

# 人の商品購買行動の多視点映像分析

知能情報学専攻 松山研究室 修士1回生 若井 祐介

平成17年3月29日

## 1 はじめに

これまで人とインタラクションを行うインテリジェントシステムが多く研究されてきたが、キーボードなど入出力デバイスを介さずに、コンピュータ側から人の状態を推定し、人に対して適切な時間・内容で何かを提案することが出来るシステムはまだ実現できていない。そこで私は、多数のカメラを用いて人を観察し、そこから計測可能な姿勢や動作の情報を用いて、人の内的な状態や行動パターンを推測する手法を研究する。

今回の研究の対象に選んだシナリオは、たとえば店舗などで人が商品を選び、決定する状況において、人がどの商品にどの程度強い興味を持ったかをコンピュータが推定するというものである。もし、明示的に商品に対する質問を行わない間に、その人の興味ある商品分野や、指向性を推定して、個人的なディスプレイや音声を通じて的確なアドバイスを与えることができれば、大規模小売店に設置された監視カメラを有効活用して販売効率を飛躍的に高めることができるであろう。

これを実現するために、24台の多視点カメラ映像を用いて、商品を選んでいる人の映像を撮影し、さらに、この映像を三次元化して人の体・頭・手などの位置や姿勢、顔の向きなど、人の行動を計測する。これらの計測値を元に、どの商品に対してどの程度の興味を保持しているかを定量的化する手法を提案する。たとえば、「立ち止まる」「目をとめる」「近寄る」「注視する」「手に触れる」という興味のレベルに応じた行動に分解し、それぞれの行動の意味と動作特徴を解析する。

現在、実験室環境で好きなものを選ぶ行動を撮影し、実現可能性検証実験を行ったので、それを

紹介する。

### 1.1 研究背景

これまで人とインタラクションを行うインテリジェントシステムが数多く研究されてきたが、それらの大半は以下の2つに集約される。

- 人からの指令を受けて新たな行動を始めるパターン。
- あらかじめ決められた条件が満たされると新たな行動を始めるパターン

そのため人間が命令したことを実行して応答を返すといった従来の情報システムから、今後は状況にあった反応を自律的・自発的に行う情報システムの実現を目指す必要がある。そして、実際にキーボードなど入出力デバイスを介さずに、コンピュータ側から人の状態を推定し、人に対して適切な時間・内容で何かを提案することが出来るシステムはまだ実現できていない。そこで私は、多数のカメラを用いて人を観察し、そこから計測可能な姿勢や動作の情報を用いて、人の内的な状態や行動パターンを推測する手法を研究する。

### 1.2 先行研究

#### 1.2.1 姿勢推定

村上らは3次元ボクセルデータに基づく人体の姿勢推定と題して人体各部の3次元位置を推定した。あらかじめ各部分に位置の情報を与えた人体モデルを設定し、3次元データとモデルをフィッ

	根ノード	関節の可動
村上	腰の位置	制限無し
田中	頭の位置	制限無し

表 1: 根ノードの決定と関節の可動範囲

ティングさせることにより，モデルの持つ位置情報から，データに含まれる各部分の座標値を推定するものである．

人の身体は胴体，頭部，腕，足などの各パーツが，関節によって繋がることで構成されていると考えることができる．そこで，人体モデルもいくつかのパーツが可動できる関節で繋がることで構成される多関節モデルを考えた．モデルの各パーツは比較的単純な形状の剛体として近似し，全体を頭，胸，腰，左右の上腕，下腕，上腿，下腿の11個の剛体パーツを想定した．各関節はいずれかのパーツの1部に組み込み，隣接するパーツ位置には従属関係があると考えたことで，あるパーツを根ノードとして，その他全てのパーツをノードとして含む木構造で表した．姿勢推定は根ノードとなるパーツから始め，近いパーツから順番にフィッティングさせることで逐次的に行われている．

### 1.2.2 熱中度の定義

また先行研究として田中らの研究では，個人の動作や姿勢から物事に対する関心の程度を判断する指標として熱中度を定義した．人が完全にリラックスした状態を基準として，意図的に身構えている際の熱中度を3次元データを用いて定量的な計測を行っている．これによりあるタスクを実行するための姿勢や動きなど「構え」を計測できたことになる．しかし，これでは姿勢や動きに表われない内面的な集中や緊張を計るのは原理的に不可能であるまた，人間の行動形態というのは本来非常に複雑なものであり，簡略化し過ぎている恐れもある．

両者の人体モデルの特徴を以下，表に示す．

## 2 研究概要

### 2.1 研究目的

店舗などで人が商品を選び，決定する状況において，人がどの商品にどの程度強い興味を持ったかをコンピュータが推定することを目指す．明示的に商品に対する質問を行わない間に，その人の興味ある商品分野や，指向性を推定して，個人的なディスプレイや音声を通じて的確なアドバイスを与えることができると考えられる．

例えばウェアラブルな機器を装着し視線を検知することで人の商品選択行動について推定することは可能であろう．しかし，商品を購入する際にそのような機器を装着することは実際にはありえないことであり，その煩わしさも半端なものではないだろう．そこで実社会において店舗に設置された監視カメラを用いて映像から人の商品購買行動を分析することを考える．映像から読み取れる情報においても視線は重要であり，視線インターフェイスを活用することも有効ではある．しかし，購買行動とは身体や顔の向き，手や足の動きなど全体で表現しているため，視線だけでなく身体全体の動きを考える必要がある．

### 2.2 予備検証

#### 2.2.1 方法

身体全体の動きを実現するために，宮崎ビルにおいて24台の多視点カメラ映像を用いて，いくつかの物を選んでる人の映像を撮影した．図のような机をセッティングし，その上にテスター，アイボ，人形，本を置いた．そしてその物体について知らない人に対して自分が説明しやすい物体を選んでもらった．このような質問内容にしたのは作為的な現象を撮影するのは意味がないと考え，選んでいる自然な現象を撮影するために心理学的な面を考慮したためである．

## 2.2.2 結果

人が商品を選択する行動において興味を持っていると感じられる2つの特徴的な動きが観測された。

- 手を使った特徴的な動作

人がある物体に対して興味を示すと、物体に手で触れることで手触りや感触、物体を手で持つことで機能性の確認という動作を行う。ここで、その物体に対してさらに興味を示した場合、手を回転させてその物体の裏側を見たり、持ち替えることでその物体の操作性を確認したりする姿が観測された。

- 時が止まる

人が手に触れて、操作をすると、その物体に対して本当に興味がある場合、その物体独特の特徴に夢中になる瞬間が観測された。目が奪われ、時が止まる瞬間である。

## 2.2.3 考察

今回の検証において、テスター、アイボ、人形、本の選択肢を用意した。しかし、本の場合には本を開いてページをめくるといった必要な動作や、アイボの場合にはボタンを押したり、アイボ自身が動くという固有の動きがある。そのような中で、選択の特徴を見極めようとした場合、選択物体特有の影響が強く出てしまうことが考えられる。つまり今回の検証において選択物体のカテゴリーを意識していなかったため物体を選択する特徴的な行動が、選択する意図ではなく、選択物体のカテゴリーの違いから生じている可能性がある。次の設定としてカテゴリーを統一する必要がある。

## 3 商品購買行動の多視点映像分析

人が商品を買う際には、欲しい商品もしくは必要な商品郡を探しだし、似かよった商品の中から選択をするという流れが存在すると考える。そこ

で商品購買の一連の流れを次のようにレベル別に分類することでその興味の度を測る。

- レベル1「立ち止まる」

興味のある商品郡や、欲しい(必要)商品郡の特定

- レベル2「目をとめる」

デザインや価格など目の引く情報の抽出  
-商品を絞りこむ

- レベル3「近寄る」

絞りこんだ商品に対する積極性の度合

- レベル4「手に触れる、操作する」

実際の機能性の確認、手触りなど

- レベル5「注視する」

興味深さ(夢中になる瞬間)の抽出

多視点映像を三次元化し人体モデルとのフィッティングを行うことで、人の体・頭・手などの位置を計測し、姿勢、顔の向きを加える。これらの計測値を元に、人が商品購買を行う中でどのレベルに達しているかを認識する。そして、どの商品に対してどの程度の興味を保持しているかをシステムが把握する。今回の検証は選択に焦点をあてたレベル4、5に値する。今後レベル1、2、3に拡張したデータを取る必要がある。