社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－CVIM－148（15）
2005／3／4

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

樋口 未来 †　小島 篤博 ‡　北橋 忠宏 ††　福永 邦雄 †

† 大阪府立大学 大学院工学研究科 E-mail : {higuchi@com., fukunaga@}cs.osakafu-u.ac.jp

‡ 大阪府立大学 総合情報センター E-mail : ark@center.osakafu-u.ac.jp

†† 関西学院大学 理工学部 E-mail : kt@ksc.kwansei.ac.jp

従来，映像理解に関する多くの研究では，動作と物体を個別に認識していた．しかしながら，人間は物体を視覚から得られる外見の特徴のみからではなく，他の人間がその物体を使用する様子を見ることで認識することができる．すなわち，人間は動作と動作対象を統合的に認識していると言える．本研究では，このような人間の認知能力に着目し，人間の動作と動作対象を統合的に認識する手法を提案する．まず，ステレオカメラから得られる動画像と距離データに加え，身体に取り付けたマーカーの 3 次元座標を取得し，人間の動作と動作対象を解析する．次に，人間の動作と動作対象といった複数の相関のある事象や事物を相補的に認識することができる協調型ベイジアンネットワークを提案し，動作と動作対象を統合的に認識する．

Recognition of Human Actions and Related Objects based on Cooperative Bayesian Networks

Mirai Higuchi†　Atsuhiro Kojima‡　Tadahiro Kitahashi††　Kunio Fukunaga†

† Graduate School of Engineering, Osaka Prefecture University

‡ Library and Science Information Center, Osaka Prefecture University

†† School of Science and Technology, Kwansei Gakuin University

In many cases, most of human actions have relation to objects. Human usually recognizes an object not only from the features of the appearance but also by observing another person using it or touching it directly. Many previous research works on image understanding, however, focus on recognizing human actions and related objects separately. In this paper, we propose an integrated method of recognizing them based on cooperative bayesian networks.

1　はじめに

近年，ロボットビジョンや撮影の自動化など人間を支援する技術の実現を目指した研究が盛んに行われている [1, 2]．人間を支援するシステムを実現するためには，人間がコンピュータをただ操作するだけでなく，コンピュータが人間の置かれている環境や状況を判断して人間にアクティブに働きかけることが必要であり，そのためにはコンピュータに人間の動作およびその対象を認識させることが重要と言える．そこで本研究では，人間の動作とその対象物体を認識することを目的とする．

人間は視覚から物体を認識する場合，主に外見の形状，色や質感などから物体を識別し，その用途を認識することができる．しかしながら，見た目の特徴のみからその物体を識別できない場合や，用途を認識できないことがある．そのような場合でも人間は，他の人間がその物体を使用する様子を見たり，あるいは自分で直接触ってみたりすることで物体の用途を認識することができる．これは，

人間が物体を認識する際に物体の外見的な特徴のみではなく，その物体が持ち得る機能を推定しているためである．また逆に人間が他の人間の動作を認識する際に，手の動き等が死角などにより正確に観測できない場合でも，動作の対象としている物体が認識できていればその動作を推定することができる．このことから，物体，特に人工物にはその物体に特定の機能・用途が存在し，物体が認識できていれば人間がその物体に対しどのような動作を行っているか推定できると考えられる．すなわち，人間は動作と動作対象を統合的に認識していると考えられる．

一方コンピュータビジョンにおいては，物体や人間の動作の認識の失敗は頻繁に起り得るため，コンピュータに人間の動作と動作対象を統合的に認識させることは有効と言える．我々はこれまでに，人間の動作と動作対象をフレーム表現により関連付けてモデル化することで，動作と動作対象を統合的に認識する手法を提案してきた [3]．この研究では人間の肢体の動きはあまり考慮していないため，“置く” などの限られた動作のみを認識対象としており，“書く” といった手の動きが重要な意味を持つ動作の認識には至っていなかった．そこで本稿では，時系列の因果関係を扱うことのできるダイナミックベイジアンネットワーク (DBNs) により，動作と動作対象を推定することを考える．このとき動作と動作対象には相関があると考えられるため，それぞれのネットワークが協調的に動作することで動作と動作対象を相補的に認識することができる協調型ベイジアンネットワーク (COBANs:COoperative BAyesian Networks) を提案する．

以下，**2** 章では関連研究を挙げ，本研究との差異を示す．**3** 章では視覚認知過程のモデル化について述べ，**4** 章で提案手法について説明する．**5** 章では提案手法を検証する実験を行うと共に考察・検討し，最後に **6** 章でまとめる．

2　関連研究

人間が環境に働きかけるシーンの動的な性質を定式化した試みとしては，Siskind らの研究が挙げられる [4, 5]．これは，机の上に置かれた物体が人の手によって持ち上げられる様子などを，力学的な解析に基づき定性的に認知する方法を提示したものである．このようなシーンを定性的に認知する試みは非常に重要であり，人間によってもたらされる対象物の移動，変形，分離などを観察し，推論するためには必要と思われる．しかしながら彼らのアプローチはボトムアップ的であり，人間の日常生活における一般的な映像中の物体識別や人物の行動認識にそのまま適用することは難しい．

人間動作の解析により物体を認識する手法が研究されている [6]．これは，人間の座るという動作から椅子の位置・姿勢を推定する手法を提案したものである．この研究では，人間の動作解析による物体の位置・姿勢の推定が目的であり，物体の用途の推定や，動作から対象物体あるいは対象物体からの動作といった双方向的な認識には到っていない．

物体と人物との関わりを，物体の機能面から考察した研究が行われている．例えば，物体の形状とその目的との関係を推論するために提案された機能モデルを，実映像中の人物の摂食動作の認識に拡張した研究 [7]，CAD データで与えられた物体の形状を解析し，その物体が持ち得る機能を推定し物体を識別する研究 [8] が挙げられる．これは，形状を解析することにより物体の機能を推定している点で興味深い．しかしながら，形状の解析による機能の推定のみしか扱っておらず，物体と動作の関連性には触れていない点で我々のアプローチとは異なる．

また，DBNs を用いた映像理解に関する研究が報告されている [9]．これは，DBNs を用いて選手の軌跡やカメラワーク等からサッカー映像のシーンを推定し，映像にインデクシングを行うものであり，複数の相関のある事象の相補的な認識は行っていない．本稿では，複数の相関のある事象を相補的に推定する協調型ベイジアンネットワークを提案する．

3　視覚認知モデル

一般に，人間が視覚から事象や事物を認識すると言っても様々なレベルが存在すると考えられる．例えば，“手を動かす” といった個別の動作から “調理する” といった高次な概念まで多岐に渡る．したがって，本研究では視覚による認識過程を階層的にモデル化する．

人間が視覚から情報を得て認識する過程を，5 段階の階層で表したモデルが考えられている [10]．このモデルの階層は，画像そのものや特徴点検出

などの画像処理のレベルに相当する符号レベルと，認識過程に相当する記号レベルに大別できる．ただし，符号は 1 つずつが意味を持たない量子的単位，記号はある意味を持つ単位とする．符号レベルは，人間の感覚器に相当するカメラ等の機器から得られる入力データと画像の特徴点抽出などの入力データから得られる知覚的特徴からなる．また記号レベルは，符号レベルの知覚的特徴を記号化した概念的特徴，物体や動作といった指示対象と直接対応する単純概念と，その単純概念の連結・合成によって得られる抽象的な事物の概念である連結・合成概念からなる．この 5 段階の階層で表したモデルでは，概念的特徴と単純概念の階層間のギャップが大きく，実際の映像に適用することは難しいと考えられる．

そこで本研究では，6 段階の階層に視覚認知過程をモデル化する．本モデルでは前述の 5 つの階層に，概念的特徴と単純概念の間に映像中の事象に直接対応する事象レベル概念を加えた 6 つの階層で構成される．各階層の詳細は下記の通りである．

1. **入力データ**
   カメラ等のセンサにより取得する入力データ.

2. **知覚的特徴**
   入力データの解析により得られる処理結果.

3. **概念的特徴**
   手の位置や移動方向，物体の形状特徴などの概念的な特徴.

4. **事象レベル概念**
   人間の身体各部位の動きとそれに伴った動作対象の状態の変化に対応する概念．例えば，“物を持ち上げる” 等である．

5. **単純概念**
   各時点における各身体部位の事象レベル概念の組み合わせにより得られる概念．例えば，“座りながら食べる” といった人間の動作と動作対象の具体的な概念である．

6. **抽象概念**
   単純概念および抽象概念を時系列で連結・合成することにより得られる，より抽象的な概念．例えば，単純概念である鍋，コンロなどの連結・合成でキッチンという概念が構成され，切る，煮るなどの概念の連結・合成で調理するという概念が構成される．

上記のように認識は，同時刻の事象を分割した事象レベル概念，同時刻の事象レベル概念の連結・合成からなる単純概念，時系列方向に概念を連結・合成することによって得られる抽象概念の 3 つのレベルからなると定義する．

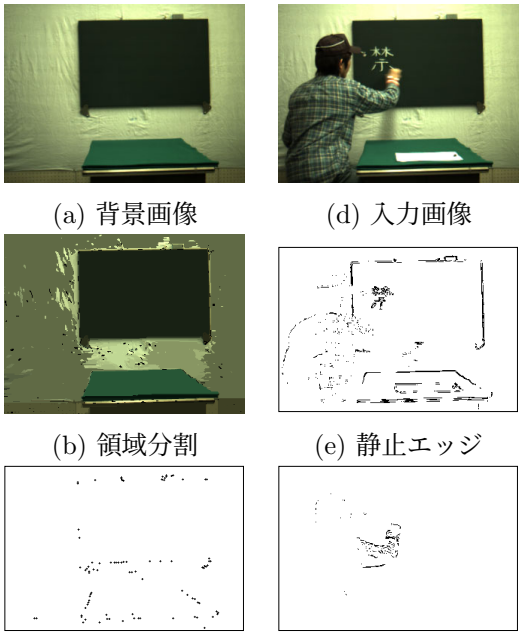
本研究では，この視覚認知モデルによる認識を最終的な目的としている．本稿では，人間の動作と動作対象の統合的認識に主眼を置き，第 4 層の事象レベル概念における手の動作とその動作対象を認識する手法を提案する．他の身体部位の動作および動作対象が認識できれば，先に述べた 6 階層の視覚認知モデルを用いてより抽象的な物体および動作を認識できると考えている．以降ではまず，第 2，3 階層の特徴抽出について述べる．その後，第 4 層の事象レベル概念における動作と動作対象の統合的認識手法について説明する．

4　動作と動作対象の統合的認識

本稿ではまず，人間が日常生活で頻繁に行う “机や棚などの水平面上での作業” や，“黒板やホワイトボードなどの垂直面上での動作” を認識対象とする．前者の動作概念を “act on horizontal board”，後者を “act on vertical board” と呼び，その動作対象を “vertical board”, “horizontal board” とする．また，“黒板やホワイトボードなどに字を書く” といった動作や “黒板やホワイトボードに紙を貼る” といった動作を認識する．それぞれの動作を “write on vertical board”, “pin up on vertical board” と呼び，その動作対象を “character”, “sheet” とする．

4.1　知覚的特徴と概念的特徴の抽出

本稿では，ステレオカメラおよびマーカーの 3 次元座標を取得できるステレオラベリングカメラの 2 つの装置を用いるため，入力データは映像，距離データ，マーカーの 3 次元座標の 3 つが得られる．マーカーは頭部，右手および左手に取り付け頭部と両手の動きを取得する．これらの入力データのうち，映像からは図 1(a) のような入力映像の人間が登場する以前の画像に対して，反復領域拡張法を用いた輝度値による領域分割 (b)，SUSAN オペレータを用いたコーナー検出 (c) を行う．さら



(a) 背景画像　(d) 入力画像

(b) 領域分割　(e) 静止エッジ

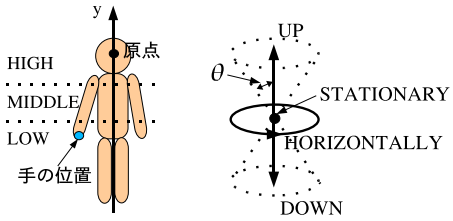
(c) コーナー検出　(f) 動エッジ

図 1: 映像からの知覚的特徴の抽出結果例

に画像を時系列に並べ，x 軸に垂直な面でスライスした断面と y 軸に垂直な面でスライスした断面に対して Sobel オペレータを用いてエッジ検出を行い，そのエッジの出現パターンから動物体によるエッジ (d) と静止物体によるエッジ (e) を抽出する [11]．またマーカーのデータから，3 次元座標系で頭部，右手，左手それぞれを追跡する．以上を知覚的特徴とし，これらの特徴から人間の動きや物体の形状に関する概念的特徴を抽出する．

まずマーカーの追跡結果から動作の概念的特徴を抽出し，次に画像の処理結果からは物体の概念的特徴を抽出する．

- 動作の概念的特徴
  マーカーの追跡結果から，右手の位置，右手の移動方向の 2 つの概念的特徴を抽出する．右手の位置は，図 2(a) のように頭部を原点とした 3 次元座標系において，高さ成分の値により HIGH，MIDDLE，LOW とする．また右手の移動方向は図 2(b) のように，速度の大きさが閾値未満のときを STATIONARY，速度ベクトルと鉛直方向とのなす角が θ 未満かつ上方に移動しているときを UP，下方に移動している場合を DOWN，それら以外の水平方向に移動している場合を HORIZONTALLY とする．ただし，右手が検出されていない場合



(a) 手の位置　(b) 手の移動方向

図 2: 手の動作特徴

いずれの動作特徴も NONE とする．

- 動作対象の概念的特徴
  “act on vertical board” と “act on horizontal board” を認識するために，右手付近での水平面と垂直面の有無 (EXIST, NONE) を概念的特徴として抽出する．また，“write on vertical board” と “pin up on vertical board” の場合は，静止エッジが増加すると考えられるため，右手付近での静止エッジの増加量 (MANY, SOME, FEW) を動作対象の概念的特徴とする．水平面，垂直面の有無の判定方法は以下の通りである．

図 1(a) のように人物が黒板に板書するといった動作を例に説明する．まず，知覚的特徴の領域分割結果に右手のマーカーを投影し，右手の動作の対象となり得る領域を決定する．図 3の例では，領域 r_8 が垂直面あるいは水平面となり得る領域であり，以降この領域を r_α と表す．ただし，$p_1 ... p_m$ は各コーナーの点を，$r_1 ... r_n$ は領域分割により得られた各領域を，m は右手に付けたマーカーを画像上に投影した点を表す．しかしながら，この段階では水平面か垂直面のいずれが存在するか，あるいはいずれも存在しないかは判定できない．そこで，ステレオ視による 3 次元計測結果と SUSAN オペレータによるコーナー検出結果 (図 1(e)) を用いて右手と接する平面が形成できるか検証する．

まず，r_α に含まれるコーナーを抽出する．図 3(右) の例では，r_α は r_8 であったとすると，r_α に含まれるコーナーのリストは以下の通りである．

$$l_8 = \{p_k \mid inc(p_k, r_8)\} = \{p_k \mid k = 8, 9, 10, 11\} \quad (1)$$

ただし，l_i を $\{p_k | inc(p_k, r_i)\}$ を満たす点のリストとし，$inc(p, r)$ は点 p が領域 r 上に存在することを表す．次に，これらの点と右手の位置を 3 次元

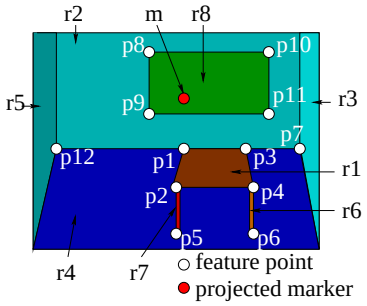


図 3: 平面探索の例

のボクセル空間に投影し，r_α から 3 次元空間中で同一平面上に存在する点のリスト s_α を式 (2) により求める．

$$s_\alpha = \{p_k \mid p_k \in l_\alpha, |z(p_k) - z(p_h)| < \theta_v\} \qquad (2)$$

ここで $z(p_h)$ は右手の 3 次元座標系での z 座標，また $z(p_k)$ は特徴点 p_k の z 座標であり，θ_v は閾値とする．この s_α の点により右手と接する平面が構成できれば，動作対象の特徴量である垂直面は “EXIST” とする．同様に右手と垂直で隣接する位置に垂直面を構成することができれば，水平面を “EXIST” とする．

4.2 DBNs による事象レベル概念の認識

以上で得られた概念的特徴を用いて，動作と動作対象を推定する．先に述べた通り，本稿では COBANs により動作と動作対象を推定する手法を提案するが，まず本章でその基本となる DBNs について説明する．次章で提案手法である COBANs について述べる．

COBANs の基本となる DBNs は，時系列の因果関係を扱うことができるベイジアンネットワークである．その基本形を図 4に示す．これは時刻 $t-1$ から時刻 t の状態遷移を表している．ただし，$S_{(t)} = \{S^1_{(t)}, S^2_{(t)}, \cdots, S^n_{(t)}\}$ は時刻 t における n 個の隠れ状態からなる隠れノードであり，$Y_{(t)}$ は時刻 t における観測値である．$P(S_{(t)} \mid S_{(t-1)})$ は時刻 $t-1$ から時刻 t へ遷移する確率 (状態遷移確率) であり，$P(Y_{(t)} \mid S_{(t)})$ は時刻 t の状態が $S_{(t)}$ である時に観測値 $Y_{(t)}$ が得られる確率 (観測確率) である．DBNs の事後確率分布を計算する方法はいくつか提案されているが，本稿では式 (3) に示す，時

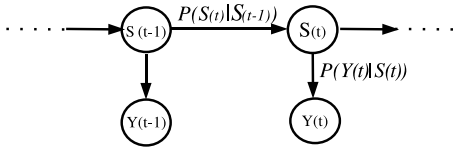


図 4: DBNs の基本形

刻 t までのすべての観測値の系列 $Y_{1:t}$ が得られたもとで時刻 t の状態 $S_{(t)}$ を求める方法を用いる．

$$P(S_{(t)} \mid Y_{1:t}) = \alpha P(Y_{(t)} \mid S_{(t)}) \cdot \sum_{S_{(t-1)}} P(S_{(t)} \mid S_{(t-1)}) P(S_{(t-1)} \mid Y_{1:t-1}) \qquad (3)$$

ただし，上式はマルコフ性を仮定しており，状態遷移は直前の 1 フレームのみを考慮している．また，α は正規化定数である．このネットワークを用いて先に述べた “act on vertical board”$(S^1_{(t)})$ と “on horizontal board”$(S^2_{(t)})$ に加え，それ以外の動作 “other”$(S^3_{(t)})$ を認識する DBNs を構成すると，隠れノードは $S_{(t)} = \{S^1_{(t)}, S^2_{(t)}, S^3_{(t)}\}$ の 3 つの状態を持つ．また，観測値 $Y_{(t)}$ は右手の位置 $Y^p_{(t)}$ と右手の移動方向 $Y^v_{(t)}$ からなる．このとき観測確率が独立であると仮定すると，$P(Y_{(t)} \mid S_{(t)})$ は式 (4) により求まる．

$$P(Y_{(t)} \mid S_{(t)}) = P(Y^p_{(t)} \mid S_{(t)}) P(Y^v_{(t)} \mid S_{(t)}) \qquad (4)$$

動作に加え動作対象を推定するには，独立にもう一つ動作対象を推定するための DBNs を用いることになる．これらの DBNs は，互いに独立に動作するため互いに影響を与えない．一方，COBANs は複数の相関のある DBNs の隠れノード間の関係を考慮に入れ事後確率分布を求める．隠れノード間の関係を扱うことで，複数の DBNs 間で相補的に推論することができる．

4.3 COBANs による事象レベル概念の統合的認識

はじめに述べたように，人間は視覚から人間の動作とその動作対象を認識する際に，動作から動作対象および動作対象から動作といったように相補的に認識していると言える．すなわち，DBNs により動作と動作対象のように相関のある複数の事象や事物を推定する場合，図 5のように認識結果を互いに反映させる必要がある．このとき，認識結果を互いに反映させる前後で状態を分けて考え，

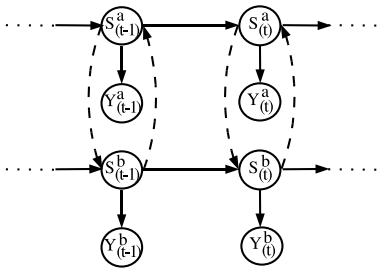


図 5: DBNs から COBANs への拡張

図 6のような動作と動作対象のネットワークを構成し，これを COBANs の基本形とする．動作の隠れノード $M_{(t)}, M'_{(t)}$，物体の隠れノード $O_{(t)}, O'_{(t)}$ の 4 つの状態からなる．$M_{(t)}$ と $O_{(t)}$ は動作と動作対象の直前の状態と観測値により決まる状態であり，$M'_{(t)}$ と $O'_{(t)}$ は $M_{(t)}$ と $O_{(t)}$ の関連性から動作と動作対象を推定し直した状態を表している．このように，COBANs では各 DBNs の隠れノードである $M_{(t)}, O_{(t)}$ の結果を互いに考慮に入れた $M'_{(t)}, O'_{(t)}$ を求めることによって相補的に認識する．これらの隠れノードの事後確率は，時刻 t の動作の観測値を $Y^M_{(t)}$，動作対象の観測値を $Y^O_{(t)}$，時刻 0 から t までのすべての観測値の系列を $Y_{1:t}$ とすると式 (4)〜(7) により求まる．

$$P(M_{(t)} \mid Y_{1:t}) = \alpha P(Y^M_{(t)} \mid M_{(t)}) \cdot \sum_{M'_{(t-1)}} P(M_{(t)} \mid M'_{(t-1)}) P(M'_{(t-1)} \mid Y_{1:t-1}) \quad (5)$$

$$P(O_{(t)} \mid Y_{1:t}) = \alpha P(Y^O_{(t)} \mid O_{(t)}) \cdot \sum_{O'_{(t-1)}} P(O_{(t)} \mid O'_{(t-1)}) P(O'_{(t-1)} \mid Y_{1:t-1}) \quad (6)$$

$$P(M'_{(t)} \mid Y_{1:t}) = \alpha \sum_{M_{(t)}} P(M'_{(t)} \mid M_{(t)}) P(M_{(t)} \mid Y_{1:t-1}) \cdot \sum_{O_{(t)}} P(M'_{(t)} \mid O_{(t)}) P(O_{(t)} \mid Y_{1:t-1}) \quad (7)$$

$$P(O'_{(t)} \mid Y_{1:t}) = \alpha \sum_{O_{(t)}} P(O'_{(t)} \mid O_{(t)}) P(O_{(t)} \mid Y_{1:t-1}) \cdot \sum_{M_{(t)}} P(O'_{(t)} \mid M_{(t)}) P(M_{(t)} \mid Y_{1:t-1}) \quad (8)$$

ただし，上式は式 3と同様にマルコフ性を仮定しており，α は正規化定数である．

動作対象の事後確率が閾値以上である物体は，シーンを 3 次元のボクセル空間で表現した空間にマッピングし，蓄積する．以降，動作対象の概念的特徴の抽出はこの 3 次元のボクセル空間中で右手と接するように “vertical board” や “horizontal board” が存在すれば，概念的特徴の垂直面および水平面は “EXIST” とする．

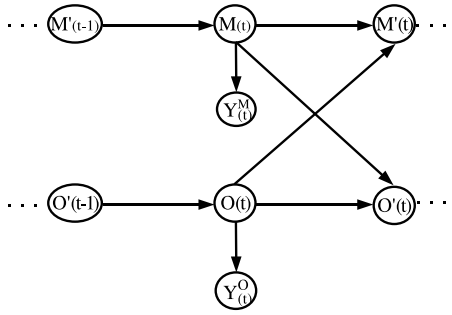


図 6: COBANs のネットワーク

5 実験および考察と検討

提案手法の有効性を確認するため，人間の一連の動作のうち，右手の動作とその動作対象を DBNs と COBANs により認識する実験を行った．ステレオカメラから取り込んだ 320×240 の解像度の画像を処理し，マーカーを頭部，右手，左手に取り付け，その 3 次元座標を取得して人間の動きを解析した．プログラムは C++で開発し，OS は Linux を使用し，CPU の処理速度が Pentium4 2.4GHz のマシンで 7 フレーム / 秒で取り込んだ画像に対して処理を行った．また，床からのカメラの高さは既知とした．

1 つ目の実験では図 7のような，黒板に板書し，黒板に紙を貼るシーンに対して実験を行った．動作の隠れ状態 $M_{(t)}$ は，“act on vertical board”$(M^0_{(t)})$ と “act on horizontal board”$(M^1_{(t)})$ の他，その他の動作あるいは動作をしていない状態を表す “other”$(M^2_{(t)})$ の 3 つ，動作対象の隠れ状態 $O_{(t)}$ は，“vertical board”$(O^0_{(t)})$ と “horizontal board”$(O^1_{(t)})$ の他，未知の物体あるいは対象物が存在しないことを表す “other objects”$(O^2_{(t)})$ の 3 つにより COBANs を構成した．また観測値には，動作の観測値 $(Y^M_{(t)})$ として右手の位置 $(Y^p_{(t)})$ と右手の移動方向 $(Y^v_{(t)})$，動作対象の観測値 $(Y^O_{(t)})$ として垂直面 $(Y^{vs}_{(t)})$ と水平面 $(Y^{hs}_{(t)})$ の有無を用いた．本稿では全ての実験において，DBNs と COBANs ともに観測確率と状態遷移確率の学習は行わず経験的に

#0190 #0330 #0480

図 7: 実験の入力映像

値を決定した．状態遷移確率と観測確率の一例を表 1および表 2に示す．1 つ目の実験による DBNs と COBANs を用いた動作の推定結果は図 8の通りとなった．動作と動作対象の推定結果はともに COBANs による結果の方が安定した認識結果が得られた．180 フレーム付近から 200 フレームにかけて机に紙を置き，400 フレーム付近から 440 フレームにかけて机に置いた紙を手に取るといった動作を行ったが，DBNs では対象物を考慮に入れていないなめ，“act on horizontal board” と “other actions” の識別に失敗している．一方，COBANs はほぼ正しく推定できた．同様に “act on vertical board” についても，“vertical board” の存在を検出できたため COBANs による認識の向上がみられた．これは，動作あるいは動作対象の特徴量を正しく抽出できなかった際に，COBANs では互いの認識結果を考慮にするため互いに認識結果を補うことができたと考えられる．

2 つ目の実験では同一のシーンに対して，$M_{(t)}$ を，“write on vertical board”，“pin up on horizontal board”，および “other actions” の 3 つ，$O_{(t)}$ を，“character”，“sheet”，および “other objects” の 3 つで実験を行った．また $Y^M_{(t)}$ は実験 1 と同じものを用い，$Y^O_{(t)}$ は静止エッジの増加量を用いた．ただし，状態遷移確率および観測確率は実験 1 とは別の値で実験を行った．その認識結果の一部を図 9, 10に示す．470 フレーム付近で行った “pin up on vertical board” は，DBNs では認識できなかったが，COBANs では動作対象である “sheet” の推定結果を考慮することで動作と動作対象を相補的に認識できていることが確認できる．一方，250 フレーム付近から 390 フレームにかけて行った “write on vertical board” は，図 9(a) の動作対象の認識が不十分だったために COBANs の方が DBNs よりも認識結果が悪化しているが，これは動作対象の観測値の抽出精度の向上や，特徴量の追加により解決できると思われる．

以上の結果から，COBANs により相関のある

表 1: 状態遷移確率 (上:右手の動作 下:動作対象)

	$M^0_{(t)}$	$M^1_{(t)}$	$M^2_{(t)}$
$M^0_{(t-1)}$	0.80	0.05	0.15
$M^1_{(t-1)}$	0.05	0.80	0.15
$M^2_{(t-1)}$	0.10	0.10	0.80

	$O^0_{(t)}$	$O^1_{(t)}$	$O^2_{(t)}$
$O^0_{(t-1)}$	0.80	0.05	0.15
$O^1_{(t-1)}$	0.05	0.80	0.15
$O^2_{(t-1)}$	0.10	0.10	0.80

表 2: 観測確率 (上:右手の位置, 下:垂直面)

	$Y^p_{(t)}$			
	LOW	MIDDLE	HIGH	NONE
$M^0_{(t)}$	0.10	0.40	0.40	0.10
$M^1_{(t)}$	0.30	0.55	0.05	0.10
$M^2_{(t)}$	0.45	0.10	0.05	0.40

	$Y^{vs}_{(t)}$	
	NONE	EXIST
$O^0_{(t)}$	0.10	0.90
$O^1_{(t)}$	0.90	0.10
$O^2_{(t)}$	0.60	0.40

複数の事象および事物を統合的に認識することが可能であることを確認し，従来の DBNs に比べて正確な，また安定した認識結果を得ることができた．今回提示した動作，動作対象は想定されるものに限定したが，観測値の改善や追加により様々な動作と動作対象が認識できると考えられる．

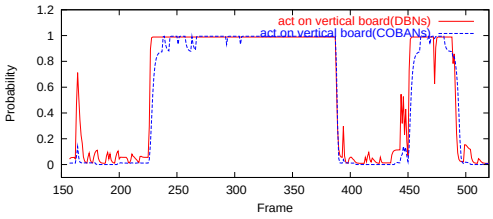
今後の課題としては，状態遷移確率等を学習することが挙げられる．また，視覚認知モデルの単純概念，抽象概念を認識するためには，他の身体部位の動作も同様に認識する必要がある．

6 まとめ

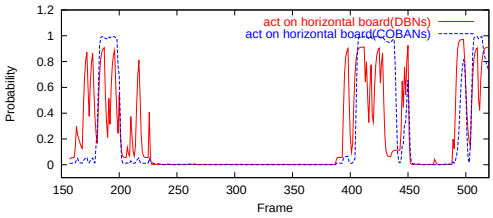
本稿では，従来の映像認識の研究とは異なり，人間の動作とその動作対象を統合的に認識する新たな手法を提案した．動作と動作対象を統合的に認識するために，複数の相関のある事象，事物を相補的に認識できる協調型ベイジアンネットワークを提案し，実験により有効性を確認した．

謝辞

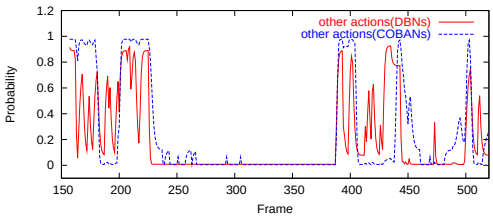
大阪府立大学大学院工学研究科の猪飼 武夫 氏には，本稿の中心的な課題である協調型ベイジアンネットワークについて，研究討論を通して有益



(a)act on vertical board

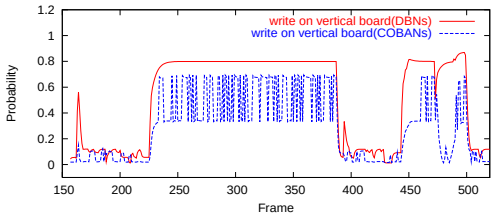


(b)act on horizontal board

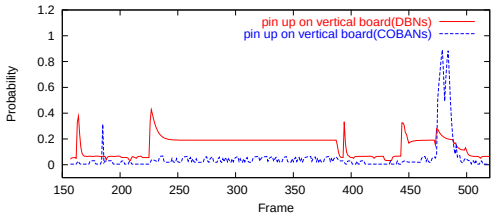


(c)other actions

図 8: 実験 1 での動作の推定結果

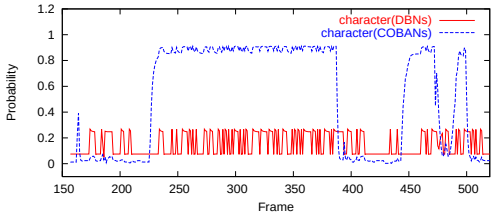


(a)write on vertical board

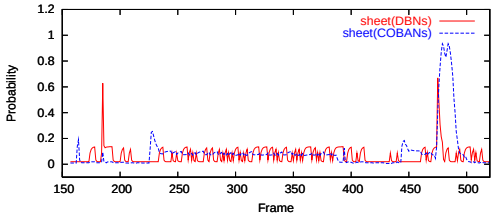


(b)pin up on horizontal board

図 9: 実験 2 での動作の推定結果



(a)character



(b)sheet

図 10: 実験 2 での動作対象の推定結果

な御助言を頂きました．また，本研究の一部は財団法人 原総合知的通信システム基金の援助によるものであり，ここに心からの謝意を表します．

参考文献

[1] H. Ishiguro, T. Maeda, T. Miyashita, and S. Tsuji, “A Strategy for Acquiring an Environmental Model with Panoramic Sensing by a Mobile Robot”, *Proc. of 1994 IEEE International Conference on Robotics and Automation*, 1994, pp.724–730.

[2] 大西 正輝，村上 昌史，福永 邦雄，“状況理解と映像評価に基づく講義の知的自動撮影，” 信学論，vol.J85-D-II, no.4, pp.594-603, Apr. 2002．

[3] M. Higuchi, S. Aoki, A. Kojima, K. Fukunaga, “Scene Recognition based on Relationship between Human Actions and Objects，” Proc. of 17th International Conference of Pattern Recognition, Vol.3, pp.73-78, Aug. 2004.

[4] R. Mann, A. Jepson, and J.M. Siskind, “Computational Perception of Scene Dynamics”, *Computer Vision and Image Understanding*, Vol. 65, No. 2, 1997, pp.113–128.

[5] A.P. Fern, R.L. Givan and J. Siskind, “Specific-to-General Learning for Temporal Events with Application to Learning Event Definitions from Video”, *Journal of Artificial Intelligence Research*, Vol. 17, December 2002, pp.379–449.

[6] 木村 晴次，Michael Hild，白井 良明，“人間の動作の解析に基づく物体認識，” 信学技報，PRMU99-202, pp.71-78, Jun. 2000．

[7] 吉田成希，北橋忠宏，“動作に基づく物体の機能的認識，” 信学技報，PRMU2002-213，pp.13–18, Feb 2003.

[8] Stark, L., and Bowyer, K.W., “Function-Based Generic Recognition for Multiple Object Categories”, *Computer Vision, Graphics, and Image Processing*, No. 1, January 1994, pp. 1-21.

[9] 上垣 直人，中辻 賢治郎，泉 正夫，福永 邦雄，“複合情報を用いたサッカーの放送型映像に対する自動インデクシング，” 画像の認識・理解シンポジウム (MIRU), pp.II-329-II-334, July 2004.

[10] 岡田直之，“語の概念の表現と蓄積，” 電子情報通信学会，1991.

[11] 大西正輝，泉 正夫，福永邦雄，“講義映像における板書領域のブロック分割とその応用，” 信学論，vol.J83-D-I, no.11，pp.1187-1195，Nov. 2000．