

# オフィスのスマート化プロジェクト\*

## 1 C-5 ～その5：オフィス機器制御における入力と環境情報の統合～

山本宏史† 平石広典† 溝口文雄†

東京理科大学 理工学部 経営工学科†

### 1 はじめに

スマートオフィスでは環境内に分散配置された移動式カメラを用いることによりオフィス内の情報を取得することが可能である。そこで本研究では、画像処理や音声認識といった要素技術を利用し、ジェスチャーや音声といったシームレスなインターフェースの観点からスマート化を目指す。しかしながら、こういったインターフェースは時としてノイズが入り認識を困難にさせるといった不確実性を伴い、スムーズな入力への大きな障害となる。一方、近年移動ロボットナビゲーションの分野では複数のセンサー情報の統合において、数多く研究がなされており[1]、ユーザ入力からの情報も一つのセンサー情報とみなせば同様に扱えると考えられる。

そこで本稿ではシステムが実行を決定する際に、ユーザ入力の認識結果のみに依らず、入力規則による区別と環境情報による認識の偏差によって入力列を限定し、システムの決定を強力にする方法を提供する。

### 2 センサー情報統合

スマートオフィスでは様々なタイプの機器が存在し、その動作も多様である。ユーザ入力としてのジェスチャー、音声認識システムは入力されるたびに各々登録された単語の認識の結果を返すがそれらは確定的ではない。そこで以下の2つの段階において入力列を限定する。

#### 2.1 入力規則

入力規則を導入することは、初期入力は命令を送る対象である、といった'カテゴリ'による区別と、この対象の時はこの動作が行なわれるという'属性'による区別を可能にする。例えば、初期入力において対象< Lights (電灯) >が動作< Right (ロボット右回転) >が同程度の認識結果であってもカテゴリから前者と断定することができる。以下に入力手順を示す。

<対象 | 動作 | 目的 | ... >

### 2.2 依存モデル

入力規則によって区別できない同カテゴリーの単語であってもその単語が環境のセンサーとの依存関係があるならば差別化は可能である。ある状態  $S$  のもとで行なわれる入力  $I$  が経験的にわかっているとす。例えば、ロボット前方が障害物が存在するという状況のもとで動作< Stop (止まれ) >はよく入力される。この事前知識を活用した同カテゴリーの単語の差別化の方法を以下に述べる。ここでジェスチャー、音声の認識結果は同スケールで扱うため、正規化するための関数を用意する。

$$f(x; a, c) = \frac{1}{1 + e^{-a(x-c)}} \quad (1)$$

$$c = Base - \alpha * Range \quad (2)$$

ここで状態  $S$  つまりセンサー情報は (2) 式において  $a$  で表現され、1~0 に正規化された値を持つ。Base, Range はそれぞれ認識結果を網羅する最大値、位相の最大範囲であり、単語別に定められている定数である。  $x$  は入力  $I$  に対応した単語の認識結果である。また関数は位相による傾きの変化を考慮してシグモイドを選んだ。位相の範囲は他の単語との相関から誤認を防ぐ最大のものとした。この位相する効果は  $\alpha$  が 0 である時は通常の認識結果がそのまま正規化されるが、  $\alpha$  が大きくなるにつれて確信度は通常より大きくなる。先ほどの例で、同カテゴリー単語< Stop >、< Go >が同程度の認識結果であるならば両者は差別化される。

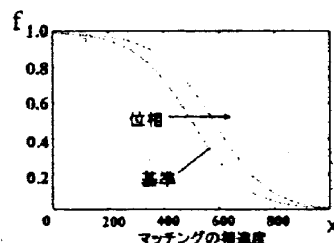


図 1: パラメータ変化による確信度の推移

### 3 システム構成

システムの実装はソケット通信を基本としてそれぞれのプロセスと通信を行なっている。メインのサーバーは

\*Smart office project. -Integration of user inputs and sensors information on office device control

†Hiroshi YAMAMOT, Hironori HIRAISHI, Fumio MIZOGUCH

†Dept. of Industrial Admin. Faculty of Sci. and Tech. Science University of Tokyo

表 1: 機器制御に使用する単語 (R,S,B,L は Robot,Blind,Screen,Lights を表す)

カテゴリー	対象	動作 (属性)	目的	メタ
単語	Robot,Screen Blind,Lights	Go(R),Right(R),Left(R),Stop(R,S,B),Move(R), Hurry(R),Back(R),Up(S,B),Down(S,B),On(L),Off(L)	Bench(R) All(L)	Reset

ジェスチャー認識, 音声認識, ロボットソナー, などの各プロセスから情報を受け, それぞれ正規化する. 入力イベントに応じて先ほど述べた規則を参照し, 対応した実行命令を各デバイスに送る.

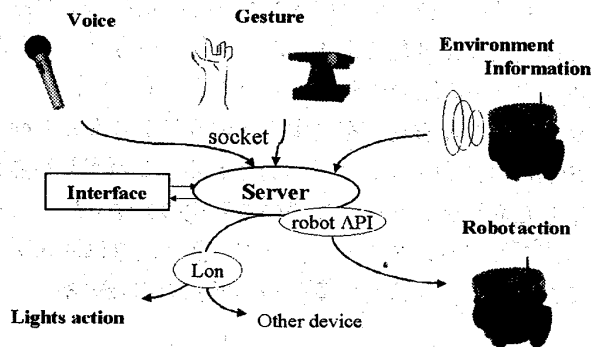


図 2: システム構成

ジェスチャー, 音声その他の統合のためにそれぞれの情報は単語レベルで認識結果を取得する必要があり, その認識はリアルタイムでなければならない. そこで本システムでのジェスチャー認識はスポッティング認識を用いた [2]. 音声認識については IBM の Via Voice の認識エンジンを Java Speech API を用いて実装した.

#### 4 オフィス機器制御における効果

スマートオフィス内の機器を制御する上で各単語の認識率および環境情報をもとに, 制御における前述した方法の効果調べる. 表 1 に機器制御で用いる単語を記す.

##### 4.1 入力規則によるフィルタ

登録した単語群においてどの単語に誤認が見られるかを調べた. 表 2 は誤認識した単語とその関係を記している. 動作 < Back > において < Robot >, < Up > との誤認が見られたが, < Robot > はカテゴリーにおいて, < Up > は属性において異なっており排除可能であった. しかしながら < Stop > では同カテゴリー, 属性の < Right > であるために規則での排除は不可能であった.

表 2: 誤認識した単語 [カテゴリー, 属性]

単語	誤認識した単語
Back [動作,R]	Robot [対象,なし], Up [動作,S,B]
Stop [動作,(R,S,B)]	Right [動作,R]

#### 4.2 環境情報による影響

ロボットはソナーセンサーを持っており, 前方ソナーは危険度として環境情報として扱う. この危険度を < Stop > に反映させることによってロボットが前進している時, この危険度が高まれば (1) 式の位相の変化によりジェスチャーの < Stop > は同じ認識結果であっても高い確信度となる. 図 3 に 10 回の < Stop > の確信度の平均の変化を示す. 確信度の平均は危険度に比例して上昇し, 危険度 1 で 0.96 という結果になった. これにより, 先ほどの < Stop > が < Right > と誤認した場合において, ロボットが危険な状態であれば < Stop > がより高い確信度を持つことがいえる.

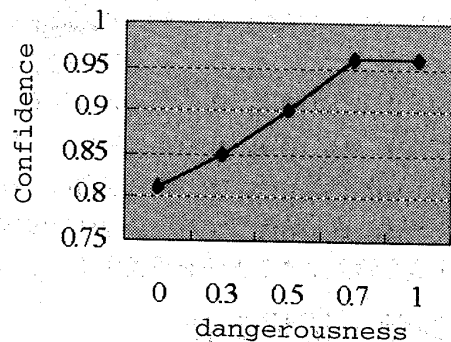


図 3: 危険度による確信度の変化

#### 5 おわりに

本稿では, 複数の入力, とりわけ不確実性をともなう入力に対して入力規則, 環境情報を考慮させるという 2 段階の方法を提案した. また実際の誤認の例からそれらを排除することが可能であることを示した. 音声とジェスチャーの組み合わせについては同時入力可能な点と表現の多様性といった点で有効である. 今後の課題として, より多くのセンサー, 入力からの複数の組み合わせを検討してみる必要がある.

#### 参考文献

- [1] 高山学: 複数センサーを用いた移動ロボットの頑健なナビゲーション, Technical Report on Advanced Intelligent Software and Technology, 1998
- [2] Ryuichi Oka, etc: Realtime Gesture-Speech Human Interface Realized on a Personal Computer of Notebook Size, RWC Symposium, 1998