区間を検出する（図 3）．ここで，得られた複数区間に重なりがある場合，これらをマージし，一つの区間として検出する．

2.4 動作識別

本ステップでは，2.3 で得られた各接触区間について動作と棚上の接触位置を決定する．従来手法[4]で識別に用いられる色・サイズ変化量は，商品によってその大きさが異なり，また，手や腕の姿勢変動による変化を含むため，商品の有無による変化のみを正確に検出するのは困難である．提案手法は，色・サイズ変化を用いず，接触区間前後それぞれの画像自体から把持の有無を識別する．そして，その組み合わせに基づいて動作を識別する．

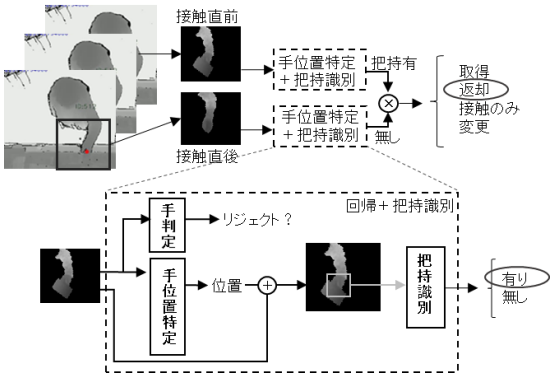

図 4　動作識別概略図

(a) センサ設置例　　(b) 商品棚

図 5　実験環境

把持の有無を正確に識別するためには，姿勢や位置の変動を除外し，手の領域のみに注目することが有効であると考えられる．したがって提案手法は機械学習により手の位置を特定し，その周辺のみを用いて把持の有無を識別する．

動作識別ステップの概略図を図 4 に示す．手を特定するための候補領域として，まず，接触点周辺のパッチを切り出す．パッチサイズはデプスセンサの距離値に基づき，実空間で一定になるように決定する．商品を取り出す場合と商品を棚に戻す場合，接触点は商品の先端となるため手の位置が変動する．そこで，ある程度広い範囲を切り出すことでパッチ内に手が含まれるようにする．

次に，得られたパッチ内にて図 4 下部に示すような手の領域特定を行う． 手の領域を座標，角度，サイズなどのパラメータで表現し，パッチを入力としてこれらを回帰学習したＣＮＮを用いる．最後に，得られた手の領域のみを切り出し，把持有無の 2 クラス識別を適用することで把持判定を行う．回帰と識別にはそれぞれ batch normalization を含む AlenxNet[6]を用いる．

上記の識別は，接触区間前後 N フレームについて適用でき，信頼度が最大のものを採用することで識別を安定化できる．具体的には，取得直後の商品の一部が一旦手で隠れてしまう場合等に精度向上が期待される．なお，2.3（接触区間検知）では接触判定された小領域を「手領域」であると仮定したが，実際には，手以外（頭部，肩，荷物など）が接触することもありうる．本システムでは，パッチに対して手位置特定と同時に手/それ以外の識別も行い，手でないと判断されたパッチについては棄却する．

こうして，得られた前後それぞれの把持有無の識別結果の組み合わせに応じて，「把持無し→把持無し＝接触のみ」「把持無し→把持あり＝取得」「把持あり→把持無し＝返却」「把持あり→把持あり＝変更」として動作識別結果を得る．そして，最後に接触位置を決定する．接触区間直前における接触点の３次元座標を，棚平面に射影して得られた 2 次元座標を接触位置とする． この位置に基づき，あらかじめ対応付けられた商品を決定することができる．

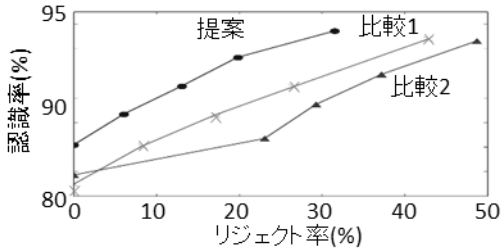

図 6　手位置特定有無による動作識別精度比較

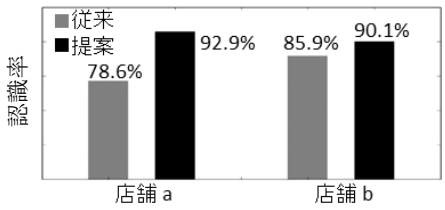

図 7　提案手法と従来手法の動作識別精度比較

3. 実験と考察

提案システムについて，棚前動作の認識精度，および，棚平面の接触位置精度の２点を評価した．従来手法との比較のため，データ撮影には RGB-D センサ（Kinect v2）を使用した．ただし，本提案システムは RGB を用いないため，同センサで取得したデプスデータのみを使用した．データは異なる２店舗にて取得した．図 5 (a)にそれぞれのセンサ設置外観を示す．設置の高さはそれぞれの天井高に合わせて，3m と 2.7m とした．

3.1 棚前動作の識別精度評価

まず，動作識別ついて２点の比較実験結果を報告する．次に提案システム全体の精度評価結果を報告する．

3.1.1 動作識別評価

A)　手位置特定の有効性検証

手位置特定の効果を検証するために提案手法と以下の比較手法 1, 2 を比較した．比較手法 1 として，パッチ全体を識別する場合，比較手法 2 として，接触点を中心に一定領域を切り出し識別する場合を用いた．比較手法 1 は入力に手や腕の姿勢変動を含み，比較手法 2 は入力に，商品に応じた位置変動を含む．

提案手法では，手位置特定のために回帰するパラメータとして，パッチ中の手の位置(x, y)を用いた．識別の際は，入力画像パッチから，手の位置を基準に固定サイズで領域を切り出して識別器に入力した．また，比較手法２における切り出しサイズも共通の値を使用した．図 4 に示すような 9798 枚のデプス画像に左右反転を施し，合計 19596 枚を学習に用いた．また，評価には接触前後１枚ずつのデプス画像 1717 ペアを用いた．図 6 に識別結果に対する信頼度閾値を変化させた場合のリジェクト率-認識率のグラフを示す．各リジェクト率において，提案手法が比較手法より高い精度を示している．この結果より，パッチから手の位置を特定して識別する提案手法の有効性が確認された．

(a)　接触のみの例　　(b)取得の例

図 8　従来法の誤認識例 (a)は色変化により取得，(b)はブロブ面積減少により返却と判定

B)　従来手法との比較

提案手法を従来手法[4]と比較し，有効性を検証した．従来手法の具体的なアルゴリズムを述べる．接触区間前後のパッチ画像上で画像を上下左右にずらしながらテンプレートマッチングを行い，最大相関値が閾値以上であれば，「接触のみ」とする．また，相関値が閾値未満の場合は，接触前後で変化が起こったと判断し，パッチ内のブロブサイズを比較する．ブロブサイズが増加していれば「取得」とし，減少していれば「返却」と識別する．

テストには実店舗にて撮影した RGB-D 画像を用いた．店舗 a での 28 動作および店舗 b での 71 動作を使用した．

図 7 に結果を示す．従来手法に関しては，閾値を変動させたうえで最大の精度を掲載している．一方，提案手法に関しては，テストセットとは異なるデータ（A と同様）にて学習した識別器を使用しており，また，信頼度の閾値による棄却は行っていない．結果から，提案手法では，RGB 情報を用いることなく，従来手法よりも高い精度を得られることを確認した．図 8 には，従来手法で誤認識し，提案手法では正しく認識した例を示す．

3.1.2 提案システム全体の精度評価

実運用ではシステム全体の精度が重要となる．本実験では 3 日間の実店舗データにおける延べ 118 名，196 動作を用いてシステムの全体フローの精度を評価した．表 1 に評価結果を示す．全動作のうちの 70.4%が正しく検出・識別されていることを確認した．なお，最下段の商品においては，動作点の頭部による隠れ，商品がカメラから見えづらい，デプスのノイズが大きくなるなどの理由により，動作識別が困難な場合が多かったため，それらを除いた場合についても評価した．その結果，全体の 75.7%について，正しい動作の検出・識別ができていること確認できた．

各ステップの精度を確認すると，頭部検出・追跡については，検出率 97.4%，適合率 100%であった．この結果により，頭部検出はほぼ漏れや誤検出なく行われていることが確認された．また，接触区間検知については，提案手法の区間検出法で検出率 83.7%であった．同実験にて固定閾値未満を接触区間とした場合[4]，最適な値を設定した場合であっても 70.3%の検出率であったことから，提案する接触区間検出法の効果が確認された．しかしながら，接触検知の失敗は全体的な精度低下の大きな要因となっている．接触検知失敗の主な理由は，デプスのノイズが想定よりも大きく，接触区間内でも棚平面への距離が大きく変動してしまうことで，接触区間が分断されてしまうことであった．これに関しては時間方向の平滑化や，より頑健な距離算出が必要だと考えられる．最後に，動作識別精度に関しては前節の実験とほぼ同等の値が得られていることが確認された．

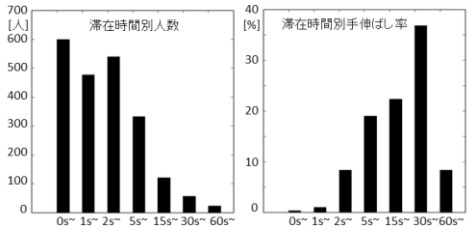

図 9 棚前の滞在時間別人数と手伸ばし率

表 1　提案システム各ステップの精度

項目	評価指標	値 (%)
頭部検出	検出率	97.4
接触区間検知	検出率	83.7
動作識別	認識率	87.7
全体フロー	検出率	70.4
動作識別（下段除く）	認識率	90.3
全体フロー（下段除く）	検出率	75.7

3.2 接触位置精度

商品特定の際に重要となる接触位置精度について評価した．図 5 (b)のような一般的な商品棚において，各段に 15cm ごとにマーカーを配置し，マーカー上に置かれた商品に計 3 回ずつ接触して精度検証した．表 2 に段ごとのx方向，および y 方向絶対誤差を，接触区間直前と直後についてそれぞれ示す．直前と直後の値を比較すると，平均的には 2cm 程度の差しかないものの，誤差最大値は直前の方が大幅に低いことがわかる．このことから，棚の接触位置算出には接触区間直前の時刻での位置を用いるのが望ましいと判断できる．

また，平均絶対誤差x方向:1.60cm, y方向:12.0cm という値に関しては，ペットボトルの幅がおよそ 7cm，平均的な棚段の高さが 25cm でことを考慮すると，ペットボトルの半分程度の幅の商品であれば，商品を特定できる精度が得られていると考えられる．なお，段ごとの平均値を比較すると，下段，すなわちデプスセンサから遠ざかるにつれて誤差が増加している傾向が確認された．

表 2 接触位置絶対誤差

段	直前 x	直前 y	直後 x	直後 y
1	1.27 cm	11.9 cm	3.00 cm	8.26 cm
2	1.19 cm	10.9 cm	2.65 cm	11.6 cm
3	1.87 cm	11.4 cm	3.66 cm	11.5 cm
4	2.04 cm	13.9 cm	3.91 cm	9.65 cm
平均	1.60 cm	12.0 cm	3.03 cm	10.3 cm
最大	5.08 cm	33.7 cm	19.2 cm	48.0 cm

4. データ可視化と応用例

提案システムのアウトプットの有用性について考察するため，実店舗 4 日間のデータの可視化を行った．3.1.2 で精度の低かった下段については可視化の対象外とした．まず，図 9(a)は滞在時間と手伸ばしした人数のグラフである．対象の棚はレジへの通路に配置されており，通過する人数が多いことがグラフに反映されている．また，図 9(b)から，30 秒以上滞在した購買客のうち，36.8%が手伸ばしに至ったことがわかる．これは逆に 63.2%が手伸ばしに至っていないことを意味しており，50%が反対側の棚を見ていたと仮定しても，残りの約 13%に関しては，「検討の結果，商品への手伸ばしに至っていない」ことになる．このことから，例えば，「長く滞在した場合，動的にキャンペーンを打つ」などの活用が考えられる．

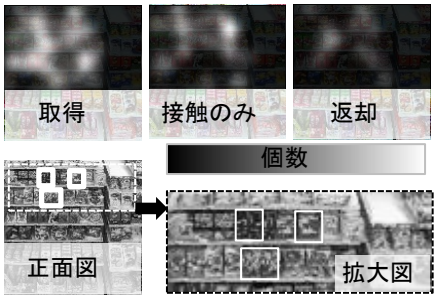

図 10　各動作の棚ヒートマップおよび棚正面画像

図 10 には動作毎の棚ヒートマップを示す．ヒートマップは接触点ごと上下左右の誤差を加味したカーネルを適用し，合計して生成した．図 10 より矩形で囲まれた商品について，接触や返却 1 が多くなっていることが確認できる．これは，購買客が商品自体には興味を示しているものの，購入に至らないことが多いことを示している．このことから，例えば「キャンペーンを打つ場合には，これらの商品を優先すると効果が高い」といった検討が可能となる．

5. おわりに

本稿では，Top-view デプスセンサにより購買客の商品に対する「取得」「返却」「接触のみ」の棚前動作を精度よく認識するシステムを提案した．複数の比較実験にて提案手法の動作認識精度優位性を示した．さらに，接触位置精度を評価し，一定サイズ以上の商品であれば，棚の接触位置の情報から商品特定が可能であることを確認した．最後に，実データを用いた可視化と活用方法を示した．

今後の課題は，ノイズに頑健な棚接触区間検知方法の開発と，より細かい商品についても安定して商品特定を行うための接触位置精度向上である．

参考文献

[1] B. Singh and T. K. Marks and M. Jones and O. Tuzel and M. Shao, “A Multi-Stream Bi-Directional Recurrent Neural Network for Fine-Grained”, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1961-1970 (2016).

[2] J. Yamamoto, K. Inoue and M. Yoshioka, "Investigation of Customer Behavior Analysis Based on Top-View Depth Camera," 2017 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 67-74 (2017).

[3] J. Liu and Y. Gu and S. Kamijo, “Customer Behavior Recognition in Retail Store from Surveillance Camera,” 2015 IEEE International Symposium on Multimedia (ISM), pp.154-159 (2015)

[4] D. Liciotti, M. Contigiani, E. Frontoni, A. Mancini, P. Zingaretti,and V. Placidi,” Shopper analytics: A customer activity recognition system using a distributed rgb-d camera network,” International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage, pp. 146-157 (2014)

[5] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," in IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 583-596 (2015).

[6] Alex Krizhevsky and Sutskever, Ilya and Hinton, Geoffrey E, “ImageNet Classification with Deep Convolutional Neural Networks,” NIPS2012, pp.1097-1105(2012).