分散視覚環境における人間の行動認識に向けた行動辞書の作成

○信田洋 (和歌山大学)　港隆史 (CREST, JST)　石黒浩 (和歌山大学)

Construction of a Human Behaviror Dictionary in Distributed Vision Environment

*Hiroshi Nobura (Wakayama Univ.), Takashi Minato (CREST, JST),
Hiroshi Ishiguro (Wakayama Univ.)

Abstract— The goal of this research is to develop intelligent systems based on a new concept called environmental intelligence. The system, consisting of intelligent agents embedded in the environemnt, supports human activities through interactions. As the first step, this paper reports human behavior recognition in a distributed omnidirectional vision system. The traditional recognition systems deal with specified gestures for the task. On the other hand, our system recognizes general human behaviors in our daily life based on a behavior dictionary that includes various primitive behaviors. In this paper we show some of constructed primitive behaviors.

Key Words: Environmental intelligence, Human behavior recognition, Behavior dictionary

1. はじめに

これまでの知能システムが実現を目指している知能の主体は，それ自体が人間と同じように環境の中で活動するエージェント上にあった．しかし人間の日常生活支援を目的としたとき，それとは異なる知能の実現方法として，知能の主体が環境の随所に埋め込まれた環境知能という考え方が重要である．本研究ではそのような環境知能の実現の場として，環境の随所に設置された複数の知覚エージェント (知覚能力, 計算能力, 通信能力を有するもの) からなる計算機ネットワークと実環境とを結ぶ知覚情報基盤 (PII:Perceptual Information Infrastructure)[1] の実現を目指している．知覚情報基盤は単にデータを通信する従来の計算機ネットワークとは異なり, 知覚エージェントによって能動的に獲得される実世界の情報を維持管理し, 人間やロボットなど実世界で行動するエージェントの認知行動を積極的に支援するものである．

このような計算資源を環境の随所に設置しネットワークで結合したシステムは，ユビキタスコンピューティング [2] の考えに基づいた種々の研究 [3] においても論じられている．これらの中には単にネットワークにアクセスするための端末が随所にあるだけでなく，それらを利用して人間行動の支援を試みるシステムもある．しかし基本的にシステムと人間との相互作用は個人を同定できる携帯端末を通して行われるため，システム側の積極的な知覚，認識行動がない．知覚エージェントがネットワークで結合された知覚情報基盤はこの点においてこれらの研究と異なる．

この知覚情報基盤を実現するために，本研究ではまず環境の随所に配置された全方位カメラ [4] から人間の行動を認識する問題を扱う．従来コンピュータビジョンの研究における人間の行動認識に関する研究は，人間の行動を追跡したり，簡単なジェスチャを認識する程度にとどまっていた．音声の研究と比較するなら，音素の同定という比較的入力に近い段階の研究が中心であった．文章認識レベルに到達するためには，音声認識の研究における単語や文法に関する辞書作りが必要である．そこで人間の行動に関してより高いレベルでの認識を実現するために，人間の行動に関する辞書を作成する．それに向けて本報告では行動認識の基本要素となる行動単語の作成について述べる．

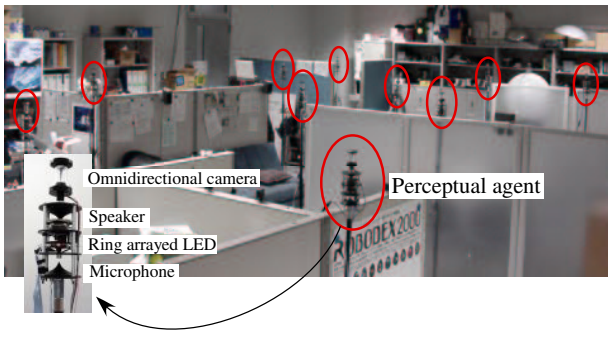

Fig.1 Prototype of PII

2. 基本的考え

2·1 分散視覚による行動認識

本研究では知覚情報基盤のプロトタイプとして，全方位カメラ，マイク，スピーカ，リング状 LED が 1 つになったデバイスを作成し，それらを環境の随所に配置してコンピュータネットワークに結合したシステムを構築した (Fig.1)．このシステムでは全方位カメラ，マイクで人間の行動を観察できるだけでなく，スピーカ，LED により環境側から人間に情報を伝えることができる．

このようにセンサを環境中に分散させることにより，センシングの死角をなくすだけでなく，同時に複数のセンサ情報を用いることにより相互に情報を補うことができる．すなわち複数のセンサを用いることにより，個々のセンサ情報処理を単純化することができ，ロバ

ストな認識システムを構築することができる．視覚においてはカメラキャリブレーションや認識対象のモデルの導入を回避し，見え方に近いレベルでの画像処理のみでシステムを構築することができる．見え方に基づく認識手法は，認識対象が複雑でモデルを構築するのが困難な場合においても容易に対応できるため，より実環境に適していると考えられる．そこで本手法でも画像処理は背景差分による人間抽出のみを行い，背景差分画像およびその時系列から行動認識を行う．

2·2　行動辞書

従来の人間の行動認識に関する研究は，人間を追跡したり，あるタスクに限られたジェスチャのみを認識する程度にとどまっていた．しかし人間の日常生活に入り込み，人間の行動支援を目指す知覚情報基盤にとって，センサデータに基づいて人間の日常生活の行動を記述できるだけの知識が必要である．それを実現するためには膨大な数の行動のモデルが必要となる．そこで本研究では人間の行動の要素を可能な限りモデル化し，その行動要素の連続性に基づいて人間の行動を認識する．そこでは行動要素間の構造のモデル，すなわち行動文法を導入することにより抽象化された認識を行う．この行動文法の基礎となる行動要素の集合は言わば行動辞書であり，行動要素は行動単語となる．

1 つの行動要素は「立つ」,「座る」など言語の動詞レベルの行動とその行動を起こした場所により記述する．各行動要素のモデルは，環境中のあらゆる場所で様々な行動を起こしたときの全ての全方位カメラの観測画像を事例として記録し，それらを教師付きクラスタリングすることにより構築する．

3.　行動要素のモデル化

3·1　手法

全方位カメラ画像から人間領域を抽出するため，背景差分を計算する．背景画像は人間がいないときの画像とし，閾値処理によって差分画像を二値化する．また背景の緩やかな変化に対応するために，背景画像を移動平均によってオンラインで更新する．

環境中のあらゆる場所で種々の行動をとったときに得られた全ての全方位カメラの背景差分画像の時系列を各行動要素の事例として記録する．各事例の時系列長さはあらかじ決められた一定の値とする．そして Fig.2 に示すように 1 事例を入力としたコホーネンネットワークを学習させることにより事例をクラスタリングする．競合層の各ノードについてそのノードが勝者になる事例集合において最も多い事例を求め，それをノードが表す行動要素とする．認識時は全ての全方位カメラの背景差分画像の時系列をコホーネンネットワークに入力し，勝者ノードが表す行動要素を認識結果として出力する．

3·2　実験

構築したシステムの認識能力を検証するために，単純な行動要素を構築する実験を行った．Fig.1 に示すような室内全体に死角がなくなるように 16 台の全方位カメラを設置した．ただし実験には部屋全体の 7 割程度の場所を使用した．計算資源の都合上，画像処理は画

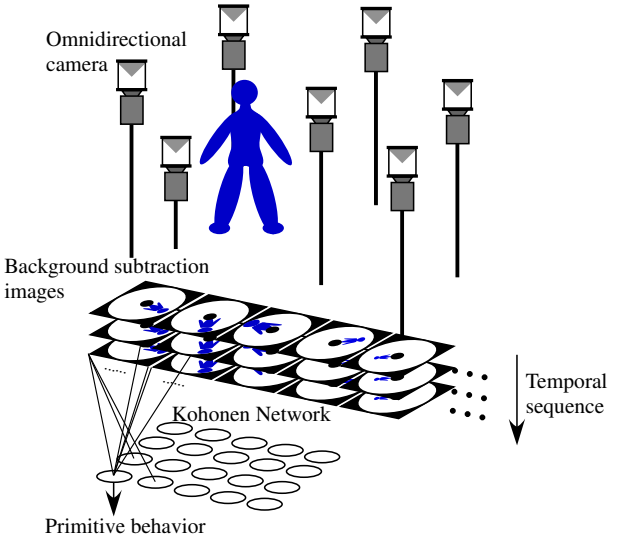

Fig.2 Modeling of primitive behavior

面分割器を使うことにより全ての全方位画像を 1 画面に統合し 1 台の PC 上で行った．1 つの全方位画像の解像度は 80×60 とした．環境中に人間は 1 人しかいないとして行動要素を全部で 25 通り定義した．具体的には実験場を 13 の場所に分割し各場所内で立っている状態および 12 の個所 (椅子あるいは床上) に座っている状態である．ここで用意した行動要素は人間のある瞬間の状態を記述するもので，事例には時系列を必要としない．したがってコホーネンネットワークへの入力は $80 \times 60 \times 16$ 次元のデータとなる．コホーネンネットワークの競合層はトーラス状の 2 次元とし，ノード数は 10×10 とした．各行動要素の事例数は 500 とし，全事例からランダムに事例を選択しネットワークに入力して学習を行った．学習回数は 10000 とした．

3·3　認識結果

学習結果を用いて人間の行動を認識させた結果，認識率は 50%程度であった．これは座っている状態，部屋の端に立っている状態，人間とカメラの間に障害物がある場合など人間がカメラに写りこむ面積が小さい場合に，画像上の人間の領域がノイズに埋もれ，他の行動要素の事例と区別できないためであった．この場合，ある行動要素を表すノードでは，その行動要素がそのノードにおいて勝者となる確率が他のノードと比較して低くなっているはずである．そこでそのようなノードを信頼度が低いとして排除すると 13 の行動要素のみが認識可能なものとして残った．この状態で認識実験を行った結果，ほぼ 100%の認識率が得られた．

本実験における認識率の悪さの基本的な原因は，全方位カメラ配置の空間的な解像度の低さであると考えられる．空間的な解像度が十分にあれば本手法により十分な認識結果が得られると考えられる．

4.　おわりに

本報告では人間の日常生活の行動認識のために，センサデータに基づく行動辞書および行動文法の構築の必要性を提案した．その先に目指すものは行動認識に

基づいて人間と相互作用する環境知能である．本報告では単なる静止状態を表す行動要素しか構築しなかったが，今後時系列を含むさまざまな行動要素を構築していく．それに伴い事例の次元数や行動要素数が膨大になると，行動要素のモデル化手法を改良する必要があると考えられる．

参考文献

1) H. Ishiguro. Distributed vision system: A perceptual information infrastructure for robot navigation. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, pp. 36–41, 1997.
2) M. Weiser. The computer for the twenty-first century. *Scientific American*, pp. 94–10, September 1991.
3) Ubiquitous computing - ubicomp links. http://homepage1.nifty.com/konomi/shinichi/ubicomp.html.
4) H. Ishiguro. Development of low-cost compact omnidirectional vision sensors and their applications. In *International Conference on Information systems, analysis and synthesis*, pp. 433–439, 1998.