

人間と共生する情報システムの実現を目指して

Developing Man-Machine Symbiotic Systems

松山 隆司
Takashi Matsuyama

京都大学
Kyoto University
tm@i.kyoto-u.ac.jp, <http://vision.kuee.kyoto-u.ac.jp/>

杉本 晃宏
Akihiro Sugimoto

国立情報学研究所
National Institute of Informatics
sugimoto@nii.ac.jp, <http://research.nii.ac.jp/~sugimoto/>

佐藤 洋一
Yoichi Sato

東京大学
University of Tokyo
ysato@iis.u-tokyo.ac.jp, <http://jpn.hci.iis.u-tokyo.ac.jp/~ysato/>

川嶋 宏彰
Kawashima Hiroaki

京都大学
Kyoto University
kawashima@i.kyoto-u.ac.jp, <http://vision.kuee.kyoto-u.ac.jp/>

keywords: man-machine symbiotic systems, human interface, man-machine interaction, wearable vision system, intelligent desk.

1. はじめに

「人間と共生する情報システム」という概念は、2001年度より5年間の計画で開始された科学研究費特定領域研究「ITの深化の基盤を拓く情報学研究（代表者：安西祐一郎）」において、第一筆者が柱長を務めている「A03柱：人間の情報処理の理解とその応用に関する研究」が研究目標として掲げているものである。

本解説では、研究の背景とねらい、A03柱における研究の概要を述べる。なお、A03柱でのより具体的な研究内容や成果、活動状況については、A03柱のホームページ [A03-URL] を参照していただきたい。

2. 人間と共生する情報システム

図1は、人間と共生する情報システムとは何かを考えるための背景となる、人間と道具、自動機械との関係を歴史的経緯に沿って左から右に向けて整理したものである。

道具は、古来より、人間にとって生活を営んでいく上で必要不可欠なものである。すなわち、道具とは、人間の身体の一部を拡張し、身体機能を増強するためのものとして捉えることができ、身体と一体化された道具によって我々人間は現在の繁栄を築くことができたといっても過言ではない。

一方、産業革命後飛躍的な進歩を遂げた（自動）機械は、主体である人間に対して従であるものとして位置づけられ、人間の命令を受けその活動を支援するためのものとして捉えられてきた。道具とは異なり、機械はそれ自身の固有の内部状態やダイナミクスを持ち、それら

に基づいて主体からの命令に反応・応答する^{*1}。人間と機械との関わり合いにおいて、この命令と反応・応答を自然に行うことを目指した研究として（広い意味での）インタフェースの研究が進められてきた。

コンピュータは非常に複雑かつ多様な内部状態、ダイナミクスを持つ機械であるため、そのインタフェースの良し悪しはコンピュータの活用度合い・範囲を左右する決定的要因となる。

コンピュータに対するインタフェースの研究は、CUI (Character User Interface) に始まり、現在は、デスクトップメタファを取り入れた GUI (Graphical User Interface) が主流になっている。また、これからのインタフェースとして、力学メディアなど視覚以外のモダリティも含めたマルチモーダルインタフェース (MMUI: Multi-Modal User Interface) や、MMUIに機械が人間のジェスチャや音声を認識する機能を付加した PUI (Perceptual User Interface) が提唱されている [Turk 00]。しかし、これらのインタフェース研究は、人間と機械とは主従の関係にあるという立場、すなわち、人間からの命令とそれに対する機械の反応・応答というコンセプトに立脚している点ではすべて同一の命令 - 応答インタラクション・モデルに基づいていると言える。

一方、21世紀社会が情報社会として発展するためには、ネットワーク結合されたコンピュータ群と人間が一体となって様々な活動を行うことが必要であり、そのためには従来の命令 - 応答モデルを超えた新たなインタラクション・モデルの開発が不可欠となる。

*1 機械については、エネルギー的な視点からの特徴付け——つまり、機械は道具と異なり、内部にエネルギー源を持つ——もあるが、ここでは情動的側面に絞って議論を行う。

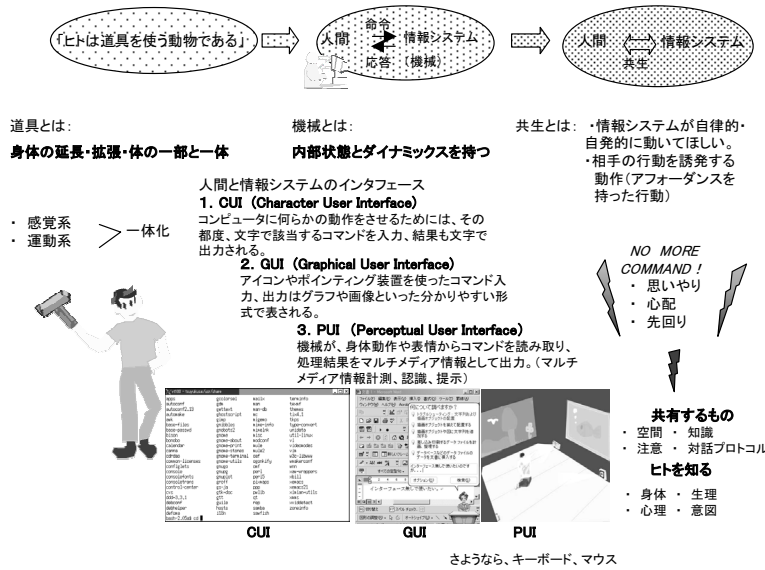


図 1 人間と共生する情報システム (機械) の位置づけ

そうした次世代のインタラクション・モデルとして A03 柱では、「人間と共生する情報システム (機械)」というコンセプトを提案している。すなわち、日常生活環境において人間と共生する情報システムを実現するには、情報システム自身が主体性、自律性を有し、パートナーとして人間との間で双方向の動的インタラクションを行うというモデルが必要となると考えられる。具体的には、人間と共生する情報システムは、人間からの命令を受けて動くだけでなく、情報システム自身が自律的に人間の行動や意図を理解し、それに基づいて人間への指示や情報提供を自発的に行うといった「双方向の (主従関係ではない) 動的インタラクションをスムーズに行う」ことが必要である。

人間との共生を実現するには、まずその基盤として情報システムが人間と空間・時間・知識・注意・対話プロトコル (たとえば、間 (ま) の取り方) を共有することが不可欠となる。また、人間と共生する情報システムを実現するには、人間が情報システムを見てパートナーとして認知することが重要となり、情報システムの持つ身体性 (物理的な身体形状とその動作機能) が情報処理機能とどのような関係を持てば良いのかを解明することも重要な研究課題となる。

こうした問題意識に基づき A03 柱では、日常生活環境において「人間と共生する情報システム (機械)」を実現するための基礎研究として、以下の課題に焦点を当てて研究を行っている (図 2)。

(I) ヒトを知る： 情報システムが自律的に人間の行動や意図を理解するには、まずヒト (生物としての人) の特性や行動パターンを解明し、得られた知見に基づいて共生のために必要な機能を考える必要がある。本課題では、知覚と行動の動的相互作用に焦点を当てた、神経生理学および心理学の立場からの人間の

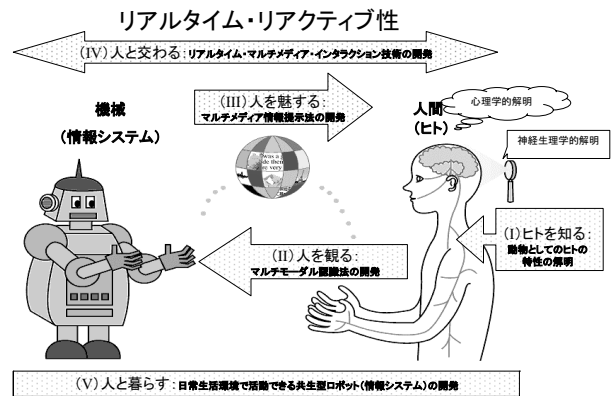


図 2 人間と共生する情報システム実現のための課題

情報処理機能の解明を行う。

(II) 人を観る： 共生を実現するには、情報システムが人間の状態をリアルタイムかつ能動的に観測する必要がある。本課題では、日常生活環境における人間の行動や動作・仕種をマルチメディア情報として観測し、その心理状態や意図、くせなどを理解するための言語・音声・画像・力学メディアを統合したマルチメディア情報認識法・システムを開発する。

(III) 人を魅する： 共生を実現するには、情報システムが自発的かつ効果的に人間に働き掛けを行うとともに、人を魅するような情報提示を行うことが必要となる。本課題では、そのための、力学メディアも含めたマルチメディア情報の生成・提示法およびシステムの開発を行う。

(IV) 人と交わる： 共生を実現するには「間の合った」コミュニケーションがなされなければならない。本課題では、身体およびマルチメディアを介したマル

チモーダルなヒューマン マシン・リアルタイム・インタラクションの実現法、および、それを利用した共生型情報システムの開発を行う。

(V) 人と暮す：人と共に暮すには、情報システムが日常生活環境にうまく馴染まなければならない。本課題では、身体性の持つ情報学的意味の解明（「共生型情報システムには身体が必要か？」）、および、知覚系・行動系を動的に統合化した共生型ロボットシステムの実現を目指す。

さらに、本柱では、(II)～(V)の課題の共通基盤となる新たな情報処理機構として、「自律ダイナミクスを持つ情報システムによる動的イベントの表現・学習・認識・生成」に関する研究を進めている。具体的には、

- (1) 微分方程式系による連続状態変化とオートマトンなどの離散状態遷移との融合・統一的表現が可能な自律ダイナミクスを持つ情報システムのモデル化（課題(II)～(V)の基礎理論）
- (2) 同システムによるマルチメディア時系列信号の認識システム（たとえば、リップリーディング）の実現（課題(II)）
- (3) 連続状態空間中におけるアトラクタへの引き込み現象を利用した滑らかなロボットの行動生成（課題(III)）
- (4) 身体を介した環境との相互作用に基づく感覚運動情報の構造化による行動学習（課題(IV)）(V)の研究を進めている。

3. 研究の概要

以下では、上記の研究課題(I)～(V)、および、自律ダイナミクスを持つ情報システムに関する課題に関して、課題ごとに、A03柱に参加している研究者が進めている研究の概要を整理する。これを通して、それぞれの研究者が「人間と共生する情報システム」というコンセプトに対して、どのような問題意識を持ち、どのような取り組み方をしているかを紹介する。それぞれの研究に関して、より詳細には、[松山 02, Matsuyama 02, 松山 03]を参照していただきたい。なお、筆者らの研究に関しては、その内容を特に詳細に説明する。

3.1 ヒトを知る

(i) サルは道具をどのように見なして扱っているのか?：道具を使った作業の行動学的解析，脳画像解析，脳神経生理学的解析（計画研究：入来班）

ニホンザルが目的を達成するために複数の道具を計画的に組み合わせて使うとき、行為の各要素間の因果関係や論理構造をどのように知覚して全体としての行為を予測的に発現するかという視点から脳内機構を検索した結果、頭頂葉ニューロン群によって分散的に表象された行為の高次構造が前頭葉皮質活動によって合目的に制御され

ていることを示唆する知見が得られた。

(ii) 物・環境はヒトの身体行為に何を与えているか?：物や環境をヒトの行為で記述する方法の開発（計画研究：佐々木班）

重度の運動障害者の運動獲得過程に物がどのような制約を与えているのかを靴下一頸椎損傷者、魚一脳卒中者、卵の表面一高次脳機能障害者という単位で記述した。また、乳児の行為が利用する環境を長期に渡り撮影し、行為の解析に利用可能な動画データベースを作成した。さらに、プロのバイオリン奏者の身体運動に現われるさまざまな水準の協調に着目し、楽器演奏の「芸術性」、「個性」の記述・解析を目指した。

(iii) ヒトは他者をどう捉えているのか?：身体性に基づく行為理解と人間のアンドロイド認知（計画研究：國吉班）

ヒューマノイド全身行動の情報構造分析と身体性学習メカニズムの構築，行為認識の学習モデルの構築，相互作用実験用アンドロイドの構築を行い，さらに，アンドロイドの振舞いを観察している幼児の脳活動を計測し，身体性と他者認知に関する新たな研究アプローチを展開した。

(iv) ヒトが理解しやすい規則とはどのようなものか?：有限状態文法の獲得過程と脳表象（公募研究：岡ノ谷班）

人間の統計的予想に対応する特徴的な脳波波形を発見した。この波形を指標に、さらに、文法的なルールが脳でどう理解されるのかを解明する。

(v) コミュニケーションにおいてヒトはどのように息を合わせるのか?：発話・身振り・呼吸の協調メカニズムの解明（公募研究：古山班）

発話と身振りを使ったマルチモーダル・コミュニケーションの動的メカニズムを、発声の下位運動系である呼吸運動との関連で捉えることを試みた。その結果、1人の人間における場合および人との対話状況における場合のいずれにおいても、発話、身振り、呼吸における動的相互関連性、協調性を示唆する結果が得られつつある。今後は、より日常的な会話場面において、言い淀みなど流暢な相から非流暢な相への転移が起こる際、あるいは逆に非流暢な相から流暢な相へ移行する際に、発話が身体運動及び呼吸とどのように再組織化されていくかに関する解明も進めて行く。

3.2 人を観る

(i) 人は何を見ているのか?：装着型ビジョンセンサの開発（計画研究：松山班）

日常生活の中で人間と自然にインタラクションする情報システムは、その入力情報を人に近い視点で得る必要がある。行動している人の意図や興味等をとらえるという観点からは、その人の視点からの入力情報が必要であると考えられるからである。そこで、本研究では、視線測定装置とコンピュータ制御可能な2台の能動カメラで

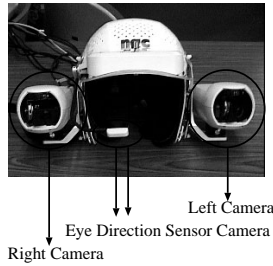


図 3 センサのヘッド部



図 4 ヘッド部を装着した様子

構成される装着型アクティブ・ビジョンセンサを開発した [Sugimoto 02, Sugimoto 03b]. このセンサは、カメラが装着者と視点を共有することで、装着者の視線情報を獲得することができるようになっている。さらに、このセンサを用いた視線情報検出のための自動キャリブレーションアルゴリズムを考案した。

まず、構築したセンサは、計算機とヘッド部に大別される。ヘッド部 (図 3) には、視線測定装置とコンピュータ制御可能な 2 台の首振りカメラを装備した。2 台のカメラの投影中心と左右の眼球中心は全て、ほぼ一直線上に並ぶように設計している。ヘッド部を実際に装着した様子を図 4 に示す。

視線測定装置は、非接触型で、瞳孔/角膜反射法を実装した Nac 製 EMR-8 を用い、右眼の視線情報を獲得する。また、カメラは Sony 製 EVI-G20 を用いた。EVI-20 は、内部にモータを 2 個搭載していて、コンピュータ制御によって、パン (水平回転角度 $\pm 30^\circ$)、チルト (垂直回転角度 $\pm 15^\circ$) 方向に、カメラの撮像系全体を物理的に回転することができる。また、駆動系の回転中心とカメラの投影中心とが一致するように設計されている。一方、計算機は PC (Pentium III 750MHz) を用いた。

視線測定装置で検出された、装着者の右眼の視線情報と 2 台のカメラでそれぞれ撮影された画像が計算機に入力される。装着者の右眼の視線情報は、右カメラで撮影した画像に重ね書きされるようになっている。

この重ね書きを実現するために、EMR-8 には、通常、右眼球の回転中心を原点とする視線角度座標系と右カメラのカメラ座標系とを、同一平面上の点を用いてキャリブレーションするアルゴリズムが提供されている。しかし、そこでは「装着者からキャリブレーションに用いる点が存在する面までの距離はそれほど遠くなく、かつ、装着者が見ている点は、常に、その面上に存在する」ことを前提としている。このため、見ている対象までの距離が変化するような日常生活環境下では、画像上に重ね書きされる点は、装着者が実際に見ている点から大きくずれるという問題が生じていた。

本研究で開発したキャリブレーションアルゴリズムでは、2 台のカメラによってステレオ視を実現させることによって、見ている点までの奥行きを算出し、その奥行きに応じて、重ね書きする点を自動補正する。この補正

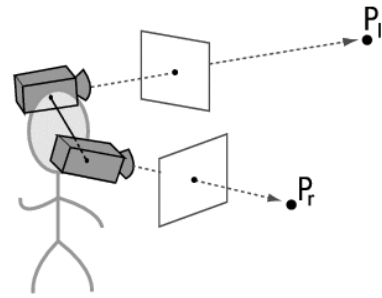


図 5 2 眼独立注視点制御

により、これまで抱えていた上記問題を解決し、見ている点の奥行きが変わるような現実的な環境下でも、正確にその視線情報を検出することが可能になった。

(ii) 人は作業のどこを注視しているのか?: 視線情報と視覚情報の統合による注視点推定手法の開発 (公募研究: 池内班)

マルチモーダルな知覚を用いて人が行う手作業の様子を観察し、その方法や手順を獲得・学習する人型ロボットを開発した。これを用いて、言語化が困難だとされてきた熟練を用する作業の電子保存のための技術、および人間とロボットとの協調作業における作業支援のための技術の開発を進めている。本研究では、データグローブに加えて視線計測装置を利用し、誤差のある視線情報から生成される仮想視覚情報と実際の視覚情報とを位置合わせすることによって、真の視線情報を推定して世界座標系における注視点を得る技術を開発している。

(iii) 人はどう動いているのか?: 装着型ビジョンセンサを用いた人物の移動軌跡推定 (計画研究: 松山班)

人間と動的にインタラクションする情報システムを実現するためには、人間が環境中のどこにいて、どのように移動している/したかをシステムが理解する必要がある。本研究では、2 台の装着型アクティブ・ビジョンセンサを注視点制御 (空間中のある点 (注視点) に常にカメラの視線を向けるカメラ制御) することによって、人物の位置・運動を推定する手法を提案した [杉本 02, Sugimoto 03b]。これによって、装着者の 3 次元位置や運動が自動的に記録できるようになり、位置に基づいた情報提示 (たとえば、博物館・美術館での展示物案内) が人間からの指示なしで行えるようになる。

本手法では、2 台の装着型アクティブ・ビジョンセンサ (図 3 参照) をそれぞれ独立に注視点制御 (2 眼独立注視点制御) し (図 5)、それによって得られる点対応および直線対応を利用することによって、3 次元世界における装着者の回転運動および並進運動を逐次的に推定する。すなわち、異なった時刻 t_1, t_2 に撮影された 2 枚の左画像組 $I_{t_1}^l(x, y), I_{t_2}^l(x, y)$ および右画像組 $I_{t_1}^r(x, y), I_{t_2}^r(x, y)$ のそれぞれの組における注視点対応とその周辺情報としての 2 本以上の直線の対応から、人物の 3 次元

世界での回転運動および並進運動を推定するに十分な幾何学的拘束式を導くことができることを示した。さらにこれを発展させ、直線の代わりに注視点周辺のオプティカルフローから、人物の運動に関する拘束式を導出し、それと注視点対応を用いて人物の運動を推定する手法も提案している [Sugimoto 03a]。

ロボットビジョンの分野では、移動ロボットの位置や運動を推定するために、ステレオ視を利用する手法が一般的である [Davison 98, Molton 00, Murray 96]。しかし、ステレオ視を利用する場合、2台のカメラの視野を共有させる必要があるうえ、運動推定のためには、2台のカメラで撮影された画像において複数の特徴対応を得る必要があり、その処理の安定性に問題が生じる。さらに、運動の推定精度は2台のカメラの基線長に依存するが、装着型の場合は基線長を長くすることがむずかしい。

これに対し、2眼独立注視点制御では、各カメラを独立に制御して空間中の点を注視させればよい(注視する点を2台のカメラで一致させる必要はない)ので、2台のカメラの視野を必ずしも共有させる必要がない。さらに、カメラが注視している2点からみると、人物を対象としてステレオ視をしている状況であると解釈できるので、その場合の基線長はカメラが注視している2点間の距離になる。すなわち、空間的に離れた点をそれぞれのカメラに注視させることにより、装着した2台のカメラの基線長に依存せず、高精度の運動推定が可能になる。このように、2眼独立注視点制御によって、高精度に人物の運動を推定することが可能になる。

(iv) 人は何を体験しているか?: ウェアラブルコンピュータを用いた、主観的体験記録の研究(公募研究: 廣瀬班)

日常生活において、体験者の行動を極力妨げない状態で、人間の視野角範囲の記録および身体、頭部の姿勢角の記録を行うウェアラブルコンピュータのプロトタイプを開発した。また、ウェアラブルコンピュータを用いて、個人の主観的体験を記録することによって、人間の体験構造を知覚的認知的側面から解明する研究も行った。この研究では、大量の体験データを蓄積することが可能になり、体験とは何か、記憶とは何かについて解明するための手がかりを得ることができると考えられる。

(v) 人の姿勢、動作を観る: 人体動作の実時間3次元形状計測システムの開発(計画研究: 松山班)

多数のカメラによって撮影した映像から、人間の詳細な3次元姿勢や動作を実時間で計測できるシステムを開発した(図6)。こうしたシステムを利用することによって、(a)人間が行う指差し動作を分析して3次元指示方位を求め、人間が何を指差しているかを理解する、(b)ジェスチャを使ったコンピュータとのPerceptual User Interfaceの実現、(c)3次元的人間の移動経路や速度の計測、(d)計測された人間の動きを基に人間に追従するロボットの実現、などができるようになった。開発したマルチカメラシステムは、2 cm程度の空間解像

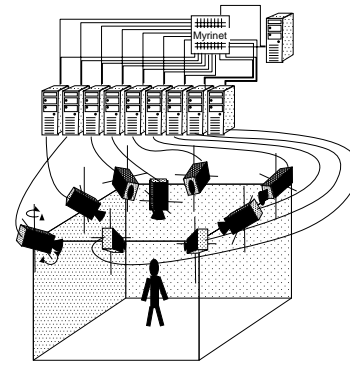


図6 実時間3次元動作計測システム

度で人体の姿勢や動作をリアルタイムで計測することができ、コミュニケーションに伴うジェスチャの定量的解析など「しっくりとした」コミュニケーションの実現を目指したインタラクション・プロトコルの開発にも利用できる。

(vi) 人の動作を観る - 動作に基づくインタラクション: ビジョン技術を基にした誘発型ユーザインタフェースの開発(公募研究: 谷口班)

ビジョン技術を用いて非接触で人間の身体動作を実時間で計測する手法を開発し、その動作情報と仮想化されたオブジェクトの持つ属性情報を基にインタラクションを誘発的に行う方式を開発した。

(vii) 人の行動を知る: 人間の行動観察のためのセンシングシステムの構築(公募研究: 八木班)

周囲360度が実時間観測できる全方位視覚センサを入力手段として、日常生活における人間の行動を非接触で観察できる実時間人間情報計測システムを構築し、自然歩行画像から個人認証を行うシステムを構築した。

(viii) 観て動作が分かるとは?: 人の日常生活動作認識モデル(公募研究: 森班)

モーションキャプチャで得られるデータを分析し、同一動作と認識される運動に共通に現れる特徴量を抽出し、それらを注目量とする認識アルゴリズムを構成した。また、日常生活を行う部屋の中に分散配置した圧力センサと磁気式のモーションキャプチャシステムから構成される日常生活行動記録システムを構築し、計測されたセンサ情報に基づいた人間の動作認識アルゴリズムを考案した。こうしたシステムを利用すれば、人間が無意識に行っている日常行動のパターンが把握でき、機械がさりげなく人の生活をサポートすることが可能となる。

(ix) 行動と視覚と空間認識のむすびつき: 能動視覚による動的な空間知覚と立体形状認識機構の解明とその応用システムの構築(公募研究: 出口班)

人は能動的な行動とそれによって引き起こされる視覚情報の変化を対照することで、空間の知覚を行っている。この機構をロボットに模倣させることで、ロボットの視覚の諸パラメータを自己校正するとともに脳内空間マッ

ブを自動生成するシステムを開発した。

(x) 人は人の存在から何を感じているのか?: 音声・音響情報処理に基づく話者情報抽出技術の高精度化(公募研究: 峯松班)

音声からの年齢推定技術を構築した。高齢者のみならず、子供音声までを対象とし、聴取実験に基づいて各話者の知覚年齢を分布として定義し、それに基づいて未知話者の知覚的年齢(分布)の推定を行った。

(xi) 人は何を見、何を感じながら物を把持し操るのか?: 遠隔操作プラットフォームを用いた人間の把持繰り動作時の視覚情報処理過程の解明(公募研究: 横小路班)

人間の把持様式を身の回りの日用品 50 品目で徹底的に計測し、把持物体を模擬するデバイスの機構設計に役立てた。また人間への視覚情報提示を完全に把握しかつ選択的にコントロール可能な遠隔操作プラットフォームの基本部分を構築した。

3.3 人を魅する

(i) 部屋中どこにでも重畳表示: 実世界環境モデルの自動獲得とユビキタスディスプレイへの利用(計画研究: 松山班)

実世界のあらゆる表面に情報を重畳表示するユビキタスディスプレイ機能を実現するために不可欠となる、レンジセンサとカラーカメラを含む複数センサによる自動キャリブレーション手法を提案し、家具などの配置が動的に変化する室内環境において基礎的な評価実験を行った。

(ii) 言葉では表せない意図を伝えるために: 非言語的情報の実世界への提示とコミュニケーションの円滑化(公募研究: 日浦班)

プロジェクタや人型ロボットを用いて実世界中に遠隔地からの非言語的情報を提示することで、共同作業では必須の指差しや手真似で行われる意図伝達を実現し、遠隔地間のコミュニケーションを円滑化する。

(iii) 人が見ることが出来る領域をどこまで広げられるか: 高精細全方位ビデオカメラとネットワークによるテレプレゼンス(公募研究: 山澤班)

高精細全方位ビデオカメラとネットワークを組み合わせ、人間があたかも遠く離れた場所にいるかのように周りを見ることが出来る遠隔臨場感システムを開発した。

(iv) 映像文化とマルチメディア技術の融合: 人物行動を伝えるための映像文法を用いた知的映像撮影・編集システムの構築(公募研究: 中村(裕)班)

「効果的に情報を提示するために、どのような映像部品(単語)を用意し、どのような映像編集規則(文法)にしたがって提示すればよいか」という問題に対し、自動撮影システム、自動編集システムを構築しながら、映画等の種々の例との比較を行い、その有効性を確かめた。

(v) 耳元で語りかける: 直交 2 軸スピーカーアレイによるスポット状局所的高音圧領域の形成(公募研究: 溝口班)

直交する 2 軸のスピーカーアレイによるスポット状高音圧分布の生成に取り組み、2 軸スピーカーアレイのシミュレータと実システムとを開発した。この結果、実システムによるスポット状高音圧分布生成、すなわち「耳元で語りかける」効果の実現に成功した。

(vi) 動作により気持ちを伝えられるロボット: 動作により情感伝達が可能なヒト型ロボットハンドの設計(公募研究: 星野班)

ヒトの手指と同じ大きさ、同じ重さの 5 指ロボットハンドを設計した。同ハンドは拇指以外の 4 指間の開閉(アブダクション)が可能であり、また拇指と他指(たとえば小指)で輪を作ることも可能であるため、手話動作や、指先が細やかに動く舞踊動作が生成できる。さまざまな動作により伝達される印象や情感のデータベースを構築し、同知識に従って情感の籠もった動作をロボットハンドに生成させる。

3.4 人と交わる

(i) しっくりとしたコミュニケーションの実現(非拘束なマルチモーダル対話の実現): 手で自由にコンピュータを操作できる拡張机型インタフェースシステムの開発(計画研究: 松山班)

人の手・指の動作を非接触かつ実時間で安定に計測するための技術として、カルマンフィルタによる複数指先位置予測とフレーム間対応付けを用いた両手複数指先軌跡計測法を開発した [Oka 02]。提案手法の特徴としては、赤外線画像の利用と、指先の幾何形状に注目した高速な指先位置検出が挙げられる。通常の可視領域光を検出するカメラではなく中赤外領域(5 μm 程度の波長領域)に感度を持つカメラを利用することにより、特定の温度領域からの放射光を検出することができる。その結果、人体表面における反射光による影響を受けずにユーザの肌領域を安定に抽出することが可能となる。これは、机上に様々な電子情報を投影しながらインタラクションを行う拡張机型インタフェースなどにおいてユーザ行動を計測・理解する際には特に有効である。このようにして得られた赤外線画像に対して正規化相関にもとづくテンプレートマッチング等の処理により指先位置を検出すると同時にモルフロジー処理により手のひらの中心位置を検出する。

さらに、各フレームで検出された指先位置について、フレーム間の指先位置の対応関係を求める。具体的には、各画像フレームの指先位置を検出する際に直後のフレームでの各指先の位置をカルマンフィルタにより予測する。ここでは、各指先位置の運動を仮定し基本システムを定式化している。さらに、各フレームにおいて検出された指先位置を観測値として入力する。また、各フレームにおける指先検出の際の評価値をもとにして観測誤差を推定し、カルマンゲインの計算に用いている。次に、このようにして予測される各指先位置と直後のフレームで実

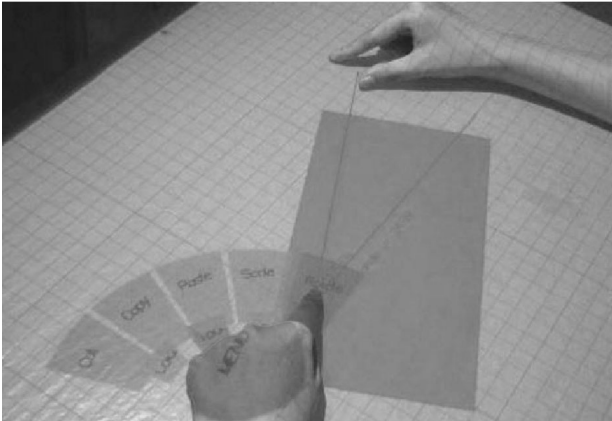


図 7 両手による描画作業の評価

際に検出された指先位置との対応関係を求めることにより、安定かつ高速に複数の指先位置の軌跡を計測することを可能としている。

さらに、この指先動作計測手法を、机上に様々な情報をプロジェクタで投影しながら人がコンピュータとインタラクションを行う拡張機型インタフェースに適用している。身体動作によるコンピュータアプリケーションの直接操作を考えた場合、従来のマウスなどの入力デバイスを用いた方法と比較していくつかの重要な違いが存在する。その大きなものとして両手動作の持つ自由度の高さがある。マウス、スタイラスペン、タッチパネルなどの入力デバイスによる操作は1点のポインティング動作が基本であり、2つの自由度に限られている。そのため、図形の回転動作や拡大・縮小など、より多くの自由度を必要とする操作を実現するためには、補助的なボタンやメニューなどの操作を含めた2つ以上の操作ステップが必要となってしまう。このようなことから、両手動作の高い自由度を有効に用いることは、さまざまな作業をより効率良くかつ自然に行なう操作環境を実現するために重要であると考えられる。

このような観点から、ユーザの身体動作計測にもとづくインタラクションにおいて、両手による直接操作を効率的に利用するための方法について検討した [Koike 02]。特に本研究では、2次元図形の描画作業における両手の役割の違いに注目し、利き手とは反対の手（現在のシステムでは左手）によってパイメニューと呼ぶ円周上に配置された階層メニューを操作しながら、両手もしくは片手で描画することを可能とするツールを実装している。このツールを利用して、マウス入力+ CRT 画面表示、マウス入力+ 机上投影表示、両手による直接入力+ 机上投影表示のそれぞれについて作業実行時間を比較した。この結果、ほとんどの場合で両手による直接操作が最も短時間で作業を実行できることを実験的に示した (図 7)。

(ii) 指コン・ロボット: 指示動作に基づく人間とロボットとの動的なインタラクション (計画研究: 松山班)

身体動作を介した人間とロボットとのインタラクシ

ョンの実現を目指し、指示動作によって人間の意図通りにロボットを誘導するインタラクションモデルを検討し、人間の指示動作によるロボット誘導システムを構築した。

(iii) 人にあわせて対話する: 音声対話におけるユーザモデルを用いた適応的応答の生成 (公募研究: 河原班)

音声対話インタフェースにおいて、ユーザのモデルを知覚し、それに応じて適応的に応答を生成する枠組みを研究した。具体的には、システムに対する習熟度、タスクドメインに関する知識レベル、性急度の3つのユーザモデルを導入し、京都市バス運行情報案内システムにおいて有効性を確認した。

(iv) 人に合わせて音声で会話するシステム: 訂正発話の検出に基づく対話の誤解解消 (公募研究: 北岡班)

人対システムの対話におけるシステムの誤認識に対し、人間が自然に行う繰り返し発話による訂正を検出することにより、システムが自身の誤解状態を検出し、また認識にフィードバックして誤解解消に役立てることができた。

(v) 複数の話者と同時に交わる: 聴覚と視覚情報を複数のレベルで実時間統合による複数話者追跡 (公募研究: 奥乃班)

自然で柔軟なコミュニケーションを実現するには、環境知覚が重要であるが聴覚による環境知覚は従来あまり取り込まれてこなかった。本研究では、聴覚と視覚情報を統合した方向通過型フィルタを能動的に使用するアクティブオーディションを実現し、ヒューマノイドロボットで複数話者同時認識の可能性を実証した。

(vi) 力学メディアを介したインタラクション: 皮膚感覚インタフェースの開発 (公募研究: 篠田班)

皮膚感覚のメカニズムや人間の心理・知的活動との関係を解明し、皮膚によるインタフェースの可能性を検討した。

(vii) 仮想連続体を触る: アクティブ多面体によるフィジカルマンマシンインタラクション (公募研究: 鈴森班)

マイクロセンサとアクチュエータを集積することにより、様々な形状や動きを提示するリンク機構 (アクティブ多面体) を開発し、PC 内に構築した多面体との力学、運動学のインタラクションを実現した。

(viii) インタフェースにおける間とは何か?: 「間」の共創メカニズムに関する研究 (公募研究: 三宅班)

「間」はコミュニケーションにおいて共創されるものと捉え、そのメカニズムを解析している。具体的には、同期タッピング課題を用いて、予測的タイミング制御との関係から実験的解析を進めた。その結果、Working Memory 等の脳高次機能に関わるプロセスとそれが関わらない身体化されたプロセスの2種類の過程から間が創出されることが明らかにされた。今後は、間の共創を支援するインタフェースの設計論へ進む予定である。

(ix) 円滑な対話と「間」の取り方: 円滑な対話を形成する「間」の役割 (公募研究: 中村 (敏) 班)

対話を円滑にするために「間」が果たす役割の解明、お

よび、「間」による対人関係の評価を行うことによって、日本文化の根源ともいわれる「間」の感性を備えたインタフェースを設計する。

3.5 人と暮す

(i) 体があることにどんな意味があるのか? : 身体の持つ情報学的意味の解明 (計画研究: 國吉班)

ヒトであれロボットであれ、それが持つ身体の物理的特性によって、それ固有の感覚運動情報の時空間構造が生み出されることが身体性の情報学的意味であるという仮説を検証するため、転がりながら起き上がる運動といった人の複雑な身体動作の動的構造の分析、ニューラルネットワークを用いた複雑な身体動作の学習シミュレータの作成などを行った。

(ii) ヒトはロボットを何と思っているのか: 脳反応計測を用いた幼児によるロボットの認識 (計画研究: 國吉班)

人間が人型ロボットの動作を見たときの脳内活動を光トポグラフィ装置によって計測し、人型ロボットが、人間を写したビデオ映像よりもリアリティを持った「相手」として認知されているという知見を得た。この結果は、人と共生する機械が人と同じようにして動作を行うことによって、人のパートナーとして認知され親しみが持たれることを生理学的に示したことになる。

(iii) 人から学ぶロボット: 人型ロボットの開発 (公募研究: 池内班)

マルチモーダルな知覚を用いて人が行う手作業の様子を観察し、その方法や手順を獲得・学習する人型ロボットを開発した。これによって、言語化が困難だとされてきた、熟練を要する技の電子的記録・保存が可能となり、技の伝承にも役立つものと思われる。

(iv) 暮らすうちに賢く使いやすく: 人と暮らしながら人と環境を知り行動するロボット (公募研究: 久野班)

身体の不自由な人に音声やジェスチャで指示されたものを取ってくるロボットの実現のために、視覚と音声の統合システムを開発した。ロボットは、人に音声で頼まれた指示を実行するうちに、頼まれたものと環境に固定された家具などの物体とを関係付けて環境の配置を知り、頼まれたものをさがす視覚情報処理が次第に効率的になる。

(v) からだの動かし方を学んでゆくロボット: 超多自由度駆動ヒューマノイドの自己感覚モデルの形成とそれを利用した環境適応行動の実現 (公募研究: 水内班)

人のように複雑な身体構造を有するロボットが、実世界で実際に身体を動かして、自身のアクチュエータ指令とそれによる状態の変化の関係を学んでゆくソフトウェアシステムの構築を目指しており、足先位置・姿勢と筋長の実機で試行錯誤により自動的に獲得してゆき、常時学習が行われるシステムを構築した。

(vi) すべての人に快適に: すべての人を見るロボット車椅子 (公募研究: 久野班)

車椅子はユーザに使いやすいだけでなく、回りの歩行

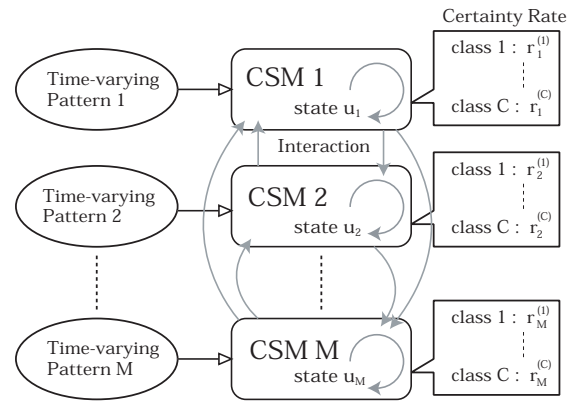


図 8 マルチメディア動的イベント認識システム

者や介護者にも負担にならないものでなければならない。そこで、ビデオカメラと距離センサで周囲を見ることにより、これらの人の行動に応じて動く車椅子を開発した。

(vii) 人の体験と記憶を支援する: 行動や体験の記録にもとづく時空間表現メディアの研究 (公募研究: 間瀬班)

人間と行動をともにする人工物パートナーであるぬいぐるみ型ロボットとの共同体験や相互作用を通じて体験を記録し、記憶想起などを支援する方法について研究を行った。

3.6 自律ダイナミクスを持つ情報システムによる動的イベントの表現・学習・認識・生成

(i) 自律ダイナミクスを持つ情報システムによるマルチメディア動的イベントの認識 (計画研究: 松山班)

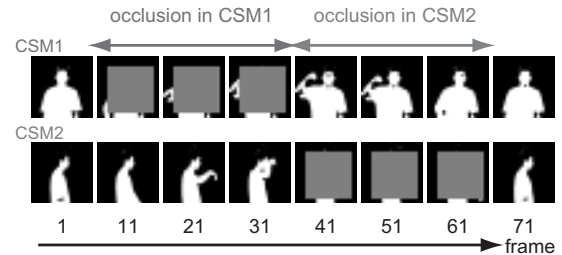
自律ダイナミクスを持ったシステムが、観測された時系列データと共鳴・同期することによって認識を行うことができる動的イベント認識モデル (Continuous State Machine, CSM) を考案した [川嶋 02, Kawashima 02]。認識時には、あらかじめ統計的に学習したダイナミクスに従って自律的に状態を遷移することが可能であり、外部からの入力の変化が、システム内部のダイナミクスと同様の変化である場合には、システムの状態は学習時に連続状態空間中に形成された軌道アトラクタに沿って遷移していく。この軌道アトラクタを、識別するクラスの数だけ状態空間中に形成しておき、認識時には、システムの状態がどのアトラクタに沿って遷移したかを観測することにより、入力パターンの識別結果を得ることができる。さらに、実際のセンサから得られる特徴量はノイズなどの影響によって信頼性が一時的に低くなることもある。そこで、あらかじめ学習したダイナミクスと外部からの入力のどちらを重視するかを、入力の予測誤差の分散に基づいて常に変化させることで、ノイズが多い区間では自律的に状態を遷移させ、入力パターンの時間的伸縮が起こった場合には外部からの入力を重視するといった、頑健な時系列パターン認識が可能となる。

次に、異なったモダリティの観測データをそれぞれ個

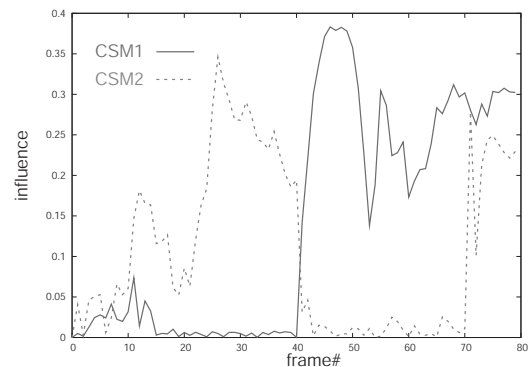
別に認識するように、上述したダイナミカルシステムを単一メディア認識モジュールとして用いる(図8)。そして、これらモジュール間で動的に情報交換を行うための相互作用を導入することによって、全体として安定なマルチメディア動的イベント認識システムを実現し、手話の認識やリップリーディング実験によってその有効性を実証した。マルチモーダルな情報を統合した認識を行う際には、リップリーディングにおける発話内容と口の動きなどのように、同時に起りうる状態の共起性に注目する必要がある。そこで、あらかじめ学習したモジュール間の状態の共起性に基づいて、認識時に互いに相手の状態を推定し、状態レベルで情報を統合する機能を実現する相互作用を導入した。さらに、実際のセンサから得られる各モダリティの信頼性は時間的に変化していて、どの特徴量の信頼性が相対的に高いとは一概に言えないという問題に対して、各モジュールが他のモジュールに与える影響力を、識別の信頼性に依りて動的に制御することで対処し、識別に重要な情報への動的なフォーカシングを実現した。具体的には、認識時に遮蔽やノイズによって識別に重要な入力情報が欠落したモジュールでは、特定の軌道に沿って状態遷移を続けていくことが困難になることに着目し、各々のモジュールの状態が特定の軌道に近づこうとする強さを重みとすることで、各モジュールの影響力の制御した。

ここで提案しているモジュール間の相互作用および動的なフォーカシングの有効性を、不完全な入力データに対する認識の頑健性の観点から検証するために、2視点カメラによる手話の識別によって評価を行った。まず、被験者の正面と側面に設置されたカメラを用いて各フレームで同期させて撮影を行い、各フレームで得られる上体のシルエットを特徴量とした。すなわち、正面カメラと側面カメラでそれぞれ特徴系列が得られることになる。そこで、2つのカメラから得られた特徴系列をそれぞれ異なるモジュールへの入力とし、3人の被験者による7単語の手話を用いて、各々のモジュール内ダイナミクスおよびモジュール間の共起性を学習した。次に、いずれのモジュールへの入力も不完全な情報となった場合の評価を行うために、図9(a)に示すように、前半(フレーム11から中間まで)には正面カメラによって得られるシルエット特徴系列に人工的に遮蔽を加え、後半(中間から終了の10フレーム前まで)には側面カメラより得られるシルエット特徴系列に同様の遮蔽を加えた。これを入力してシステムを動作させたときの、各モジュールの影響力の変化を図9(b)のグラフ(横軸がフレーム番号、縦軸が各モジュールの影響力)に示す。グラフから、入力する系列に対する遮蔽の加え方の変化に対応して、モジュールの影響力が制御されていることが分かる。

信頼性の低いモジュールを動的に抑制し、識別に重要な情報を持つモジュールの影響力を高めることで、他のモジュールの情報を補完し、システム全体として頑健に



(a) Input Sequences



(b) Influence of consisting modules

図9 多視点手話認識において入力に人工的な遮蔽を加えたときの動的フォーカシング

動作することが可能であり、単一のモジュールで識別を行う場合に比べて認識率が高くなることを確認した。

(ii) アトラクタに基づいてロボットの行動を設計する: 力学的情報処理系による行動と言語の相互発達の実現(公募研究: 岡田班)

複数の非線形ダイナミクスを状態空間に配置することによって構造を持った力学系を設計し、状態空間における引き込み現象を用いることでロボットの運動を生成するモデルの構築を目指す。

(iii) 体を使った学習の実現: 身体を介した環境との相互作用に基づく感覚運動情報構造の獲得(計画研究: 國吉班)

身体性の本質は、人間やヒューマノイドロボットの持つ身体ダイナミクスが環境と相互作用することによって感覚運動空間が構造化されることであると考え、情報構造の獲得を行う知的システムの構成法を解明することを目的とする。具体的には、非線形リカレントニューラルネットを用いたモデルによって知覚運動情報や文脈情報に内在する情報構造の獲得・利用を目指す。

4. おわりに

本稿では、2001年度より5年間の計画で進められている特定領域研究「ITの深化の基盤を拓く情報学研究」の一環として、第一筆者が柱長を務めているA03柱における研究の目的、研究の概要を述べた。

我々が研究目標としている「人間と共生する情報シス

テム」という考え方は、

- 第 2 章の研究課題 (I) で述べたように、ヒトに関する心理学、生理学的知見に対する情報学的解釈、活用法を開発することが、今後の情報システムでは不可欠であり、そうした研究を基盤とする情報学と従来の人間とは独立に成り立つ計算機科学・工学との本質的違いを明らかにしているといえる。
- 「人にやさしい」、「バリアフリーな」情報社会の構築に大きく寄与するものと考えられ、IT 社会の深化が人間中心の社会の構築を目指したものであることを広く社会にアピールすることができ、A03 柱の研究は社会的にも大きな意義があると考えられる。

また「人間と共生する情報システム」実現のための情報処理機構として提案している「自律ダイナミクスを持つ情報システム」の研究は、Norbert Wiener が提唱したサイバネティクスと、Alan Turing の提案した計算機械とを融合することによって、実世界における動的イベントの表現・学習・認識・生成機構を構築しようとするもので、今後の研究の進展によっては、従来にはない全く新たな情報処理の体系ができるのではないかと考えている。

最後に、誌面の都合で十分詳細な議論はできなかったが、本稿を契機に研究の輪が広がって行けば幸いである。

謝 辞

研究会などを通じて、日頃から議論していただいている A03 柱の班員に感謝する。本研究の一部は、科学研究費補助金 13224051 の補助を受けて行った。

◇ 参 考 文 献 ◇

- [A03-URL] <http://vision.kuee.kyoto-u.ac.jp/Informatics-A03/Main.htm>
- [Davison 98] Davison, A. J. and Murray, D. W.: Mobile Robot Localisation Using Active Vision, *Proc. of ECCV*, Vol. 2, pp. 809–825, 1998.
- [川嶋 02] 川嶋宏彰, 松山隆司: 連続状態モデル間の相互作用に基づく多視点動作認識, 電子情報通信学会論文誌 Vol.J85-D-II, No.12, pp.1801–1812, 2002.
- [Kawashima 02] Kawashima, H. and Matsuyama, T.: Integrated Event Recognition from Multiple Sources, *Proc. of the 16th International Conference on Pattern Recognition*, Vol.2, pp.785–789, 2002.
- [Koike 02] Koike, H., Chen, X., Nakanishi, Y., Oka, K. and Sato, Y.: Two-Handed Drawing on Augmented Desk, *Extended Abstract of 2002 ACM Conf. Human Factors in Computing Systems (SIGCHI 2002)*, pp. 760–761, 2002.
- [松山 02] 松山隆司 編: IT の深化の基盤を拓く情報学研究平成 13 年度研究成果報告書: A03 人間の情報処理の理解とその応用に関する研究, 2002.
- [Matsuyama 02] Matsuyama, T. (ed.): *Proc. of the First International Workshop on Man-Machine Symbiotic Systems*, 2002.
- [松山 03] 松山隆司 編: IT の深化の基盤を拓く情報学研究平成 14 年度研究成果報告書: A03 人間の情報処理の理解とその応用に関する研究, 2003.
- [Molton 00] Molton, N. and Brady, M.: Practical Structure and Motion from Stereo When Motion is Unconstrained, *Int. J. of Computer Vision* Vol. 39, 1, pp. 5–23, 2000.
- [Murray 96] Murray, D. W., Reid, I. D. and Davison, A. J.: Steering and Navigation Behaviours Using Fixation, *Proc. of British Machine Vision Conference*, 1996.
- [Oka 02] Oka, K., Sato, Y. and Koike, H.: Real-Time Fingertip Tracking and Gesture Recognition, *IEEE Computer Graphics and Applications*, Vol. 22, No. 6, pp. 64–71, 2002.
- [Sugimoto 03a] Sugimoto, A. and Ikeada, T.: Diverging Viewing-Lines in Binocular Vision: A Method for Estimating Ego Motion by Mounted Active Cameras, *Proc. of Australia-Japan Advanced Workshop on Computer Vision*, pp. 126–133, 2003.
- [Sugimoto 03b] Sugimoto, A. and Matsuyama, T.: Active Wearable Vision Sensor: Detecting Person's Blink Points and Estimating Human Motion Trajectory, *Proc. of IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 539–545, 2003.
- [杉本 02] 杉本晃宏, 長友渉, 松山隆司: 装着型能動視覚センサを用いた人物の運動推定, 画像の認識・理解シンポジウム (MIRU2002), Vol. I, pp. I-215–I-222, 2002.
- [Sugimoto 02] Sugimoto, A., Nakayama, A. and Matsuyama, T.: Detecting a Gazing Region by Visual Direction and Stereo Cameras, *Proc. of the 16th International Conference on Pattern Recognition*, Vol. III, pp. 278–282, 2002.
- [Turk 00] Turk, M. and Robertson, G.: Perceptual User Interfaces, *Communications of ACM*, Vol.43, No.3, pp.33–34, 2000.

著 者 紹 介

松山 隆司 (正会員)

1976 年京大大学院修士課程修了。京大助手、東北大助教授、岡山大教授を経て 1995 年より京大大学院電子通信工学専攻教授。現在同大学院情報学研究科知能情報学専攻教授。2002 年学術情報メディアセンター長、評議員。工博。画像理解、人工知能、分散協調視覚の研究に従事。1980 年情報処理学会創立 20 周年記念論文賞、1990 年人工知能学会論文賞、1993 年情報処理学会論文賞、1994 年電子情報通信学会論文賞、1995 年第 5 回国際コンピュータビジョン会議 Marr Prize、1996 年国際パターン認識連合 Fellow、1999 年電子情報通信学会論文賞、2000 年画像センシングシンポジウム優秀論文賞。

杉本 晃宏

1987 年東京大学工学部計数工学科卒業。1989 年同大学院工学系研究科修士課程修了 (数理工学専攻)。同年、日立製作所基礎研究所に入社。1991 年～95 年、ATR に出向。1999 年京都大学講師、2002 年国立情報学研究所助教授、現在に至る。博士 (工学)。視覚情報処理や離散システム・アルゴリズムなどに興味をもち、数理的手法に基づいたコンピュータビジョンの研究に従事。2001 年情報処理学会論文賞。

佐藤 洋一

1990 年東京大学工学部機械工学科卒業。1991 年よりカーネギーメロン大学計算機科学部ロボティクス学科博士課程に在籍。1997 年 Ph.D. in Robotics 修了。同年より東京大学生産技術研究所研究機関研究員、講師を経て、現在同研究所助教授。コンピュータビジョン、HCI、およびコンピュータグラフィックスに関する研究に従事。平成 11 年山下記念研究賞、MIRU2000 最優秀論文賞、平成 11 年度日本 VR 学会論文誌論文賞等を受賞。電子情報通信学会、情報処理学会、日本 VR 学会、ACM、IEEE 各会員。

川嶋 宏彰

1999 年京都大学工学部電気電子工学卒。2001 年同大学院修士課程了 (情報学)。2002 年同大学院博士後期課程中退し、同大学院情報学研究科知能情報学専攻助手。時系列パターン認識、マンマシンインタラクションに関する研究に従事。