DBN を用いた人物の動作パターンに基づく物体認識

三木　博史† 　小島　篤博†† 　宮本　貴朗†† 　黄瀬　浩一†

† 大阪府立大学工学研究科　〒 599-8531 大阪府堺市中区学園町 1-1
†† 大阪府立大学総合教育研究機構　〒 599-8531 大阪府堺市中区学園町 1-1
E-mail: †miki@m.cs.osakafu-u.ac.jp, ††ark@las.osakafu-u.ac.jp, ††aki@center.osakafu-u.ac.jp
†kise@cs.osakafu-u.ac.jp

あらまし　従来, 室内環境で自律的に動作するロボットのための環境認識の研究では, 主に物体の形状モデルに基づいた物体認識手法が提案されてきた. 本研究ではこういった物体の形状モデルを用いることなく, 人物の動作パターンに着目した, 物体の確率的な推定手法を提案する. このような人物動作と物体の関連性は, 人物が特定の物体を扱う際に, 姿勢や手の動きなどに特有のパターンとして表れる. この人物動作のパターンから物体を推定するため, Dynamic Bayesian Networks (DBN) を用いて人物動作の特徴と物体との関係を確率的に表現する. さらに, この DBN に物体に対する人物動作を与えて学習させることで, 物体の認識が可能となる. 最後に, 本手法の有効性を検証するため, 事前に物体の形状モデルを与えることなく室内の物体を認識する実験を行った.

キーワード　環境認識, 物体認識, Dynamic Bayesian Networks, 人物動作認識

Object Recognition Based on Human Actions by Using DBNs

Hiroshi MIKI†, Atsuhiro KOJIMA††, Takao MIYAMOTO††, and Koichi KISE†

† Graduate School of Engineering, Osaka Prefecture University　1-1 Gakuen-cho, Naka, Sakai, Osaka 599-8531, Japan
†† Faculty of Liberal Arts and Sciences, Osaka Prefecture University　1-1 Gakuen-cho, Naka, Sakai, Osaka 599-8531, Japan
E-mail: †miki@m.cs.osakafu-u.ac.jp, ††ark@las.osakafu-u.ac.jp, ††aki@center.osakafu-u.ac.jp
†kise@cs.osakafu-u.ac.jp

Abstract　On conventional research for enviornment recognition which is necessary for robots working autonomously in an indoor enviornment, most of previous methods are based on shape models. In this paper, we propose a method for object recognition focused on the relationship between human actions and objects. Such relationship becomes obvious on human action patterns when he or she handles an object. To estimate object categories by using action patterns, we represent such relationship probabilistically in Dynamic Bayesian Networks (DBN). By learning human actions toward objects statistically, objects can be recognized. Finally, we performed experiments and confirmed that objects can be recognized without shape models.

Key words　enviornment recognition, object recognition, Dynamic Bayesian Networks, human action recognition

1. はじめに

近年, 家庭用ロボットなどの室内環境で自律的に動作するロボットの開発が多く進められている. このような自律活動型ロボットの要素技術の一つとして, 室内の物体配置などの環境認識がある. 従来, 環境認識の研究では, ロボットが目標地点へと到達するまでの経路探索が目標の一つとなってきた [1]〜[3]. しかしながら, こうした手法では, 環境中の物体を障害物として回避することに主眼が置かれ, 個々の物体の認識は必ずしも対象とされていなかった.

ロボットが自律的に動作し適切なサービスを提供するためには, 机や椅子などの物体の種別や, その配置等を認識する必要がある. 物体を認識する手法として, 従来は物体形状モデルに基づいた手法が多く用いられてきた. しかしながら, 一般に, 同じカテゴリの物体であっても, その形状は多様であり, これらすべてを識別する形状モデルを用意するのは困難である.

一方, 形状などの外見的特徴から対象物を識別する従来の手法に対し, 人物の行為・行動の観察を通して対象物の機能属性

を認識し，その結果に基づいて対象物の認識を行うという新しい考え方が提案されている [4]. Moore らは物体に対する人物の手の動きから人物の行動と物体とを推定する手法を提案している [5]. 同様の研究として, 映像中の人物動作から物体領域を推定する手法も提案されている [6]. この手法では, 固定カメラから得られる 2 次元画像を人物動作を基にして領域分割し, ラベル付けを行っているが, 3 次元空間上での物体の認識は行っていない.

人物と物体の関連を用いて 3 次元空間の個々の物体を認識する手法として, 樋口らによるものがある [7]. この手法では, 人物の動作と物体の機能や用途といった概念を階層モデル化することで, 人物動作と物体の形状や機能の統合的な認識が試みられている. また, 高谷らは, 樋口らの手法を自律移動ロボットに適用し, 複数の視点から人物動作を解析することにより動作の対象物を認識し, その結果を環境マップとして構築する手法を提案している [8]. これらの手法では, 人物の動作と物体との関連性を, あらかじめ詳細な概念モデルとして用意する必要がある. また, 動作と物体の関連性を確率的なものとして扱う手法も提案されている [9]. こうした手法では, 人手により概念モデルを設計する代わりに, 人物の動作と物体との関連性を学習により取得することができる.

そこで本研究では, 人間が個々の物体の認識に対応する詳細なモデルを用意するのではなく, 確率ネットワークによって人物動作と物体との関係を表現し, これを学習させることで, 人物動作から物体を推定する手法を提案する. 確率ネットワークは, 時系列上の動的な事象を確率的に扱うことのできる DBN(Dynamic Bayesian Networks) [10] によって表現し, 人物の動作と物体との関係を学習させることで, 対象物体を確率的に推定する.

以下, 2 章では本手法の概要について説明する. 3 章では人物の動作を特徴量として抽出する手法について説明する. 4 章では, 物体と人物動作との関連性を表す DBN の構成について述べ, 物体を推定する手法を説明する. 5 章では, 提案手法を検証する実験とその結果を述べる. 6 章では, 実験に対する考察を述べ, 最後に 7 章でまとめとする.

2. 提案手法の概要

本手法では, 人物の動作と物体との関係を確率ネットワークによって表現し, 人物動作の特徴量を入力として物体を推定する. この確率ネットワークには, 時系列データを扱うことのできる Dynamic Bayeisian Networks (DBN) を用いる [10]. 確率ネットワークは, 3 種類の DBN から構成し, それぞれ人物の移動や姿勢を推定する姿勢 DBN, 人物の物体に対する動作を推定する基本動作 DBN, 物体を推定する関連物体 DBN とする.

認識処理の手順を図 1 に示す. 本手法では, 環境マップ中の物体領域が既に抽出されていると仮定し, これらの領域を識別することを目的とする. 文献 [8] の手法では, 物体が平面で構成されているものと仮定し, ステレオカメラを使って無人の状態での室内の 3 次元マップを作成した上で, 3 次元 Hough 変換により抽出した平面を物体領域としている. 本手法では, 環境の 3 次元マップと物体領域を水平面上に投影し, これを環境マップ

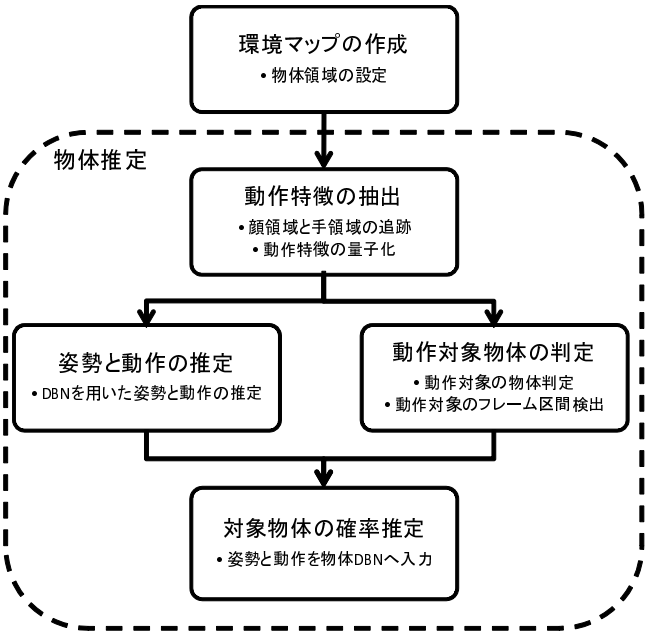


図 1　物体推定手順

とする. 次に, 人物の行動を観察し, 人物の動作をもとに対象領域ごとの物体種類を推定する. 簡単のため, 対象人物は 1 名であるとし, また同時に複数の物体に関連する動作は行わないものとする. 以下に, 物体推定処理の手順を示す.

（1） 動作特徴の抽出

室内の環境マップを作成した後, ステレオカメラを室内の定点に設置し, 人物が行動する様子を撮影する. 最初に, 画像中の肌色領域を抽出し, 人物の手と顔を識別する. そして, これらの領域を追跡し, 頭の位置や, 手と頭の位置関係などを人物の動作特徴として抽出する. これらの動作特徴を量子化し, 姿勢 DBN と動作 DBN の入力とする.

（2） 姿勢と動作の推定

続いて, 人物の姿勢と動作を確率的に推定する. これらは, 人物の動作特徴を入力とする姿勢 DBN, 動作 DBN を用いて推定する. 各フレームでの姿勢と動作の確率値を物体 DBN への入力とする.

（3） 動作対象物体の推定

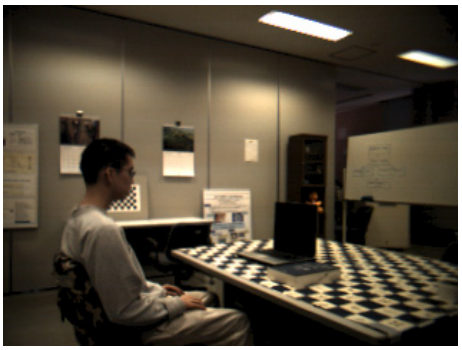
(2) と並行して, フレームごとに環境マップ中に人物の位置を投影し, 人物が近接している物体を, そのフレームでの動作対象物体として判定する. そして, 各物体が動作対象となっているフレーム区間を検出する.

（4） 対象物体の推定

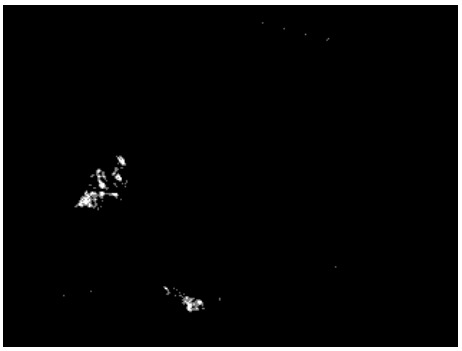
(3) で検出されたフレーム区間での (2) によって得られる姿勢と動作の確率値を対象物体の DBN に入力し, 物体を推定する. 以上の処理を繰り返すことで, それぞれの物体領域が確率的に推定される.

3. 動作特徴の抽出

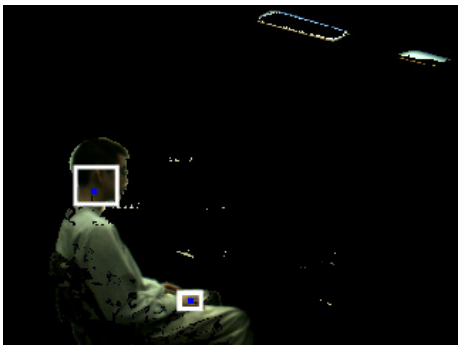
本研究では, 顔と手の動きを追跡することで, 人物の動作特徴を抽出する. 顔と手に注目する理由の一つは, 顔の位置を追跡することで, 立っている, 座っているなどの人物の姿勢が得られ



(a) 入力画像



(b) 肌色領域画像



(c) 顔と手の領域

図 2　肌領域抽出

ることである．また, 物体と関連する人物動作の多くが手を使う動作であるため, 手の位置を追跡することでそれらの手を使う動作を抽出できると考えられる．

3.1　顔領域および手領域の追跡

顔と手の動きを追跡するため, まず画像中の肌色情報を抽出する．肌色情報抽出に関する研究としては, 画像中の肌色情報を抽出し, それらの領域を追跡する手法がある [11]．これは, 肌色の確率分布を肌色モデルとして作成し, 画像中の肌確率を求めることで肌色領域を抽出する手法である．本手法では, $L^*a^*b^*$ 表色系のうち, a^*, b^* を用いて肌色モデルを作成し, 図 2(a) のような入力画像から, 図 2(b) のような肌領域を抽出する．

表 1　動 作 特 徴

動作特徴	ラベル	量子化数
顔の高さ	H_f	3
手と顔の距離	D_{fh}	3
手から顔への方向	A_{fh}	5
手と顔の距離変化	R_{fh}	3
顔の動く速さ	S_f	3
手の動く速さ	S_h	3
顔の動く方向	M_f	6
手の動く方向	M_h	6

本手法では, 人物の動作を観察する際にはカメラの位置は固定されているものとし, 肌色情報抽出において人物の領域以外からのノイズを減少させるため, 図 2(c) のように, 背景差分によって人物領域内のみの肌色情報を抽出し, こうして得られた肌色領域の 3 次元位置をステレオカメラによって取得する．得られる肌色領域のうち, 最も高い位置にあるものを顔領域とする．その他の領域で, 顔からの 3 次元距離が一定範囲内のもので, 最も領域サイズの大きいものを手領域とする．

3.2　人物の動作特徴

前節の処理で得られた顔と手の 3 次元位置から, 人物が物体を扱う際の動作の特徴量を抽出する．本手法では, 表 1 に示される動作特徴を, 特徴量として考える．これらの動作特徴には, 単一のフレームから算出できるもの $(H_t, D_{fh}, A_{fh},)$ と, 現在のフレームと直前の T フレームにおける動作から算出されるもの $(D_{fh}, R_{fh}, S_f, S_h, M_f, M_h)$ とがある．カメラ座標系で, 水平方向を x, 鉛直上向きを y, 奥行きを z とし, 時刻 t での顔と手の位置を $F(t) = (f_x(t), f_y(t), f_z(t))$, $H(t) = (h_x(t), h_y(t), h_z(t))$ として, 各動作特徴を次式のように定める．

$$H_f(t) = f_y(t) \tag{1a}$$

$$D_{fh}(t) = |F(t) - H(t)| \tag{1b}$$

$$A_{fh}(t) = \arccos \frac{f_y(t) - h_y(t)}{D_{fh}(t)} \tag{1c}$$

$$R_{fh}(t) = D_{fh} - \frac{1}{T} \sum_{k=t-T}^{t-1} D_{fh}(k) \tag{1d}$$

$$S_f(t) = |F(t) - F^{'}(t)| \tag{1e}$$

$$S_h(t) = |H(t) - H^{'}(t)| \tag{1f}$$

$$M_f(t) = \arccos \frac{f_y^{'}(t) - f_y(t)}{S_f(t)} \tag{1g}$$

$$M_h(t) = \arccos \frac{h_y^{'}(t) - h_y(t)}{S_h(t)} \tag{1h}$$

ただし,

$$F^{'}(t) = \frac{1}{T} \sum_{k=t-T}^{t-1} F(k), \qquad H^{'}(t) = \frac{1}{T} \sum_{k=t-T}^{t-1} H(k)$$

ここで, A_{fh}, M_f, M_h は, 鉛直下向きからの角度である．すなわち, 水平方向への動作であれば, これらの角度は $\pi/2$ である．

次に, 得られた動作特徴を DBN への入力とするため, 次のよ

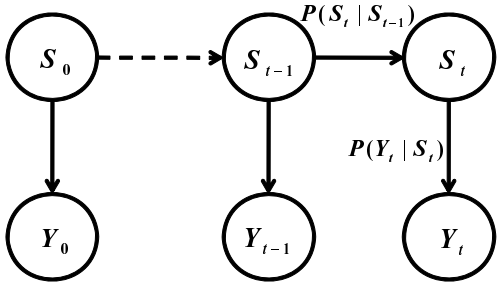


図 3　DBN の基本モデル

うに量子化する.

- 顔の高さ (H_f)

人物が立っている状態, 椅子に座っている状態, しゃがんでいる状態の 3 状態を区別するための閾値を設定する. 立っている状態と座っている状態での顔の高さの平均値, また同様に, 座っている状態としゃがんでいる状態での顔の高さの平均値を閾値とする.

- 顔と手の動く速さ (S_f, S_h)

現在のフレームより前の数フレームと比較して, 静止している, ゆっくり動いている, 速く動いている, 3 状態を区別できる閾値を設定する. ゆっくり動く速さは, 物をとるときの手の動く速さを基準とする.　また, 速く動く場合の速さは, 歩くときの顔の動く速さを基準とする.

- 顔, 手の動く方向, 手から顔への方向 (M_f, M_h, A_{fh})

鉛直下向きからの角度を 5 段階に量子化したものとなる. このときの閾値は, $\frac{\pi}{8}$, $\frac{3\pi}{8}$, $\frac{5\pi}{8}$, $\frac{7\pi}{8}$ である. また, 顔, 手の動く方向については, 静止している状態を含め 6 段階である.

- 顔と手の距離 (D_{fh})

顔と手の距離が近い状態, 肘を曲げた状態, 肘を伸ばした状態の 3 種類が区別できるように閾値を設定する.

- 顔と手の距離の変化 (R_{fh})

顔と手の距離が, 変わらない, 近づく, 遠ざかる, の 3 状態を区別できるように閾値を設定する.

以上のように動作特徴を量子化したものを, 次章で説明する DBN への入力とする.

4. DBN を用いた物体推定

本手法では, 物体と人物動作との関係を確率ネットワークを用いて表現する. 確率ネットワークは 3 種類の DBN を組み合わせて構成する. 最初に DBN の基本モデルについて簡単に説明し, その後, 本手法で構成した DBN について述べる.

4.1 DBN

DBN の基本モデルを図 3 に示す. これは, 時刻 0 から時刻 t までの状態遷移を表したものである. S_t は時刻 t において推定する事象, Y_t は時刻 t における観測値である. 時刻 t までに得られた観測値の系列を $Y_{1:t}$ とすると, 事後確率 $P(S_t|Y_{1:t})$ は, 式 (2) によって計算される [10].

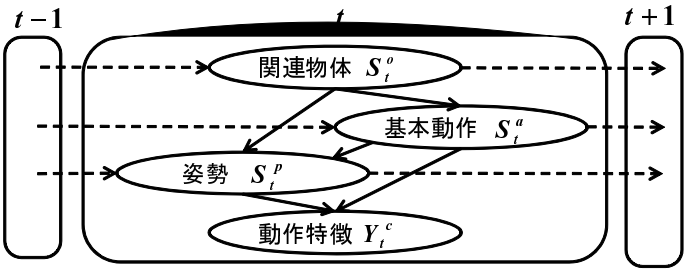


図 4　DBN の構成

表 2　姿勢の状態

姿勢 S_t^p
立っている
座っている
歩いている

表 3　基本動作の状態

基本動作 S_t^a
何もしていない
座る
立つ
物を取る
プリンタに紙補給
板書する
棚に物を置く
ゴミを捨てる

$$P(S_t|Y_{1:t}) = \alpha P(Y_t|S_t) \sum_{S_{t-1}} P(S_t|S_{t-1}) P(S_{t-1}|Y_{1:t-1}) \tag{2}$$

ここで, α は正規化定数であり, $S_t = \{s^0, s^1, ...\}$ に対し $\sum_i s^i = 1$ となるように定める. また, 時刻 $t-1$ での状態 S_{t-1} から時刻 t での状態 S_t へ遷移する状態遷移確率を, $P(S_t|S_{t-1})$ と表す. さらに, 状態が S_t であるときに観測値が Y_t となる観測確率を, $P(Y_t|S_t)$ と表す.

また, 時刻 t での互いに独立した観測値が複数得られる場合, それらを $Y_t^1, Y_t^2, \ldots, Y_t^n$ とすると, 観測確率 $P(Y_t|S_t)$ は式 (3) によって求められる.

$$P(Y_t|S_t) = \prod_{i=1}^{n} P(Y_t^i|S_t) \tag{3}$$

状態遷移確率 $P(S_t|S_{t-1})$ と, 観測確率 $P(Y_t|S_t)$ を, 統計などから事前に学習させることで, 観測値の入力系列から事後確率が求められる.

4.2 動作と物体の DBN

前章で説明した手順で得られる人物の動作特徴を観測値 Y_t^c として, その動作と関連する物体の事後確率を計算する DBN を構成する. 図 4 に構成した DBN を示す.

本手法では, 人物動作から安定して物体種別を推定するため, 歩行や姿勢変化といった人物が姿勢を定めるまでの段階と, 目的の物体に対して手を用いた操作を行う段階とに分けて考える. そして, 人物の状態を表すノードをそれぞれに対応させ, 姿勢 S_t^p と基本動作 S_t^a の 2 種類としている. これは, それぞれの段階において, 物体に対する手の動きの特徴が大きく異なるためである. このような人物の行為・行動の階層性については, 文献 [4] でもその重要性が指摘されている.

以下に説明する DBN を用いて, 各ノードの事後確率を算出

する．

- 姿勢 DBN

動作特徴 Y_t^c を観測値として, 人物の全身の姿勢を推定するためのノード S_t^p の事後確率を算出する DBN である．S_t^p は, 表 2 に示される姿勢の状態を表している．また, 図 4 での動作特徴 Y_t^c は前章で述べたとおり, 実際には顔の高さ H_f, 手と顔の距離 D_{fh} などの複数のノードで構成されている．

- 基本動作 DBN

動作特徴 Y_t^c と, 姿勢 DBN によって得られる姿勢 S_t^p の事後確率値とを観測値として, 人物の物体に関連する瞬間的な動作を推定するためのノード S_t^a を算出する DBN である．S_t^a は, 表 3 に示される基本動作の状態を表している．

- 関連物体 DBN

姿勢 DBN, 基本動作 DBN によって得られる姿勢 S_t^p および基本動作 S_t^a それぞれの事後確率を観測値として, 関連物体 S_t^o の事後確率を算出する DBN である．S_t^o は, 認識対象とする物体の種類に相当する数の状態を表している．

4.3 観測確率行列と状態遷移確率行列

以上の DBN に必要な観測確率行列及び状態遷移確率行列は, あらかじめ撮影した人物行動から得られる統計データを元に算出する．学習用動画中の各フレームでの人物の姿勢, 及び基本動作, その動作対象となる関連物体は手動でラベル付けする．

まず, 姿勢 DBN の確率行列を算出する．連続したフレーム間で姿勢状態が推移する確率 $P(S_t^p|S_{t-1}^p)$ を統計的に求め, 姿勢 DBN の状態遷移確率行列とする．また, 各姿勢状態において観測される人物の動作特徴の確率 $P(Y_t|S_t^p)$ を姿勢のラベルから統計的に求め, 姿勢 DBN の観測確率行列とする．

次に, このようにして得られた姿勢 DBN を学習用動画に適用し, 各フレームでの人物姿勢の確率を算出する．得られた人物姿勢の確率及び動作特徴から, 姿勢 DBN と同様に, 基本動作 DBN の確率行列を算出する．関連物体 DBN の観測確率行列についても同様の手順で求める．ただし, 関連物体 DBN の状態遷移確率行列については, 物体が時間と共に変化することはないため, 本来対角成分が 1 の行列であるが, 時間変化による推定結果の推移を考慮し, これを対角成分が 0.7 で, 残りの確率を均等に割り振る．

4.4 DBN と物体領域の関連付け

本手法では, 同時に行動する人物は一人であると仮定しており, 基本動作 DBN, 姿勢 DBN はそれぞれ単一の DBN を用いる．一方, 物体領域は複数存在するため, 各領域に対して個別に関連物体 DBN を用意し, それらを並行して推定する．

複数の物体領域を認識するためには, 人物が動作の対象としている物体領域を判定しなくてはならない．そのために, 本手法では人物が物体領域に近接している状態を検出する．まず, 人物の手と顔の位置を, 環境マップに投影し, 物体領域との交差を $x-z$ 平面上の位置で判定する．ここで, 人物の顔が物体領域に出入りする場合, その物体領域に対して人物が何らかの動作をしていると考えられる．そこで, 顔が物体領域に出入りする場合, その前後 N フレームを人物がその物体に関連した動作をしているフレーム区間であるとして検出する．

表 4　H_f の観測確率

		H_f		
		低	中	高
S_t^p	立	0.043	0.056	0.891
	座	0.092	0.903	0.005
	歩	0.156	0.008	0.836

表 5　姿勢 DBN の状態遷移確率

		S_t^p		
		立	座	歩
S_{t-1}^p	立	0.981	0.006	0.012
	座	0.004	0.995	0.000
	歩	0.017	0.003	0.980

また, 人物の手が物体領域内にある場合は, 顔と同様に人物がその物体に関連した動作をしていると考えられるので, そのフレーム区間を検出する．ただし, 4.2 節で述べたとおり, 人物が移動しているときと停留している時では手の動きの特徴が違うため, 手の動きによるフレーム区間を検出するのは, 顔の移動による区間が検出されていない場合のみとする．

以上の手順で対象の物体領域に対して人物が動作を行った区間が検出される．この区間での動作と姿勢の確率値を対象の物体領域の関連物体 DBN へ入力することで, 関連物体 DBN の事後確率値が推移する．また, 物体 o に関連した区間 w の最終時点で得られる関連物体 DBN の事後確率値ベクトルを, 区間 w での 物体 o の推定結果 $P_w^o = \{p_0, p_1, ...\}$ (p_i は物体 i であると推定された確率) とし, 物体 o の最終的な物体推定結果 P^o を式 (4) に従って求める．

$$P^o = \beta \sum_w P_w^o \tag{4}$$

ここで, β は正規化定数であり, $\sum_i p_i = 1$ となるように定める．

5. 実　　験

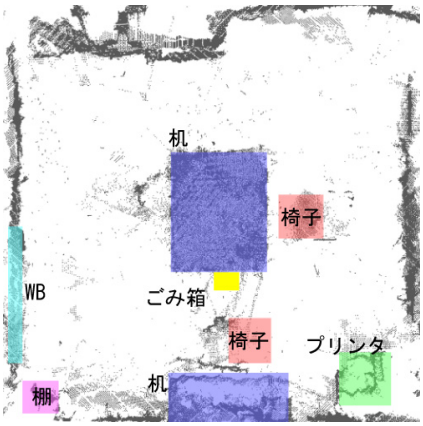
提案手法の有効性を確認するため, 環境マップ内の既知の物体を推定する実験を 2 種類行った．最初は, 環境マップ上で物体領域が互いに重ならないような環境で実験した．続いて, 物体領域が重なる条件のもとで, それぞれの物体がどの程度判別できるのかを調べた．

実験には, ノートパソコン (intel Core 2 Duo T7700 2.0GHz, メモリ 4GB) と, SRI International 社製ステレオカメラを用いた．カメラから得られる画像の解像度は, 320 × 240 ピクセル, ステレオカメラによる奥行き値計算には, ステレオカメラに付属のライブラリを使用した．動画撮影のフレームレートは 15fps で行った．動作特徴抽出手順における, 基準フレーム数は $T = 5$ とした．また, 人物が物体と関連した動作をしているフレーム区間の検出では, フレーム数 $N = 15$ とした．

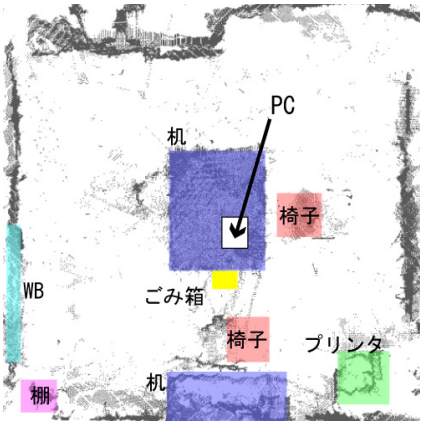
実験の前に学習用動画をあらかじめ撮影し, 4.3 で述べた方法で各 DBN の確率行列を作成した．この手順で作成された確率行列の例として, 表 4 に姿勢 DBN での頭の高さ H_f に対する観測確率行列, 表 5 に姿勢 DBN の遷移確率行列を示す．

5.1 物体領域が重ならない環境での実験

最初の実験では, 図 5(a) に示されるように, 机に推定対象の物体がなく, 物体領域に重なりがない環境で物体の推定を行った．認識対象物体の種類は, 机, 椅子, プリンタ, ごみ箱, 棚, ホワイトボード (WB) とした．実験は, 室内の異なる地点で人物の行動を撮影した 10 本の動画のうち, 安定して人物の動作特徴が抽出され, 人物が物体に近接する区間を正しく検出できて



(a) 環境マップ



(b) PC を追加した環境マップ

図 5　環境マップ

いる 7 本の動画を使用した．提案手法に従い，物体領域の DBN に観測値を入力した．この実験での様子を示したものが図 6(a)，図 6(b) である．

この実験における，机と WB のフレームの変化に対する物体確率の推移を示したものが図 7, 8 である．図 7 では，フレーム区間 240-260 フレーム，320-340 フレーム が該当領域対して人物が動作を行った区間である．図 8 では，グラフに示されるフレーム全体が，該当領域に対する人物動作の区間である．また，表 6 に各物体領域の認識結果を示す．表の対角成分が物体推定の正解確率である．この実験では，すべての物体で，正解の物体確率が最も高くなった．机の推定結果をみると，認識結果は 87.5% と，前物体領域の中で最も高い正解確率が得られた．しかしながら，椅子の正解確率は 26.0% と，机に比べて低い結果も見られた．その他の物体については，正解確率が 約 40% 程度であったが，ごみ箱や WB などは別の物体の確率が正解確率とほぼ同等であり，明確に物体を識別できないものもあった．

5.2　物体領域が重なる環境での実験

続く実験では，図 5(b) のように，机の上に置いたノートパソコン (PC) を認識対象の物体領域として追加した．基本動作の状態にも，表 3 の状態に PC を操作する 状態を加え，あらかじめ学習させた．机と PC の領域が重なる環境で，同様の実験を行い，物体が区別されるかどうかを検証した．このとき用いた動画は，5.1 節での実験で用いたものと同じものであるが，ノートパソコンに対応する領域を追加し，その領域に対する動作を机とは別に検出している点で異なる．

この実験では，机と PC 以外は，ほぼ最初の実験と同様の結

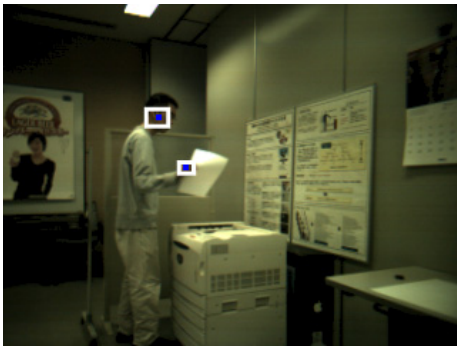
表 6　物体領域の認識結果

領域	物体確率 (%)					
	机	椅子	プリンタ	ごみ箱	棚	WB
机	87.5	1.8	1.0	7.9	1.7	0.1
椅子	11.0	26.0	13.7	25.0	18.5	5.7
プリンタ	3.6	16.1	40.0	9.6	25.7	5.0
ごみ箱	0.9	5.8	12.6	39.5	38.9	2.3
棚	0.2	0.9	9.1	20.1	44.9	24.9
WB	0.1	0.2	1.8	2.4	47.6	47.9

表 7　PC を追加した場合の認識結果

領域	物体確率 (%)						
	机	椅子	プリンタ	ゴミ箱	棚	WB	PC
机	28.3	2.5	0.8	7.7	1.7	0.1	59.0
椅子	6.8	43.1	8.2	25.2	14.1	1.2	1.3
プリンタ	1.0	16.0	36.8	9.9	25.9	5.0	5.4
ゴミ箱	1.1	5.0	10.9	40.8	39.5	2.0	0.6
棚	0.5	0.8	8.7	20.0	45.4	24.7	0.0
WB	0.2	0.2	1.7	2.2	47.5	48.2	0.0
PC	7.4	0.3	0.0	0.2	0.0	0.0	92.1

(a) frame 565



(b) frame 591

図 6　プリンタから紙をとる動作のフレーム画像

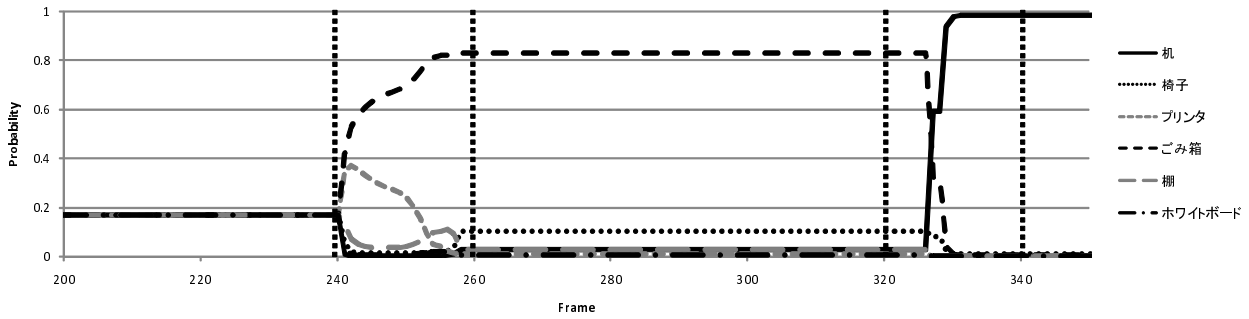


図 7　机領域の物体確率推移

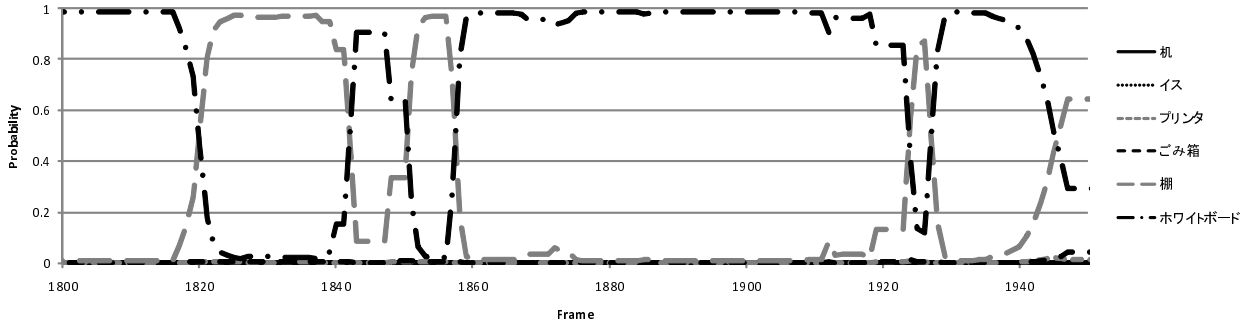


図 8　ホワイトボード領域の物体確率推移

果が得られた. 机領域での PC 確率は 59.0%, PC 領域での PC 確率は 92.1% と, PC の正解確率が高かったが, 机と PC が正しく判別されない場合があった.

6. 考　　察

5.1 節の実験結果からは, すべての物体で正解の物体に対する確率が最も高くなっており, 本手法が物体認識に有効な手法であることが確認された. また, 5.2 節の実験からは, 領域の重なる物体に対しては有効な結果が得られなかったが. これらの実験におけるいくつかの項目について考察する.

6.1　動作対象とする物体領域の判定について

机領域に対する動作区間に関しては, ほとんどの区間で正解確率が高くなり, 結果として机は正しく推定されているが, 一部の区間では正解以外の確率が高くなるものがあった. 例として, 図 7 に示される机領域に対する動作区間 240-260 フレームに注目する. この区間は, 机に対する動作ではなく, 椅子に座る動作をしている際に机に近接した区間を誤って検出したものである. このように本来の物体とは関係ない動作をしている区間の抽出割合が増加すると, 誤認識する可能性が高くなると考えられる. そのため, 正確な物体領域の抽出に加え, 物体領域に対する人物動作をより正確に検出できるよう手法を改善する必要がある.

6.2　類似した動作を受ける物体について

類似した動作を受ける物体は, それらの物体の確率が高くなる結果となった. 本実験では, 特に WB と棚の組み合わせでその傾向が見られた. 表 6 に示した WB の推定結果では, 正解確率が 47.6%, 棚であると誤認識した確率が 47.9% であった. これは, WB に対して板書する動作と, 棚に対して手を伸ばす動作が類似しているためであると考えられる. 図 8 に示される WB 領域の物体確率推移でも, 1830 フレーム, 1850 フレーム, 1925 フレーム付近など, 頻繁に棚の確率が上昇していることからも確認できる. 逆に, 棚の推定結果が, 正解確率 44.9%, WB 確率が 24.9% と, 棚の推定結果が WB の推定結果と比べて良好なのは, WB に対する動作が板書する動作のみであるのに対し, 棚に対する動作が多様であり, 棚を判定する基準が多いことが理由として考えられる. こうした類似した動作を受ける物体を区別するためには, より詳細な動作を分類できるよう, 動作特徴の分類や DBN の構造を検討することなどが必要である.

6.3　動きの少ない動作を受ける物体について

5.2 節の実験では, 結果は PC 領域の正解確率が 92.1% と高い確率であった. しかしながら, 机領域の正解確率は 28.3% であり, 机を PC であると誤認識した確率が 59.0% と高い確率であった. これには, PC に対する人物動作が, 他の物体に対する動作に比べて, 動きの少ない動作である点が原因の一つとして挙げられる. PC に対する動作のように長時間同じ姿勢をとり続ける動作は, 短い時間で終了する動作よりも, 確率推移に与える影響が大きいと考えられる. これは, 同じ姿勢をとり続ける動作は, 学習の際に特定の動作特徴が観測される割合が高くなるためである. こうした長時間同じ姿勢を取り続ける動作を考慮して適切に物体を推定するためには, 動作の分類, DBN の構造についての検討などが必要である.

6.4　従来法との比較

従来, 人物の動作を手がかりとした物体認識手法としては, 人物の動作による対象物や環境の変化を観察するもの [5], [8], また, これに加えて手の動作パターンを補助的に用いるもの [4], [9]

が提案されている．

しかしながら, 人物が手で扱う物体領域の抽出や追跡は困難な場合が多く，条件が限定されるという問題点があった．本研究では, これらに比べて比較的追跡が容易である人物の顔と手を対象とし, その動きを詳細な特徴量として捉えることで, 物体の外見的な情報に依存することなく, ある程度の物体識別が可能であることを示した．

6.5 今後の課題

本手法の課題としては, DBN の構造の検討が挙げられる．人間が考えた抽象的な動作を状態とする他に, DBN の構造自体を学習させる手法についても検討すべきである．

また, 物体領域の抽出法についての検討も必要である．今回は物体を静的な領域としてあらかじめ与えたが, 単に物体領域を静的に切り出すだけではなく, 人物の動作と関連付けて物体領域を抽出していく手法などが考えられる．これは, 物体領域に対する人物の動作と複合的に考慮していかなければいけない問題である．

さらに, 本手法は自律移動ロボットの開発に際して考えられる問題に対する手法の提案であり, 実際に自動的に室内の環境認識を行うロボットに組み込んで本手法を検討していくことが課題として考えられる．

7. ま と め

本研究では, 人物行動と物体との関係を確率ネットワークによって表現し, 室内環境の物体を確率的に推定する手法を提案した．確率ネットワークには DBN を用い, 人物の姿勢や動作といった状態を考慮した確率ネットワークを構成した．また, 実験によって, 人物の動作特徴を抽出し DBN に入力することで物体を推定し, すべての物体で正解の物体に対する確率が最も高くなり, 本手法が物体認識に対して有効な手法であることを確認した．

文　　献

[1] 根岸, 三浦, 白井 : “全方位ステレオとレーザレンジファインダの統合による移動ロボットの地図生成”, 日本ロボット学会誌, **21**, 6, pp. 690–696 (2003).

[2] J. M. Saez and F. Escolano: “Entropy minimization SLAM using stereo vision”, IEEE InternationalConference on Robotics and Automation (2005).

[3] P. Z. Biber, S. Fleck and T. Duckett: “3D modeling of indoor environments for a robotic security guard”, IEEE Workshop on Advanced 3D Imaging for Safety and Security (2005).

[4] 北橋, 樋口, 小島, 福永 : “動作と物体の統合的認識とそのモデル化”, 情報処理学会研究報告, pp. 109–116 (2005).

[5] D. Moore, I. Essa and M. Hayes: “Exploiting human actions and object context for recognition tasks”, IEEE International Conference on Computer Vision 1999 (1999).

[6] P. Peursum, S. Venkatesh, G. A. W. West and H. H. Bui: “Object labelling from human action recognition”, IEEE International Conference on Pervasive Computing and Communication 2003, pp. 399–406 (2003).

[7] 樋口, 小島, 福永 : “人間の動作と動作対象の関連性に基づくシーンの統合的認識”, 画像の認識・理解シンポジウム (MIRU2004), pp. I–469–I–474 (2004).

[8] 高谷, 三谷, 小島, 福永 : “移動ロボットによる人物の動作観察に基づく環境認識”, 2006 Meeting on Image Recognition and Understanding (2006).

[9] 樋口, 小島, 北橋, 福永 : “協調型ベイジアンネットワークを用いた動作と動作対象の統合的認識”, 情報処理学会研究報告, pp. 117–254 (2005).

[10] K. P. Murphy: “Dynamic Bayesian Networks: Representation, Inference and Learning”, PhD thesis, University of California (2002).

[11] 浅沼, 大西, 小島, 福永 : “色情報と領域追跡情報を用いた人物の顔と手の領域の抽出”, 電気学会論文誌, **119-C**, pp. 1351–1358 (1999).