

# 多数センサによる観測データの相関を手がかりとする人間行動判別

服部 傑<sup>†</sup> 小林 智行<sup>†</sup> 亀田 能成<sup>††</sup> 大田 友一<sup>††</sup>

<sup>††</sup> 筑波大学大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1  
E-mail: <sup>†</sup>{hattori,kobayashi}@image.esys.tsukuba.ac.jp, <sup>††</sup>{kameda,ohta}@iit.tsukuba.ac.jp

あらまし 高付加価値型住環境システムを実現するためには、人間の行動観測を安定して行う必要がある。我々は、その一環として多様なセンサを用いた人間の行動認識の実現に向けて研究を行っている。本稿では、カメラやマイクロフォンといった複数種類のマルチメディアパッシブセンサが非均一、非校正につけられた環境下でそれらを複合利用して大量の情報を取得し、そこから得られる情報を組み合わせることで人間行動判別に役立てる枠組みについて述べる。  
キーワード センサフュージョン, 行動認識, クラスタリング, マッシブセンシング

## Human Action Classification by Utilizing Correlation of Observation Data from Massive Sensors

Takashi HATTORI<sup>†</sup>, Tomoyuki KOBAYASHI<sup>†</sup>, Yoshinari KAMEDA<sup>††</sup>, and Yuichi OHTA<sup>††</sup>

<sup>††</sup> Graduate School of Systems and Information Engineering, University of Tsukuba .  
Tenoudai 1-1-1, Tsukuba, Ibaraki, 305-8573 Japan  
E-mail: <sup>†</sup>{hattori,kobayashi}@image.esys.tsukuba.ac.jp, <sup>††</sup>{kameda,ohta}@iit.tsukuba.ac.jp

**Abstract** A stable and robust human action recognition is necessary for realization of smart living environment. Our purpose is to develop a sensing system for human action recognition by utilizing a number of multimedia passive sensors. In this paper, we describe a method to utilize a large number of data from multimedia sensors such as cameras and microphones that are set un-uniformly and uncalibrated.

**Key words** Sensor Fusion, Human Action Recognition, Clustering, Massive Sensing

### 1. はじめに

情報社会における見守り技術として、環境中に様々なセンサを埋め込み、その観測値をもとに人間行動を分類したり認識したりする研究が進められている。

例えば、アウェアホームやインテリジェントルーム [1]、ユビキタスホーム [2]、センシングルーム [3]、インテリジェントクラスルーム [4] など国内外で様々なプロジェクトが進められている。

このような多数のセンサを同時に利用して人間行動の分類や認識を行う際に問題となるのが、センサのプロパティ管理である。一般に多数のセンサを同時に用いる場合には、それらを等間隔に並べたり、センサの位置を正確に測定したりしておくことが必要とされる。均一に配置され、校正されたセンサ群はデータ処理としては扱いやすいが、現実問題としてこのようなセンサ群を実際の日常環境中に用意するのは容易なことではない。また、複数の種類のセンサを用いる場合、従来はそれぞれで高次レベルまで特徴抽出を行い、それから結果を統合するのが一般的であった。

これに対し、本研究では非均一、非校正、非均質なセンサ群に対して統一的な特徴抽出法と、それに基づく人間行動の分類、認識法について提案する。このような非均一、非校正、非均質なマルチメディアパッシブセンサ群に基づくセンシングを、我々はマッシブセンシングと呼ぶ。

マッシブセンシングにおいては、各センサは固定されているものとし、環境も認識対象以外は静止しているものとするが、センサの配置は任意であり、それらの空間的配置も不明であるとする。また、センサの種類もカメラやマイクなど数種類が用意されているものとする。一方で、観測することになる人間行動がどのようなものであるかは事前に予想はしない。すなわち、どのような人間行動であっても対応可能であるとする。

本アプローチにおいてまず問題になるのは、受動的センサから得られる情報からの有用な部分情報の抽出である。カメラやマイクロフォンは受動的センサとして多くの情報を獲得可能であるが、日常生活環境ではセンサ観測量にノイズが混入しやすい。これは環境中の背景とみなされる構成要素の中に、定期的なノイズ発生源や、ゆるやかな変動要因が存在するためである。そこで我々は、背景の動的特性に注目し、そのような環境でも

非定常部分である有用な部分情報を抽出する方法を採用する。

もう一つの問題は、センサフュージョンの方法である。我々は特定行動の認識を想定しているわけではないため、その行動が映像的に観測できるか、ないしは音響的に観測できるのかを予見することはできない。そこで、本稿では、映像特徴量と音響特徴量を統一的に扱うための情報の粒度の均質化について述べ、観測された時系列情報を判別するための距離の定義を行う。また、実際に取得したデータにおいて、定義した距離を用いてのクラスタリングを行い、その結果について考察を行う。

## 2. 有意部分の抽出とデータ粒度の均質化

カメラやマイクロフォンのようなマルチメディアパッシブセンサは、大量のデータを出力し続けるため、そのすべての出力について記録をしたり高次処理をすることは現実的ではない。そこで、意味があると見なせるデータ系列のみを抽出する。

センサからみた日常環境は静止した固定的なものとは限らず、定常的なノイズや緩やかな変動要因を含んでいる。そこで本手法では、映像・音声データ各々に対し過去一定期間のデータを保持し、動的背景モデルを更新するとともに、その背景モデルに従わない部分を有意部分として抽出する。

### 2.1 M 推定を用いた背景画像モデルの推定と前景画像抽出

我々の想定している日常環境でのセンシングでは、室内であつても屋外の日照状況の変化や照明器具の変化の影響を受けたり、家具や物品等の移動にも影響を受けたりするので、何らかの方法で背景画像を更新していく必要がある。そこで、本研究では口バスト統計に基づく M 推定 [5] を用いて動的背景画像モデルを構築し、同時に前景の抽出を行う。この手法では、照明状況の変化・対象環境の変化に対して適応的に背景推定モデルを更新していくことができる。

### 2.2 定常ノイズモデルを用いた前景音像抽出

音像についても、映像データに対する処理と同様に、定常音に対応した背景音モデルを定義し、前景を取り出す方法を用いる。背景音スペクトルを利用し、前景音抽出を行うという手法は、Boll ら [6] によって提案され、その後様々な改良がなされている。本手法では、Spectral Subtraction 法 [7] に M 推定を用いて背景音スペクトルを推定し、それを用いた前景音抽出を行う。

## 3. 有意データ節抽出と特徴量

異種のセンサを統一的に利用する場合、各センサの出力する特徴量の時間方向の粒度と次元数の違いは、しばしば認識器の性能に影響を与える。一般的なセンサの出力には、時間方向の分解能と、一回のセンシングで得られる観測値自体の分解能がある。

一般のマルチメディアセンサは、もともと人間の知覚に合わせて設計されているため、時間方向については可聴域の音の観測は数十 KHz 程度で量子化され、映像は数十 Hz で量子化される。一方、我々は人間行動の観測に必要な時間的分解能についても考慮し、映像・音声の特徴量とも時間方向の粒度を 100msec 刻みに調整する。

この 100msec 幅の各時間データセグメントにおける特徴量次元数については、その数が著しく異なるとセンサフュージョンが難しくなると予想されるので、それぞれの次元数をおおよそ同じオーダーに収めるものとする。具体的には、映像は各カメラの撮影領域を 25 分割し、RGB の 3 成分の計 75 次元を特徴量とした。また、音声は 10 ~ 13 KHz の範囲から、27 個の代表周波数を定義し、27 次元の特徴量とした。

我々の研究では、連続するいくつかの時間データセグメントをまとめた、一定の長さを持つ時系列データを単位としてその後の判別処理に用いる。具体的には、入力となる時間データセグメント群から前景となるデータを含むものを取り出し、定常状態から外れていると判断される区間を含む一定時間長を有意データ節として抽出する。

## 4. 有意データ節間の相違度

3 節までの手法によって抽出された各有意データ節は、多次元の時系列データである。これらをクラスタリングすることによって行動の判別をするためには、時系列データ間の相違度を定義する必要がある。

時系列データをマッチングする場合には、データ間の位相のずれが問題となる。また、時系列データ間のデータの伸縮も同様に問題となる。位相のずれやデータの伸縮を吸収するような適切な相違度を用意することで、それに基づいたクラスタリングとイベント判別が可能になる。本研究ではこれを吸収することができる相違度として Running Spectrum [8] を導入した。

Running Spectrum について以下に示す。

有意データ節中のある特徴量次元  $n$  の時系列データ  $a_n$  を離散フーリエ変換したものを  $A_n$  とし、その各周波数成分を  $\lambda$ 、周波数成分数を  $L$  とする。そして、次元数  $N \times$  周波数成分数  $L$  の空間でのユークリッド距離を有意データ節間の距離として定義する。具体的には、有意データ節  $a$  の離散フーリエ変換  $A_{RS}$  を式 (1) および (2) によって求める。これらは、Running Spectrum と呼ばれる。そして、式 (3) によって相違度を定義する。

$$A_{RS} = [A_1, A_2, \dots, A_n, \dots, A_N] \quad (1)$$

$$A_n = Re(DFT(a_n)) \quad (2)$$

$$d_{RS}(A, B) = \left( \sum_{n=1}^N \sum_{\lambda=1}^L (A_{n\lambda} - B_{n\lambda})^2 \right)^{\frac{1}{2}} \quad (3)$$

## 5. クラスタリング手法

### 5.1 指 針

本研究で得られた有意データ節をクラスタリングによって分類する際には、以下の点について留意する必要があると考えられる。

- クラスタリングの手法に先験知識を持ち込まない
- クラスタリングの結果が視覚的に分析しやすい

以上二点を考慮し、本研究では階層型クラスタリングを用いて階層状にクラスを構成する。階層型クラスタリングにおいては先験知識を用いないクラスタリングが可能であり、また、クラ

スタリング結果としてデータの階層構造が得られるため、結果の視覚的な分析が容易であると考えられる。

本研究では、階層型クラスタリングのための、クラスタ間相違度  $D$  の定義について Ward 法 [9] を用いるものとする。

Ward 法について以下に示す。

$$D(C_1, C_2) = E(C_1 \cup C_2) - E(C_1) - E(C_2) \quad (4)$$

$$E(C_i) = \sum_{X_i \in C_i} d(X_i, \bar{X}) \quad (5)$$

ここで、各変数は以下のような意味を表す。

- $C_i$ : あるクラスタ
- $\bar{X}_i$ : クラスタ  $C_i$  の平均
- $X_j$ : ある一つのデータ (ベクトル)
- $d(X_i, X_j)$ : 二つの有意データ節間の相違度 (ここでは RunningSpectrum を利用)

連続している有意データ節は、特徴空間では近辺にまとまって存在すると考えられる。Ward 法を用いた逐次的クラスタ併合では、新しいクラスタの中心から各要素の二乗の総和が小さくなるような 2 つのクラスタからクラスタの併合が進んでいく [9]。このため、単リンク法などより一連の動作シーケンスが同じクラスタに結合される可能性は低くなり、一連の動作の部分部分が独立した状態にクラスタリングが行われると考えられる。

## 5.2 階層型クラスタリング

階層型クラスタリングの具体的な処理手順は以下に示すとおりである。

- (1) 各データセグメント一つ一つを初期クラスタとする
- (2) クラスタ間の相違度をすべて求め、最小値を探索する
- (3) 該当クラスタのペアを統合し新しいクラスタを作成
- (4) クラスタ数が 2 以上であれば 2 に戻る

これによって、各ノードをクラスタ、各リーフをデータセグメントとする二分木の形でクラスタリング結果が得られる。このクラスタリング結果を、クラスタ間相違度を横軸に取り、クラスタの位置が子クラスタ間の相違度を表す位置となるようなツリーで表示したものをデンドログラムと呼ぶ。

## 6. センサフュージョンとその評価手法

本研究では、各センサから得られる特徴量を等価に扱い、人間行動判別を行う枠組みの構築を目標としている。

ここで、RS を用いた相違度を用いる場合、各センサの RS 特徴量は多次元のベクトルデータとして定義される。

そのため、単一センサから多センサへの拡張は単に個々のセグメントの特徴ベクトルの次元が拡張されるだけ、という扱いで考えることができる。

先に挙げた理由から、カメラ (映像) データの特徴量とマイクロフォン (音) データの特徴量とで、時間方向の粒度をそろえ、分解能も概ね同じオーダーに収めているので、カメラとマイクの統合も、単純に特徴ベクトルの次元を拡張する形で実現する。

このとき、複数センサから得られる有意データ節を統合して

行ったクラスタリングに於いて、どのようにセンサフュージョンがなされたのかを評価する指標が必要となる。

そこで、クラスタリングにおける各センサの重要度と、各センサ同士がどの程度類似していたのか、の二項目に着目してセンサフュージョンの評価を行うこととする。

本研究では、矢入ら [10] の手法にならい、情報量基準に基づいた評価基準を導入する。

クラスタリングにおけるセンサの重要度については、情報量基準の観点からセンサ特徴量とクラス集合間の相互情報量で定義する。クラス集合とは、クラスタリングの結果として得られるクラスタ群のことを示す。これは、言い換えれば、センサ特徴量  $S_j$  が明らかになったとき、それがクラス集合  $C$  に対する情報をどの程度与えるかという観点に基づいてセンサの重要度を決定することを意味する。具体的には、式 (6) の形で表現することができる。

$$\begin{aligned} Impc(S_j) &= I(S_j; C) = H(C) - H(C | S_j) \\ &= \sum_i \sum_l P(C_i) P(S_{jl} | C_i) \log \frac{P(S_{jl} | C_i)}{P(S_{jl})} \quad (6) \end{aligned}$$

ここで、 $C_i$  はある  $i$  番目のクラスタを、 $S_j$  はある  $j$  番目のセンサを表し、 $S_{jl}$  は同センサの出力値  $l$  を表す。また、 $I(S_j; C)$  はセンサ入力とクラス集合間の相互情報量を表す。 $H(C)$  は、クラス集合  $C$  のエントロピーを表し、 $H(C | S_j)$  はセンサ  $S_j$  の状態集合が得られたときの条件付きエントロピーを表す。

一方、センサ同士の類似度については、式 (7) のようにセンサ間の相互情報量によって定義する。

$$\begin{aligned} Sim(S_j, S_k) &= I(S_j; S_k) \\ &= \sum_{l,m} P(S_{jl}, S_{km}) \log \frac{P(S_{jl}, S_{km})}{P(S_{jl})P(S_{km})} \quad (7) \end{aligned}$$

## 7. システム構成

本研究で用いるシステムの構成について以下に述べる。

実装したシステムは、ネットワーク接続されたセンサ群と処理用 PC 群から構成されている。システムの詳細については文献 [11] を参照されたい。

実験環境として、本学実験室の半分のスペースおよびその周辺廊下を対象としている。各センサは図 1 のように配置されている。

- ネットワークカメラ群: 実験室天井と周辺廊下天井に計 35 台を設置する。これは、実験環境床面を概ね死角無く撮影できる数である。

- マイクロフォン群: マイクロフォンは 8 台を実験室天井に設置する。マイクロフォンはカメラと異なり、オクルージョンが発生しても観測が可能であるため、実験室の広さに対して十分な数であると考えられる。

- 画像処理用 PC 群: ネットワークカメラ群に対して画像取得・処理をする。7 台の PC を用いて、各 PC で 5 台の AXIS に接続した。

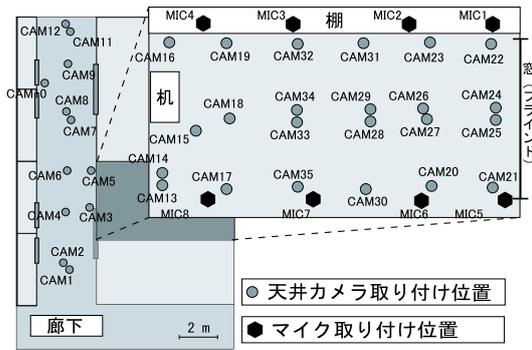


図 1 センサ配置平面図

なお、今回の実験では、データ取得以外の処理はすべてオフラインで行っている。

## 8. 有意データ節抽出実験

### 8.1 実験環境及び入力データ

本システムの設置環境、特に実験室では、日常的に以下のような作業（イベント）が発生する。

- 棚に入っている物品の持ち出しと返却
- 物品の探索
- プリンタでの印刷や印刷物の回収
- 撮影実験・PCの前での作業

本環境中では、これらの作業の発生頻度は日に数回程度である。本来は設置環境から長期的にデータの収集を行い、それらに対して処理を適用すべきであるが、今回は検証のために意図的に数分おきに上記のような作業を反復して行い、それに対して観測を行った。

実験データ取得中は、約 3 時間にわたって下記の作業を 5 回ずつ反復して行った。

- 機材棚からのソフトウェアの持ち出し、および返却
- 機材棚からの機材の持ち出し、および返却
- 機材棚の探索（物色）
- 印刷・印刷物の回収

それぞれの試行において、自然な日常行動に近い入力データを得るために、行動経路や振る舞いなどは統一しなかった。作業を行う人物は一人であったが、上着を二種類用意し適宜取り替えた。また、実験環境に対して特に出入りの禁止をしなかったため、意図的に行ったもの以外の日常行動も入力として観測されている。

本実験では 2 時間 48 分 13 秒分のデータを取得しており、秒に直すと 10093 秒である。本実験では、有意データ節の時間長  $T_{ds}$  を 1 秒と定めているため、有意データ節の候補数も同数あることになる。このとき、保存されたデータは JPEG 圧縮された画像データがカメラ 35 台分で計 86 GB、zip 圧縮された音声データがマイクロフォン 8 本分で 2.8 GB となった。

抽出された有意データ節数を図 2 にまとめる。横軸はカメラの番号を表し、縦軸は抽出された有意データ節の数を表している。

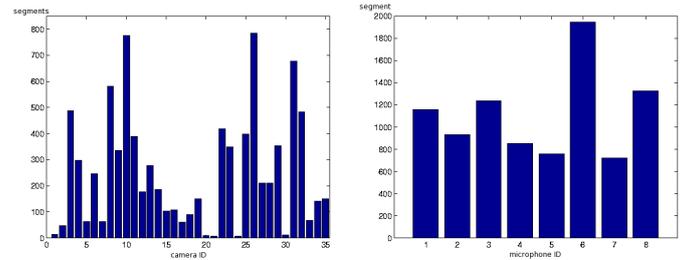


図 2 有意データ節数 (左：カメラ、右：マイクロフォン)

### 8.2 カメラでの有意データ節抽出数についての考察

カメラの設置位置は大きく廊下と実験室内に分かれるため、有意データ節抽出数についても設置領域ごとに考察する。

実験室内に設置されたカメラのうち、CAM20, 21, 24, 30 の有意データ節数が他と比較して極端に少ないのは、これらのカメラが実験室の奥の方を撮影しており、実験中に当カメラの撮影領域に人間が立ち入ることが少なかったためである。

実験室のカメラの中でも有意データ節を多く抽出している CAM26, 31 について考える。これらのカメラは物品持ち出しの対象となっている棚を主に撮影している。物品持ち出し動作では、棚の前に動作人物が停留するために他のカメラと比較して抽出数が多くなったと考えられる。

次に、比較的データ数の少なかった CAM15 ~ 18 と CAM33 ~ 35 に注目する。これらのカメラの撮影領域には、実験時に大量の物品が置いてあり通行が難しくなっていた領域を視野に含んでいるため、相対的に有意データ節の数が少なくなった。

その他のカメラについては、それぞれ戸棚に関連する動作で主に有意データ節を出力しているカメラ CAM22, 23, 32 と、物品が一時的に置かれることがあった机に関連する有意データ節を出力しているカメラ CAM25, 27 ~ 29 に分けられる。

廊下に設置されたカメラのうち、特に出力数の多い CAM10 について注目する。原因として、このカメラが本研究室入り口を撮影しており、研究室へ出入りする人間に加えて、撮影範囲となっている研究室の出入り口付近に停留していた人物が前景として抽出されたことが考えられる。

次に、廊下のカメラで出力数の少なかった CAM1, 2, 5, 7 について考える。CAM1, 2 については、実験時に繰り返し行った動作の行動範囲外のカメラとなっている。CAM1 の出力数が特に少なくなっているのは、このカメラが廊下の端を撮影しており、歩行者が撮影領域に立ち入ることが少なかったためと考えられる。CAM5, 7 も CAM1 と同様に廊下の端の方を撮影しているために、実験動作を撮影可能な廊下カメラの中でも出力数が少なくなったと考えられる。

その他のカメラについては、実験で繰り返し通過した動作と通常の歩行者が検出された結果、300 ~ 400 前後の有意データ節が得られた。

### 8.3 マイクロフォンでの有意データ節の抽出数についての考察

対象実験環境では、音に関連するノイズ源としてエアコンの稼働音や PC や各種機器の稼働音が存在する。このノイズ音

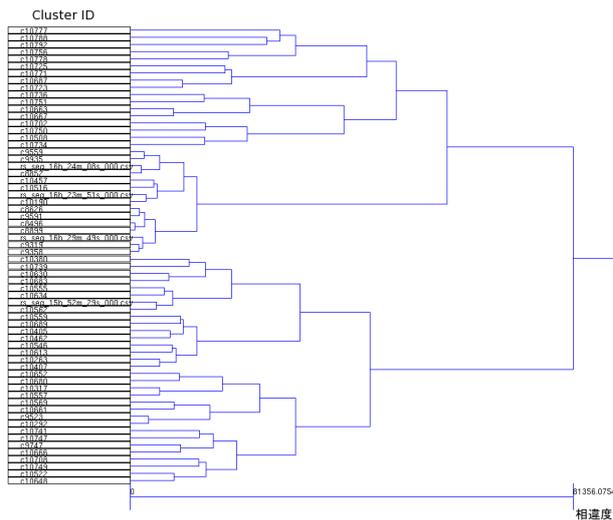


図3 64個のクラスタまでを示したデンドログラム

表1 クラスタごとの有意データ節数

NSC	1~20	21~40	41~60	61~80	81~100	101~120	...	3600
	46	11	2	1	0	1	0	1

は、普通に会話する程度の音と大差ない程度の大きさである。逆に、作業や入退室に伴う扉や柵の開閉音はそれらと比べて倍以上の大きさの入力として観測されている。

各マイクロフォンでの有意データ節数はおおむね 1000 前後に収まっている。

MIC6 の有意データ節数が他と比べて大きな値となっている。しかし、入力音波形の傾向が他のマイクロフォンと著しく異なっているわけではないので、閾値の設定値が大きく影響していると考えられる。適応的に閾値を決定するなどの対処が、今後の課題として挙げられる。

### 9. 全センサを用いた有意データ節のクラスタリング実験

本節では、4 節と 5 節で述べた相速度とクラスタリング手法を用いて、8 節の実験で得られた有意データ節集合に対してクラスタリングを行う。

クラスタリングは、全センサから得られた有意データ節の特徴量を統合したうえで実行する。

また、クラスタリング結果について 6 節で述べたように情報量基準に基づいてセンサの重要度とセンサ間の類似度を求め、それについての考察を行う。

なお、センサフュージョンする際に CAM1~35 を sensor1~35 とし、マイク 1~8 を sensor36~43 として扱った。

図 3 にクラスタ数が 64 個となったからのデンドログラムを、表 1 にそのときのクラスタ内有意データ節数の分布を示す。NSC はクラスタあたりの有意データ節数を表す。

まず、約 5400 個の有意データ節のうち、約 3600 個が一つのクラスタに集中していることが分かる。このクラスタに分類さ

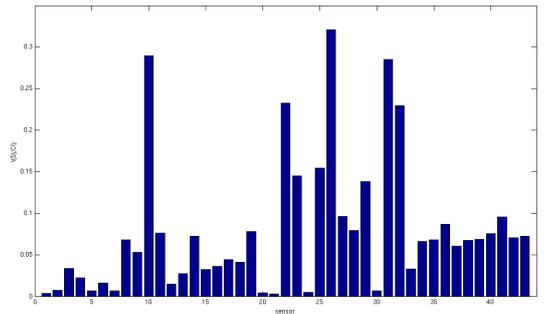


図4 全センサのクラスタリングにおける各センサの寄与率

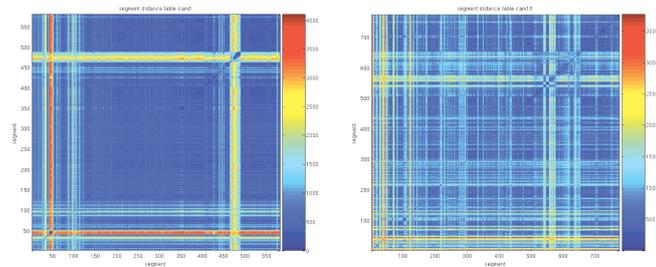


図5 単一センサにおける有意データ節間相互相速度 (左; sensor8, 右; sensor10)

れたセグメントを調べると、大半がマイクロフォンのみについて有意データがある有意データ節と、廊下のカメラのみについて有意データがある有意データ節であった。

ここで、マイクロフォンと廊下のカメラについて、RS 特徴量の相速度を調べたところ、その他のセンサに比べて相速度が全体的に低い傾向があった。このことから、センサの観測可能な事象のバリエーションが少ないものに関連する有意データ節についてはクラスタの結合が階層型クラスタリングの早い段階で進むと言える。

このような問題に対しては、クラスタ結合時に対象のクラスタの分散値の変化に注目して処理を行うといった対処法が考えられる。

また、比較的規模の小さいクラスタ内では、ある一連の行動の一部分という形で有意データ節がまとまっていることが確認できた。

しかし、運動が遅い行動の一部分であるような有意データ節では、その運動の方向に依らず同じクラスタとして結合しているものが見られた。これは、本実験で設定した有意データ節の時間長が 1 秒であり、その間に動きの少なかったデータでは、動作よりも前景成分の現れた位置の影響が大きくなってしまったと考えられる。

次に、図 4 に示す全センサを用いたクラスタリングにおける各センサの寄与率について考察する。

グラフ全体の傾向としては、センサごとの有意データ節抽出数を図示した図 2 と類似した形となっていることが分かる。これについてはセンサの出力する有意データ節の特徴量に基づいてクラスタリングを行っているため、センサの出力した有意

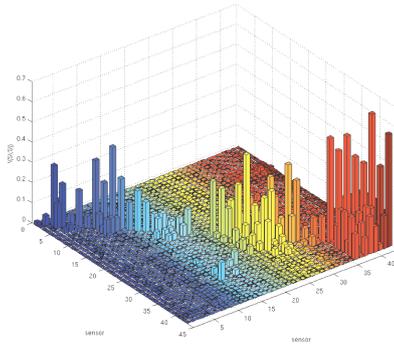


図 6 センサ反応の生起率に基づくセンサ間相互情報量

データ節の数とクラスタリング結果への寄与率に相関がみられるのは自然だと考えられる。

その一方で、局所的には有意データ節の出力数とクラスタリングへの寄与率で傾向の異なるセンサも存在している。これらについて、有意データ節の出力数に対してクラスタリングへの寄与率が相対的に減少したものと増大したものに分けて考える。

まず、前者について考えると、sensor3, 4, 5, 6, 8, 9, 11, 12, 13 のカメラ及び sensor36 ~ 43 のマイクロフォンが該当している。カメラについては、sensor13 以外はすべて廊下に設置されたカメラである。また、廊下に設置されたカメラのうち、sensor10 以外は、すべて廊下の床面を撮影している。sensor13 は実験室内だが入り口の下置き場周辺を撮影している。

これらのカメラは、観測可能な事象のバリエーションが比較的少ないのではないかと考えられる。

そこでカメラの中で、有意データ節の出力数に対してクラスタリングへの寄与率が相対的に低いものについて、センサ単体で有意データ節の RS 特徴量間の距離を調べた。

有意データ節の出力数に対して、クラスタリングへの寄与率が相対的に低いカメラでは、センサ単体で有意データ節間の RS 相違度を求めたとき、殆どの有意データ節間で相違度が変化せず、かつ比較的相違度が低い値にとどまっていることが分かった。その一例を図 5 に示す。sensor8 がクラスタリングへの寄与率が相対的に低かった例、sensor10 が出力数とクラスタリングへの寄与率の間で相関があると考えられる例である。

これらから、有意データ節の出力数が大きくとも寄与率の小さいセンサに関しては、情報の発生が少ないセンサだと言えることができる。

逆に、有意データ節の出力数に対してクラスタリングへの寄与率が相対的に高くなっている sensor22 では、有意データ節間の RS 相違度に比較的大きなばらつきが確認できた。

ついで、センサ間類似度について考察する。センサ間類似度をプロットしたものが図 6 である。

数力所ずつまとまって類似度の高い領域が存在し、それらは大きく下記の 6 種に分けられる。

- ( 1 ) sensor3, 4, 5, 6
- ( 2 ) sensor8, 9, 10, 11
- ( 3 ) sensor15, 16, 18, 19

- ( 4 ) sensor22 ~ 35
- ( 5 ) sensor36 ~ 43
- ( 6 ) sensor14, 17, 18, 19 と sensor33, 34, 35

このうち、項目 5 はすべてマイクロフォンであり、その感度特性を考えると自然な結果であるといえる。

また、それ以外のペアはそれぞれ設置位置が近いセンサ群となっている。このことは、各カメラが似通った位置を撮影していることから、人間の直感とも一致する。

それ以外に、今回の実験ではマイクロフォンとカメラという組み合わせでは特に類似度が高いペアは見られなかった。

## 10. ま と め

本論文では、生活環境における人間行動一般の判別を目的として、複数種類・多数のパッシブ型マルチメディアセンサから得られる大量のデータに基づいた人間行動判別の枠組みについて提案した。

また、オフィス環境で日常的に発生すると考えられる事象を対象とした実験によって、本論文で提案した枠組みの有効性が示された。

本研究で提案した枠組みにおいて考えられる課題としては、センサにおける特徴量の分解能・次元数について、観測データやセンサが与える情報量に基づいて決定する手法の開発や、異常が発生したセンサを検出し、センサフュージョンの段階において影響が出ないようにする処理手法の開発が挙げられる。

## 文 献

- [1] Rodney. A. Brooks. The Intelligent Room Project *Proceedings of the Second International Cognitive Technology Conference (CT'97)*, pp.271-277, 1997.
- [2] 美濃 導彦. ゆかりプロジェクトの目的と概要 -UKARI プロジェクト報告 No1- 情報処理学会 第 66 回全国大会, vol.5, no.3, TTC-2, pp.5-8, 2004.
- [3] Taketoshi Mori. Hiroshi Noguchi. Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior *First International Workshop on Networked Sensing Systems(INSS2004)*, pp.40-43, 6 2004.
- [4] 西口 敏司. 東 和秀. 亀田 能成. 角所 考. 美濃 導彦. 講義自動撮影における話者位置推定のための視聴覚情報の統合 電気学会論文誌 C vol.124, no.3, pp.729-739, 2004.
- [5] 島井 博行. 栗田 多喜夫. 梅山 伸二. 田中 勝. 三島 健稔. ロバスト統計に基づいた適応的な背景推定法. 電子情報通信学会論文誌 D-II, Vol.J86-D-II, No.6, pp.796-806, 2003.
- [6] S.F.Boll. Suppression of acoustic noise in Speech using Spectral Subtraction *IEEE Trans. Acoust. Speech Signal Process.*, vol.27 No.2, pp.113-120,1979.
- [7] 岡崎 雅嗣. 国本 利文. 小林 隆夫. 信号の定常性を考慮したスペクトルサブトラクション法, 電子情報通信学会技術報告 SP, 91-2003, pp.41-46, 2003.
- [8] 藤岡 一馬. 早坂 昇. 吉田 則信. 宮永 喜一. スペクトルの時間変化を利用した雑音除去, 電子情報通信学会技術報告 SP, 75-2003, pp.7-12, 2003.
- [9] 神鷹 敏弘. データマイニング分野のクラスタリング手法. 人工知能学会誌, Vol. 18, No.1(2003 年 1 月), pp. 59-65
- [10] 矢入 健久. 中須賀 真一. 堀 浩一. 異種冗長なセンサー情報に基づく自律的状态抽象化法 人工知能学会誌, Vol. 14, No.4(1999 年 7 月).
- [11] 服部 傑. 亀田 能成. 大田 友一. 人間行動観測のための視聴覚センサを用いた非正常部分の抽出, 電子情報通信学会技術研究報告 MVE, 2005-51, pp.31-36, 2005.