

# 赤外線センサによる行動予測の実現性のためのデータ照合

村上 雅紀<sup>†</sup> 岡部 正幸<sup>††</sup> 梅村 恭司<sup>‡</sup>

<sup>† ‡</sup>豊橋技術科学大学 情報工学系 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

<sup>† †</sup>豊橋技術科学大学 情報メディア基盤センター 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

E-mail: <sup>†</sup> murakami@ss.cs.tut.ac.jp, <sup>† †</sup> okabe@imc.tut.ac.jp, <sup>‡</sup> umemura@tut.jp

**あらまし** 本研究では人の行動をセンサで観測し、複数の学習データから観測データと類似する学習データを判別することで行動を予測するシステムを提案する。本システムでは複数の赤外線センサから人の行動を観測し、人の行動を示すオンオフの二値情報をセンサデータに記録する。この観測データと前もって準備した複数の学習データと類似判定を行うことで人の行動を予測する。センサデータの類似判定を行う尺度として、Keogh らが提案する Compression-based Dissimilarity Measure (CDM)を用いた。ここで観測途中のデータから人の行動を予測するため、学習データの比較方法について検証し、短いデータにおいて判別精度を向上する補正方法を示した。

**キーワード** 人感センサ, 行動予測

## Data Matching for prediction using infrared sensors

Masaki MURAKAMI<sup>†</sup> Masayuki OKABE<sup>††</sup> and Kyoji UMEMURA<sup>‡</sup>

<sup>† ‡</sup> Department of Information and Computer Sciences, Toyohashi University of Technology 1-1 Hibarigaoka, Tenpaku-cho, Toyohashi, Aichi, 441-8580 Japan

<sup>† †</sup> Information and Media Center, Toyohashi University of Technology 1-1 Hibarigaoka, Tenpaku-cho, Toyohashi, Aichi, 441-8580 Japan

E-mail: <sup>†</sup> murakami@ss.cs.tut.ac.jp, <sup>† †</sup> okabe@imc.tut.ac.jp, <sup>‡</sup> umemura@tut.jp

### 1. はじめに

近年、センサが反応すると照明が点灯する自動証明が多く使われている。こうした自動証明システムでは照明で照らすことが可能な範囲にセンサを取り付け、センサにより人が居ることを判断して照明を点灯させる。こうしたシステムでは、人の行き先を照らすことが出来ず、利用者は暗い場所に向かって移動する必要がある。暗所は状況の判断が難しく、危険を招く場合がある。本研究では視界を確保するため、計算機が人間の行動を分析し、行動を予測することで前もって照明を制御するシステムを提案し、その精度について検討を行った。

上記のシステムを実現するため、部屋に設置された複数の赤外線センサから人の行動を観測し、人の行動を示すオンオフの二値情報をセンサデータに記録する。こうした観測データと前もって準備した複数の学習データと類似判定を行うことで人の行動を予測する。ここでの類似判定とは与えられた1つのセンサデータに対して、複数のセンサデータの中からどれが類似しているかを判定する方法である。本研究では Keogh らが提案する、Compression-based Dissimilarity Measure (CDM)[2]を用いてセンサデータの類似判別を行い、対

応付けされた動作イベントから人間の行動を予測する。

本研究に関連する研究として、情報量の最大化に基づいて赤外線センサの方向を制御する我妻らの研究[5]が挙げられる。ここでは良いセンサの方向を用いた具体的な例として CDM による類似判別を用いた人間の行動判別アプリケーションを利用している。このアプリケーションが人間の行動を判別するものであるのに対し、本研究は観測途中のデータから人間の行動を予測するものである。

### 2. Compression-based Dissimilarity Measure

#### 2.1. 概要

Keogh らが提案する CDM は以下の式によって求められる。

$$CDM(x, y) = \frac{C(xy)}{C(x) + C(y)} \quad (1)$$

$x, y$ : 時系列データを文字列化したもの

$C(x), C(y)$ : 文字列  $x, y$  を圧縮した概念上のデータサイズ

$C(xy)$ : 文字列  $x, y$  を連結したものを圧縮した概念上のデータサイズ

本研究ではセンサデータを文字列化し、複数の文字列データに対して CDM を算出し、CDM の値が小さな値になるデータが類似していると判定する。

同じ条件から得られた 2 つの時系列データを結合して圧縮すると、モデルを共有することができるため、別々に圧縮した場合に比べデータサイズが小さくなるのが期待できる。一方、別の条件から得た時系列データを結合し圧縮しても、モデルを共有できないためデータサイズが小さくなることは期待できない。また、概念上は空文字列  $x$  の圧縮後のデータサイズは 0 となるのが自然である。

### 2.2. CDM の問題点

CDM は使用する圧縮アルゴリズムにより精度が異なるが、今回は菊地らの研究[4]により精度が高いと示された bzip2 を用いて CDM を求める。CDM では、「時系列データ（本研究でのセンサデータ）の文字列化」の方法が類似判別の精度に関連する。本研究では菊地らの研究[4]で高い精度が得られた方法を用いて実験を行う。

実際の圧縮データには圧縮によらずデータサイズが変化しない固定領域が存在する。これにより圧縮したデータサイズは式(2)で表すサイズとなる。このため、実際の CDM は式(3)で表す値となり、データサイズに固定領域が含まれると短い共有データが有利になるという性質がある。

$$C'(s) = C(s) + \beta \quad (2)$$

$$CDM'(x,y) = \frac{C(xy) + \beta}{C(x) + C(y) + 2\beta} \quad (3)$$

$s$  : 圧縮対象の文字列

$C'(s)$  : 文字列  $s$  を圧縮した実際のデータサイズ

$CDM'(x,y)$  : 文字列  $x,y$  から求まる実際の CDM の値

$\beta$  : 固定領域のデータサイズ

## 3. センサデータ

### 3.1. 使用するセンサ

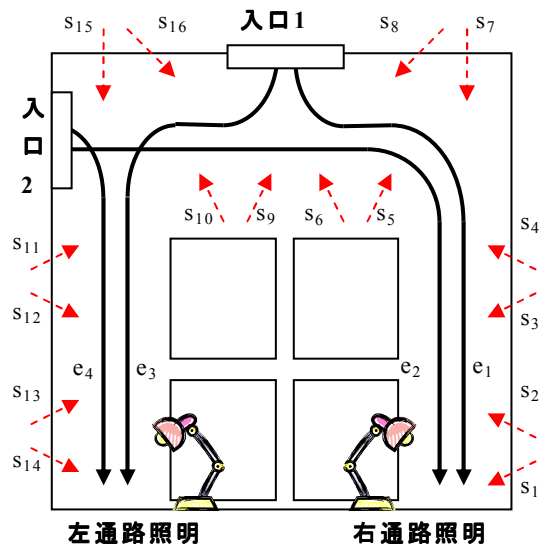
本研究では人感センサとして図 1 のような焦電型赤外線センサを用いる。このセンサはセンシング範囲内で人間が動いた時にオンとし、オン状態を 1、オフ状態を 0 とした二値情報を記録する。今回は人間の行動を想定しているため 0.2 秒間隔で記録し時系列データとする。



図 1 実験で使用する焦電型赤外線センサ

このセンサを図 2 の環境に 16 台設置し、以下の 4 種類のイベントを想定し実験を行う。

- $e_1$  : 入り口 1 から右通路に移動するイベント
- $e_2$  : 入り口 2 から右通路に移動するイベント
- $e_3$  : 入り口 1 から左通路に移動するイベント
- $e_4$  : 入り口 2 から左通路に移動するイベント



赤外線センサ	人がいる	: 1
	人がいない	: 0
センサと取得方向	$s_1 \sim s_{16}$	----->
動作イベント	$e_1 \sim e_4$	————>

図 2 実験環境

### 3.2. センサデータの文字列化

CDM を算出するにあたりセンサデータを文字列化する必要がある。文字列化の方法については下記の菊地らの方法を用いる。

菊地らの研究では同じ状態が長く続き過ぎるデータでは、CDM の精度が低下すると考えている。そこで

時間ごとに並べられた実時間間隔データにおいて図 3 のようにセンサ状態が変化した場合にのみ値を取り出し、それ以外の値を破棄する。このようにして得られたものを可変時間間隔データとよぶことにする。この可変時間間隔データを時間ごとに区切り文字を 1 文字置いて 1 行の文字列として文字列化した。

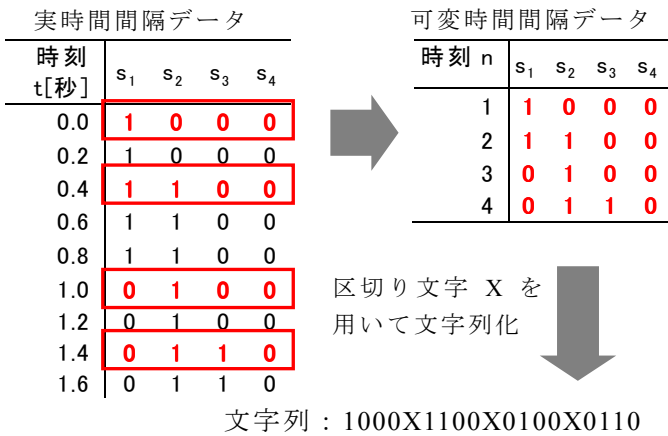


図 3 センサデータの文字列化

可変時間間隔データのある時刻において 1 回記録される長さを 1 とすると、各イベントの可変時間間隔データ 10 件における平均データ長は表 1 のようになる。e<sub>2</sub> が比較的長いデータであるのに対し、e<sub>4</sub> は短いデータであることがわかる。

表 1 各イベントのデータ長

イベント	平均データ長
e <sub>1</sub>	25
e <sub>2</sub>	33.4
e <sub>3</sub>	28.4
e <sub>4</sub>	15.5

## 4. CDM を用いた予測方法

### 4.1. 予測方法の概要

図 4 に本研究で行う予測システムの概要を示す。

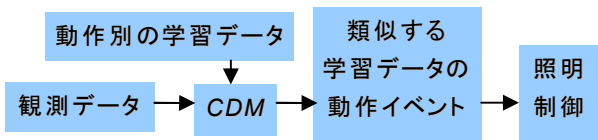


図 4 予測システムの概要

本研究では入力として観測したセンサデータを与えると、そのデータの行動がどのイベントの動作に属するか判別し、対応する照明を制御するものである。予め動作イベントでラベル付けされた学習データを与えておき、入力データと各学習データとの相違度を計算し、入力データと類似する学習データの動作イベントから対応する照明を制御する。このとき、最も類似する学習データのラベルではなく、類似する上位 3 件

のうち 2 件以上が同一ラベルの学習データがみられた場合、そのラベルの動作と判別する。また、2 件以上同一ラベルの学習データが見られなかった場合、どの動作とも判別されない。

### 4.2. 二種類の比較方法

本研究では前もって照明を制御するため、動作途中のセンサデータから次の状態を含めた動作を予測する必要がある。そのため、多くの場合で学習データに比べ観測データは短い。ここで、動作途中のセンサデータに対し学習データ全体を比較する比較方法 1 と、観測データの長さに合わせて学習データの先頭部分を比較する比較方法 2 を提案し検証する。

- 比較方法 1

観測データ : 0001X0011X0110X0101

学習データ : 0001X0010X0100X0101X1010X1100X1000

- 比較方法 2

観測データ : 0001X0011X0110X0101

学習データ : 0001X0010X0100X0101X1010X1100X1000

各イベントに対して 20 件のセンサデータを取得し、内 10 件を学習データ、残り 10 件を観測データとした。この観測データを可変時間間隔上の単位時間で分割し、観測途中のデータとすることで、観測データの長さによって適合率の変化を観察する。

以上の実験で比較方法の差が明確に出たのは e<sub>4</sub> であった、e<sub>4</sub> において比較方法 1 を用いて行ったグラフを図 5 に、比較方法 2 を用いて行ったグラフを図 6 に示す。予測を行うには短いデータで適合率が高いことが望ましく、比較方法 1 では短いデータでの適合率が低い、ある長さを越えると高い適合率を示した。対して比較方法 2 では短いデータに対しても良い適合率を示している。これは学習データの比較部分を観測データの長さに合わせて調整することで、不要なデータとの比較を回避したためと考えられる。また、比較方法 2 の場合、e<sub>4</sub> と判別された動作の適合率は短いデータでは高くなるが、一定の長さを越えると減少するという傾向が確認できた。本来ならば長いデータのほうが情報が多いので、適合率が向上するのが当然と思われるが、これと矛盾している。対して比較方法 1 の場合、e<sub>4</sub> と判別された動作の適合率はデータ長が長くなるにつれて大きくなり、減少することはない。これは、短いデータに対して実際の圧縮データの長さを用いたときに、CDM は有利になるという問題によるものと考えられる。

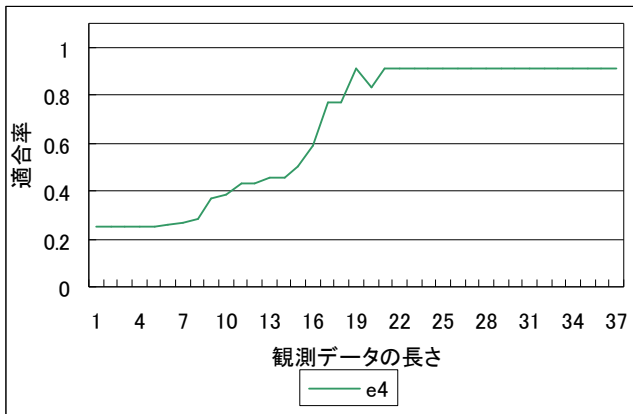


図 5 比較方法 1 (学習データ全体と比較する方法) における  $e_4$  の適合率

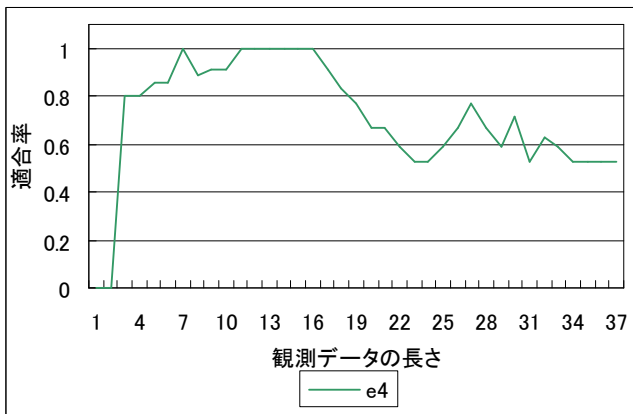


図 6 比較方法 2 (学習データの先頭と比較する方法) における  $e_4$  の適合率

## 5. 圧縮サイズの補正

### 5.1. 固定領域のモデル化

式(3)で示したように、実際の CDM には短いデータが有利になる問題がある。ここで固定領域のデータサイズを求め、式(4)の形でモデル化する。

$$CDM(x, y) = \frac{C'(xy) - \beta}{C'(x) + C'(y) - 2\beta} \quad (4)$$

### 5.2. 固定領域サイズの取得

ランダムな文字列を圧縮し、使用する圧縮アルゴリズムにおける固定領域のサイズを求める。今回使用する bzip2 において、ランダムな文字列のサイズを変化させ圧縮後のサイズにどのように影響があるか調べた。圧縮後のサイズは文字列によって変わるため、同一サイズのランダムな文字列 10 件の平均をそのサイズにおける圧縮後のサイズとした。グラフを図 7 に示す。圧縮後のデータサイズは 37[byte]を切片に単調に増加していることがわかる。常に圧縮データは 1[byte]以上

存在するため、固定領域がバイアスとして 36[byte]存在すると推定できる。

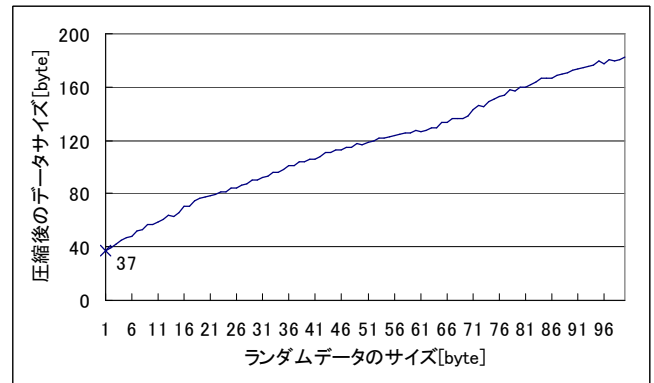


図 7 bzip2 によるランダムデータの圧縮

### 5.3. 固定領域の補正の検証

上記の方法で固定領域のサイズを求め、式(4)により  $C'$ より  $C$ を求めて CDM を計算する。4.2 と同様の実験を固定領域の補正を行うことで、結果にどのような影響が見れるか検証する。

固定領域の補正をして比較方法 1 を行った結果を図 8 に、固定領域の補正をして比較方法 2 を行った結果を図 9 に示す。比較方法 2 の場合、補正無しでは  $e_4$  と判別された動作の適合率が一定の長さを越えると減少していたが、補正有りでは  $e_4$  と判別された動作の適合率はデータ長が長くなるにつれて大きくなり、減少することはない。これは  $e_4$  が他のデータに比べ短いデータであるため、補正無しの場合では他のデータよりも過度に類似していると判別され、他のデータが  $e_4$  と判別された為だと考えられる。そのため、補正有りの場合は固定領域のサイズにより過度に評価されることがなく、正しく判別できている。比較方法 1 の場合、補正をすることで僅かな適合率の向上が見られたが、短いデータに対する適合率が低い問題は解決できなかった。これは比較方法 1 では動作途中のセンサデータに対し学習データ全体を比較するため、長いデータとの比較が多くなり、他のデータよりも過度に類似していると判別される問題が発生しなかった為だと考えられる。そのため、他のイベントにくらべデータ長の短いイベント  $e_4$  ではわずかに精度の向上が見られた。

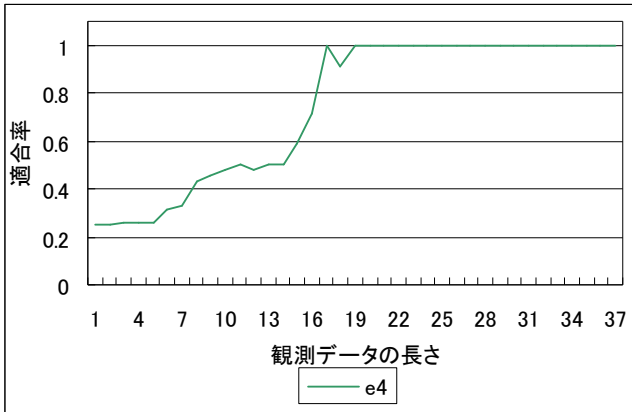


図 8 比較方法 1 (学習データ全体と比較する方法) 圧縮サイズの補正有りにおける  $e_4$  の適合率

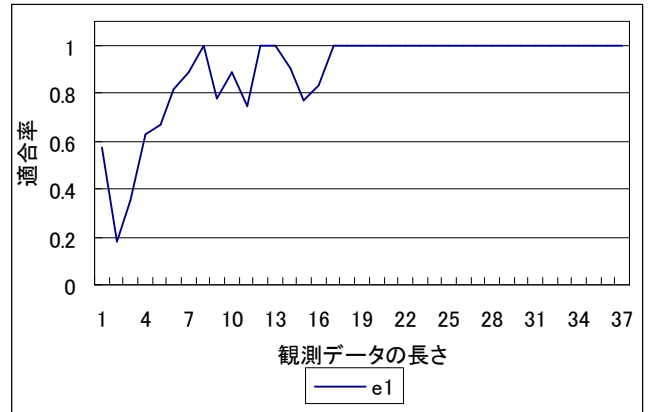


図 10 比較方法 2 (学習データの先頭と比較する方法) 圧縮サイズの補正有りにおける  $e_1$  の適合率

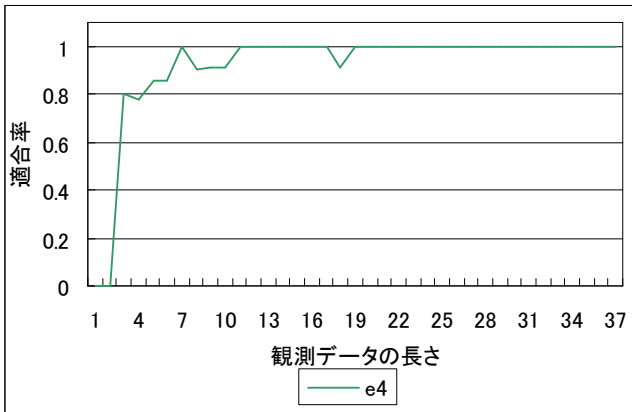


図 9 比較方法 2 (学習データの先頭と比較する方法) 圧縮サイズの補正有りにおける  $e_4$  の適合率

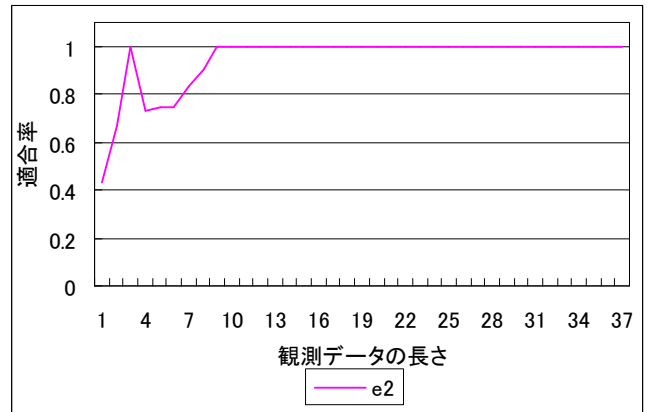


図 11 比較方法 2 (学習データの先頭と比較する方法) 圧縮サイズの補正有りにおける  $e_2$  の適合率

## 6. 補正を加えた場合の比較方法の検討

図 8 と図 9 のグラフを比較するとイベント  $e_4$  では比較方法 2 の先頭との比較を行う方法のほうが短い長さでも精度良く判別できていることがわかる。

また、図 10、図 11、図 12、図 9 に比較方法 2 において圧縮サイズの補正を行う場合の  $e_1$ 、 $e_2$ 、 $e_3$ 、 $e_4$  の結果を示すと、どのイベントにおいても安定して短い長さで判別できていることがわかる。このことから、観測途中のデータから予測を行うには、比較方法 2 の先頭との比較を行う方法に圧縮サイズの補正を行う方法が妥当だといえる。

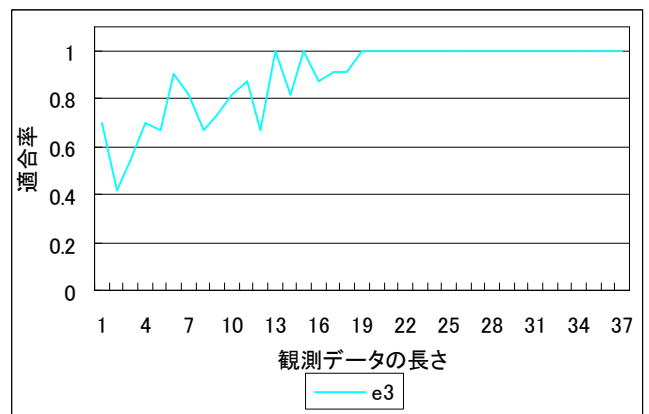


図 12 比較方法 2 (学習データの先頭と比較する方法) 圧縮サイズの補正有りにおける  $e_3$  の適合率

## 7. まとめ

本研究では CDM による類似判別を用いた人間の行動予測を行った。観測途中のデータを精度良く類似判

別するため、学習データ全体と比較する方法と学習データの先頭と比較する方法を提案し、比較を行いそれぞれの方法の特長を調べた。また、短いデータにおける CDM の問題点を調査し、補正をする必要性と補正の方法を明らかにした。以上の結果から、観測途中のデータから予測を行う問題において、比較方法 2 の先頭との比較を行う方法に圧縮サイズの補正を行う方法の妥当性を示すことができた。

### 参 考 文 献

- [1] Deutsch, L. P.: DEFLATE Compressed Data Format Specification version 1.3, *Network Working Group RFC 1951*, (1996).
- [2] Keogh, E., Lonardi, S. and Ratanamahatana, C. A.: Towards parameter-free data mining, *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, (2004), 206-215.
- [3] Shannon, C.E.: A Mathematical Theory of Communication, *SIGMOBILE Mob. Comput. Commun. Rev.*, Vol. 5(2001), 3-55.
- [4] 菊地 誠、阿部 洋丈、岡部 正幸、梅村 恭司: Compression-based Dissimilarity Measure (CDM) を用いた人感センサ情報の類似判定、*情報学ワークショップ 2009(WINF2009) 論文集*, (2009)、185-188.
- [5] 我妻裕樹、阿部洋丈、岡部正幸、梅村恭司: 情報量の最大化に基づく指向性センサの方向制御、*DEIM2010 D5-4*.