

# インスタ・マーチャндаイジングのための カメラ映像を活用した来店客行動認識

Shopper Behavior Recognition for In-Store Merchandising using Camera Image

大坪 紹二  
Shohji Ohtsubo  
荻部 朋幸  
Tomoyuki Karibe

守山 隆昭  
Takaaki Moriyama  
スギリ プラナタ  
Sugiri Pranata

石原 健  
Takeshi Ishihara  
イエン シュイ  
Yan Xu

## 要 旨

カメラ映像から来店客の行動特性を抽出し、小売店舗での売り上げ向上施策の仮説を提案するために、来店客の店舗内行動取得・分析システムを開発している。混雑する売場では、複数の人物が重なり合う機会が多く、画像認識による人物検出・追跡に誤認識が発生しやすい。そこでRGB-Depthカメラにより輝度情報と深度情報を取得し、画像中の輝度勾配と深度勾配を用いた機械学習により、人物検出を行う方式を導入した。その結果、輝度情報のみを用いる手法では適合率99%のときの再現率1%のケースに対しても、再現率97%での認識が可能となり、実店舗環境における人物検出において深度情報の利用が有効であると結論付けることができた。また、RGBカメラによる人物追跡における、画質と記録フレームレートによる誤結合発生率を評価し、記録フレームレートを高めることが誤結合発生を低減させることを明らかにした。

## Abstract

We are developing a system to detect and analyze human behavior for retail stores, which extracts undiscovered behavioral features of shoppers from camera images, in order to propose innovative ideas for in-store merchandising. Human detection and tracking based on image feature matching is likely to fail in an environment with various backgrounds, such as in crowded stores. Therefore, we have implemented a method of human detection based not only on Histogram of Oriented Gradients (HOG) descriptors but also on Histogram of Depth Difference (HDD) descriptors acquired from machine learning an image from an RGB-Depth sensor. As a result we have succeeded in reaching a precision of 99% and recall rate of 97% for human detection, even for cases which can rarely be detected only by using HOG from an RGB image. We also discovered that we could reduce the error rate of human tracking by increasing the frame rate of the image rather than the quality.

## 1. はじめに

小売・流通業界では、ISM (In-Store Merchandising) と呼ばれる、市場の要求に合致した商品の品揃えと陳列を、科学的手法を用いて構成して来店客に提示することにより、店頭での売場生産性を向上させる活動が一般的に行われている。最も実施されているISMの1つは、POS (Point Of Sale) データ分析であり、POSデータ分析は、POSレジスタで取得した物品販売の売り上げ実績を分析し、来店客の購買行動特性を客観的に把握することにより、来店客により受け入れられる商品や売場を提供し、売り上げ向上を図る活動である。

店舗における購買行動モデルは以前から研究されており[1]、第1図に、購買に至る店内行動と各行動を捉えるためのデータ、および、データを活用した売り上げ拡大施策例の関係の一例を示す。店舗に入った来店客は、興味のない商品棚、興味をもった商品棚に対してそれぞれ「通過」と「立ち寄り」を行い、各商品やPOP広告に対する「視認」、気になった商品に対する購入の「選択検討」

を行った後、最終的に購入判断した商品の「購買」を行う。これら5つの行動は店舗内の購買行動モデルと定義することができる。従来から行われているPOSデータ分析は、購買に至る店舗行動モデルのうち、第5の「購買」後の結果を分析する行為である。

	取得データ	データ活用方法
店内 回遊	店舗内の客動線データ	・商品をどこに備えるか、何と何を並べるか
	ステップ1 通過	・各売場を通ったか否か (通過率) ・他の売場に対する通過率の比較 ・ターゲットが意図どおりに移動したか
	ステップ2 立ち寄り	・各売場に立ち寄ったか否か (立ち寄り率) ・配置する商品カテゴリーの選択 関心の有無
	ステップ3 視認	・各商品やポップに視線を向けたか否か (視認率) ・接触率の高いコミュニケーション 方法の改善 (POPや商品の高さなど)
	ステップ4 選択検討	・商品を手にとったか否か ・商品をかごに入れたか否か ・商品を棚に戻したか否か ・選ばれたのになぜ購入されなかったか、購入されるためにはどうすれば 良いかの検討
ステップ5 購買	・商品を購入したか否か ・なぜ商品が売れたかの検討	

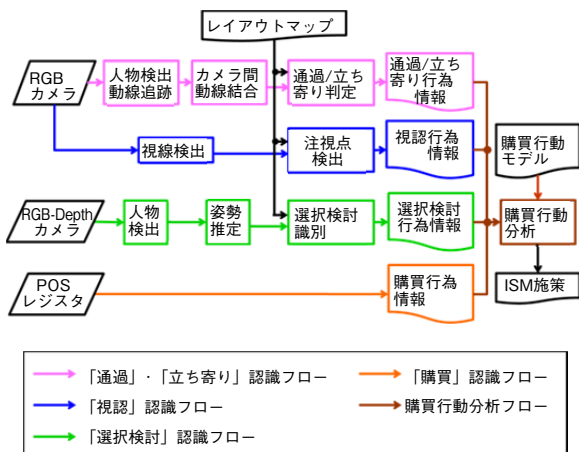
第1図 店舗における来店客購買行動のモデル化  
Fig. 1 Shopper behavior modeling

しかし、POSデータ分析は、商品の売れ行き、傾向を分析するためには有効であるが、商品に対する来店客の行動は読み取れないため、興味をもたれたか否か、選択検討されたか否かなどを分析することは困難である。また、POSレジスタの普及によりPOSデータは容易に取得できるが、それ以外の行動を捉えるための取得方法は普及しておらず、通過、立ち寄り、視認、選択検討に関する購買行動は、ISMにおいて重要であるにも関わらず、店内購買行動の分析は十分ではなかった。そこで筆者らは現在、店舗における来店客の購買以外の行動を取得し、新たなISM施策立案による売り上げ向上を小売企業、商品メーカーに提供することを目的に、購買結果以外の店舗内行動取得・分析システムの開発に取り組んでいる。

本システムにおける1つの特徴は、映像データを元に購買行動の自動検出を行い、さまざまな購買行動データを取得することにある。しかし複数人物が重なり合う映像が撮られる混雑した店内においては、従来の技術では誤認識が容易に発生してしまう課題があった。そこで、認識が難しい人物の姿勢推定のため、RGB-Depthカメラにより輝度情報と深度情報を取得し、画像中の輝度勾配と深度勾配を用いた機械学習により、人物検出を行う方式を導入した。また、RGBカメラを用いた動線認識における誤結合発生要因の評価を行った。

## 2. 店舗内行動取得・分析システムの概要

購買行動データを効率的に取得するためには、自動的に購買行動を認識する仕組みが必要である。来店客の購買行動を取得し、分析するための店舗内行動取得・分析システムの機能フローを第2図に示す。本システムでは、



第2図 店舗内行動取得・分析システムの機能フロー  
Fig. 2 Function flow in shopper behavior recognition system

第2図で示す機能により、店舗における来店客の「通過」・「立ち寄り」・「視認」・「選択検討」・「購買」という購買行動を認識する。

本章では、本システムにおける、これらの機能フローについて説明する。

### 2.1 「通過」・「立ち寄り」認識 (来店客動線認識)

通過・立ち寄り行為情報は、店舗内レイアウトが来店客にとってスムーズに店内を移動できるような構造になっているか、購買目的の商品に辿りつく間などのようなものに興味を示したかなど、レイアウトや商品と来店客行動との関連性の分析に有効な来店客の動線情報である。この動線情報の取得には、店舗内全域を対象とした来店客の位置を観測する必要があるため、店舗内天井に設置された全方位監視カメラの映像を用いる。動線情報の取得は、カメラで撮影したRGB画像から人物検出を行い、動きを追尾することにより行う。そして動線情報と店舗内のレイアウトマップとを照合することにより、来店客がどの棚の前を「通過」したか、あるいはどれくらいの時間「立ち寄り」したかに関するデータを取得する。

#### [1] 人物検出・動線追跡

人物検出は、背景差分と、人体の頭部から肩の稜(りょう)線にかけての輪郭形状を表現するAOM (Active  $\Omega$  Model) を用いた頭部検出[2]により行う。次に、検出した頭部のカメラ座標を店舗座標に変換し、頭部座標の移動ベクトルや頭部領域の輝度特性などを用いて人物の動線追跡を行う。

しかし、複数の来店客がすれ違う場合のカメラ映像に対して、異なる来店客の動線を誤結合してしまう課題があった。誤結合を低減させるための適切なカメラ撮影パラメータを決定するための取り組みについて、3.2節に記載する。

#### [2] カメラ間動線結合

店舗内天井に設置された全方位監視カメラ1台で店舗内全域をカバーできるケースはまれである。そこで、店舗内の来店客の動きを追い続けるため、複数の全方位監視カメラを天井に設置し、カメラ間で同一来店客の動線を結合する。カメラ間動線結合は、各カメラが生成した動線の移動ベクトルや座標情報の類似性をもとに行う。

### 2.2 「視認」認識

来店客が店舗内でどのように商品に接触したか、来店客が商品に対して関心をもっているかを理解するためには、「通過」・「立ち寄り」などの動線の他に、来店客の視認に関する情報が有効である。視認に関しては、自社の

みならず、他社の視線検出技術の導入を検討中である。

### 2.3 「選択検討」認識（姿勢推定）

来店客が棚の前で「立ち寄り」を行った際に、具体的にどの商品に興味を示しているか、またどのような観点から購入を迷っているかなどの購買意図を理解するためには、視線の他、姿勢（動作）に関する情報が有効である。姿勢の認識は、まずカメラ映像から人物検出を行い、次に腕や胴体など部位ごとの位置と姿勢を検出し、各部位の位置と姿勢の組み合わせから姿勢を推定することにより行う。そして、商品位置と推定した来店客姿勢との位置・姿勢関係から、「選択検討」行為を識別する。

しかし、姿勢を推定するのに要する頭部以外の部位は、部位同士によるオクルージョン（隠れ）が発生しやすいため、カメラで撮影したRGB画像からでは認識精度が不十分だった。そこで本システムでは、各部位を計測しやすい、商品棚の上に取り付けたRGB-Depthカメラで撮影した、来店客のRGB画像と深度情報を用いた。

#### 〔1〕人物検出

混み合った店舗内において、商品棚の上に設置したRGB-Depthカメラにより撮影した画像を用いて棚前の来店客の人物検出を行う場合、背景の映像が他の来店客の移動により移り変わるため、背景差分とAOMを用いた人物検出の精度が低下してしまうという問題があった。そこで、RGB-Depthカメラで撮影したRGB画像と深度情報を用いた機械学習により、背景差分とAOMでは頭部検出が困難なシーンに対しても、人物検出を可能とした。本取り組みの詳細は、3.1節に記載する。

#### 〔2〕姿勢推定

姿勢推定は、前項で述べた人物検出により検出した人体領域に対して、RGB-Depthカメラにより撮影したRGB画像と深度情報を取得し、あらかじめ学習した人体の腕や胴体などの部位ごとの特徴量をもとに、機械学習により部位ごとの位置と姿勢を検出し、各部位の位置と姿勢の組み合わせから、しゃがみや手伸ばしなどの購買姿勢を推定する。

#### 〔3〕選択検討識別

「選択検討」行為は、姿勢推定により購買に関する姿勢を推定した後、事前にシステムに登録された店舗内3Dレイアウトマップ上の棚や商品の位置と来店客の姿勢との位置関係をもとに識別する。選択検討識別により、“しゃがんで一番下の棚をのぞき込んだ”“商品を手に取ったが棚に戻した”など、ISM策定に有効な「選択検討」行為情報を取得することが可能となる。

### 2.4 「購買」認識（購買記録）

来店客の購買行動は最終的に、いかに「購買」につながるかが重要な要素である。そこで従来ISM施策と同様に、店舗内の購入記録であるPOSデータを取り込むことにより購買行為を認識する。

### 2.5 購買行動分析

システムで認識することにより取得した「通過」・「立ち寄り」・「視認」・「選択検討」に関する行為情報と、POSデータから取得した「購買」に関する行為情報に関する行為情報のタイムスタンプを照合することにより、店内に来店した来店客が最終的にどのような購買を行ったかを分析する。ただし、「通過」・「立ち寄り」・「視認」・「選択検討」・「購買」という購買行動を認識したとしても、認識した結果全てが意味のあるものになるとは限らない。ISM施策の立案を行うには、実際にどのような行動が購買につながるかなど、マーケティング的に意味のある情報がわからなければ、具体的に認識すべき来店客行動は定義できない。

今回、本システムの開発に際し、公益財団法人流通経済研究所とともに、来店客の店舗内行動のうち、どのような行動が購買行動として意味あるものなのか、行動分類およびモデリングを実施した。このモデルをベースとして作成した映像データをもとに機械学習を実施し、意味のある来店客行動を分析している。

### 2.6 レイアウト生成

通過・立ち寄り認識や視認認識を行うためには、レイアウトマップが必要であり、あらかじめレイアウトマップを作成し、システムに登録しておく必要がある。しかしレイアウトマップの作成には、商品棚や通路が占める領域情報やカメラの設置位置などの詳細パラメータを登録する必要があり、手動設定する場合、工数が小規模店舗1件ごとに約8人日かかるうえに、人的ミスなどが起こりやすいといった課題があった。そこで、RGB-Depthカメラを用いて撮影した店舗内3D映像をデータ化し取り込むことにより、簡単な編集作業によりレイアウトマップ生成を可能にした。

## 3. カメラ映像を用いた行動認識性能の改善

### 3.1 RGB-Depthカメラを用いた人物検出の改善

来店客の行動を認識することは、店舗内における特に売り上げに対する影響が大きい米飯棚などの混雑が想定されるエリアでの分析において重要になってくる。このような混雑したエリアの映像では、人物同士の重なりや

背景映像の影響などにより監視カメラなど、RGBカメラによる映像だけでは人物の認識率が低下してしまう課題があった。そこで、このような店舗内における重要なエリアにはRGB-Depthカメラを設置することを想定し、深度情報を利用した人の重なり合いや背景映像の影響などに強い3D人物認識の実用化に向けて取り組んできた内容を本節で述べる。

RGB画像において、人物検出を行う場合によく用いられる手法として、HOG (Histogram of Oriented Gradients) が挙げられる。これは、画像中の輝度勾配を特徴量として物体の輪郭を抽出し、その形状をもとに識別器を構成し、機械学習により人物を認識する方法である。

本システムでは、それに加え、HDD (Histogram of Depth Difference) というRGB-Depthカメラからの深度情報の勾配を用いた深度情報からの人物らしさを機械学習し、人物抽出をするという取り組みを行った。具体的には、深度画像における画素(x,y)それぞれに対して、式(1)のように勾配を計算し、この深度勾配から導き出される3780次元の特徴ベクトルからSVM (Support Vector Machine) により適切な識別器を構成することにより、人物であるか人物でないかの判定を行った[3][4]。

$$\Delta_x = \frac{D(x+1,y) - D(x-1,y)}{2} \dots\dots\dots (1)$$

$$\Delta_y = \frac{D(x,y+1) - D(x,y-1)}{2}$$

ここでD(x, y)は深度画像内の座標(x, y)における深度値である。



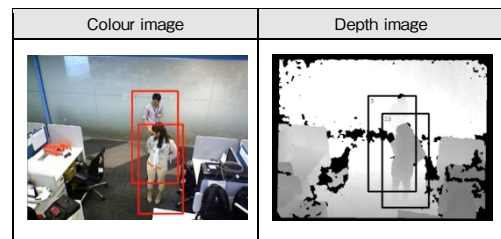
第3図 HOGとHDDを組み合わせた人物検出  
Fig. 3 Human detection from a combination of HOG and HDD

今回HOGとHDDという2種類の特徴量を組み合わせることにより、上記のような混雑した店内の領域においても実用に耐えうる人物認識が可能かどうかの検証を行った。

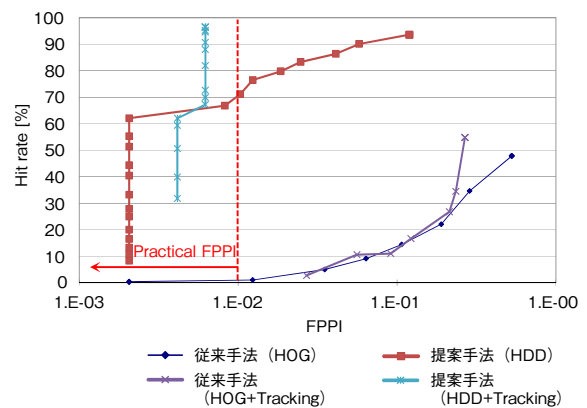
一般的に人物認識で用いられる評価用データセットは人物と同じ高さのカメラから水平方向に撮られたものが多く、さらに単調な姿勢のものを用いられることが多い。しかしながら、店舗内のターゲットとしているエリアにおいてカメラを設置する場所を考えると、棚上などの高い位置が現実的であり、複雑な人物の姿勢や背景の映

り込みが映像として用いられることが想定された。そこで、本取り組みでは天井から見下ろした角度から撮影した、複雑な姿勢や背景のデータセットを独自に用意し、それらに対する評価を行うことで検証を行った(第4図)。さらに、単純なHOGやHDDによる人物検出のみならず、独自の技術である画素ベースのLUT (Look-Up Table) トラッキング技術を導入することにより[5]、さらなる精度向上を図った。その結果を下記の第5図に示す。

ここでFPPI (False Positive Per Image) とは1枚の画像中で誤認識が発生する確率を示している。またHit Rateは認識すべきものを認識した確率(再現率)を示している。FPPIとHit Rateはトレードオフの関係にあり、認識のための識別器のしきい値を変化させることによって調整する。



第4図 評価に用いた実用データセットの例  
Fig. 4 Example from practical data set used for evaluation



第5図 実用データセット (485パターン) における評価結果  
Fig. 5 Evaluation result using practical data set (485 patterns)

誤認識が発生すると、人手による動線データの修正が必要となってくる。そこで、コストという観点で実用に耐えうるFPPI=1.E-02以下におけるHit Rateを見てみると、HOGのみを用いた場合の再現率が最大1%だったのに対し、HDDを用いた人物検出と画素ベースのLUTトラッキングを併用した結果は最大97%であり、100%に近い値が得られた。つまり、店舗内の混雑しやすい重要なエリアにおいて、人物行動認識には、深度情報の利用が有効

であると結論付けることができた。

一方で、今回の評価では、上記手法においても認識できないケースがあることもわかってきた。今回評価に用いた深度センサはそのデバイス特性により、遠方の対象の認識は難しく、また、外来光の影響を受けやすい。そのため、店舗内での設置間隔や照度に応じたセンサの選定が今後の課題の1つである。

さらに、店舗内の重要なエリアにおいては、前述した来店客行動モデリングで定義したことにより、従来では想定していなかった姿勢を認識する必要があることがわかった。例えば、来店客が商品を手に取りそれを上下に振り、音を確認するような動作を行うといった行動を行った場合、来店客は商品に対しかなり興味があるということが上記行動モデリングの取り組みでわかっている。

このような来店客の複雑な行動を認識するためには、身体の一部が胴体などによって隠れてしまうケースが多く、こういったパターンに該当するようなデータセットを用いて同様にHDDによる高精度姿勢認識技術を確立していく必要がある。

### 3.2 カメラ映像からの動線追跡の改善

ネットワークカメラの映像をサーバ側に記録する際の通信帯域や記録容量の関係から、記録フレームレートと画像圧縮率は、トレードオフの関係がある。そのため、動線追跡に適した記録フレームレートと画像圧縮率の関係に対する評価を行った。

評価対象は、店舗内の人物Aと人物Bがすれ違う際に、人物Bの動線を、人物Aの動線として誤って追跡してしまう誤結合の発生率（誤結合発生率）とし、店舗天井に設置した全方位カメラにより複数人物がすれ違う60シーンを記録した独自データセットを作成した。解像度は640×480 (VGA) に固定し、記録フレームレートを8 fpsと25 fps、画像圧縮率を標準画質モードと高画質モードとしてそれぞれ記録し、誤結合発生率を評価した。第1表に誤結合発生率の結果を示す。記録フレームレートを8fpsから25 fpsに変更することで、画像圧縮率によらず、誤結合発生率が顕著に向上した。この要因は、高フレームレートの場合、フレーム間の同一人物の距離が短くなるためと考えられる。一方、画像圧縮率に着目すると、記録フレームレートによらず、標準画質の方が、高画質よりも、誤結合発生率が低い。これは、画像圧縮により画像の高周波ノイズが抑制されることによる、頭部位置推定時の誤差減少、および、同一人物を示す輝度特性値の誤差減少によるものと考えられる。

第1表 動線追跡評価における誤検出発生率

Table 1 Error rate of the human tracking experiment

	低レート (8 fps)	高レート (25 fps)
標準画質	15 %	0 %
高画質	23 %	5 %

## 4. まとめ

本取り組みにより、姿勢認識には、RGB-Depthカメラで取得した色情報と距離情報をもとに、HOGとHDDの特徴量を組み合わせることにより、輝度情報のみを用いた手法では適合率99 %のときの再現率が1 %しか達成できなかったようなケースに対しても、97%での認識が可能となり、実店舗環境における人物検出において深度情報の利用が有効であると結論付けることができた。また、RGBカメラを用いた動線認識における誤結合発生要因を評価し、映像データを高画質で記録するよりも、高フレームレートで記録することが誤結合発生を低減させることを明らかにした。今後は店舗内環境に適切なRGB-Depthセンサの導入、人体の各部位をHOGとHDDを用いて個別に検出することによる姿勢推定技術開発、購買意図の推定精度向上の取り組みを行う予定である。

今回、本システムの開発に際し、来店客の店舗内行動のうち、どのような行動が購買行動として意味あるものなのかに対する行動分類、および行動モデリングにご協力いただいた公益財団法人流通経済研究所の山崎泰弘様、三坂昇司様には、多大なるご尽力をいただきました。心より感謝の意を表します。

## 参考文献

- [1] 神谷渉, "ショッパー・インサイトを捉えるための技法," ショッパー・マーケティング, 財団法人流通経済研究所, 日本経済新聞出版社, 東京都, 2011年, 第5章.
- [2] 橋本潔 他, "姿勢変動に伴う身体特徴変化の統計的モデリングによる遮蔽に頑健な人物追跡," ViEW2011, 11-25, pp.73-74, 2011.
- [3] Shengyin Wu et al., "An attempt to pedestrian detection in depth images," Intelligent Visual Surveillance, 2011 Third Chinese Conference, pp.97-100, Dec. 1-2, 2011.
- [4] Dalal, N et al., "Histograms of oriented gradients for human detection," Computer Vision and Pattern Recognition, 2005. IEEE Computer Society Conference, pp. 886-893, Jun. 20-25, 2005.
- [5] Cher-Keng Heng et al., "Shrink boost for selecting multi-LBP histogram features in object detection," Computer Vision and Pattern Recognition, 2012 IEEE Conference, pp.3250-3257, Jun.

16-21, 2012.

## 執筆者紹介



大坪 紹二 Shohji Ohtsubo  
AVCネットワークス社  
イノベーションセンター  
Innovation Center, AVC Networks Company



守山 隆昭 Takaaki Moriyama  
AVCネットワークス社  
イノベーションセンター  
Innovation Center, AVC Networks Company



石原 健 Takeshi Ishihara  
AVCネットワークス社  
イノベーションセンター  
Innovation Center, AVC Networks Company



苅部 朋幸 Tomoyuki Karibe  
MICE推進プロジェクト  
MICE Promotion Project



スギリ プラナタ Sugiri Pranata  
パナソニックR&Dセンター シンガポール  
Panasonic R&D Center Singapore



イエン シュイ Yan Xu  
パナソニックR&Dセンター シンガポール  
Panasonic R&D Center Singapore