

機関番号：12601

研究種目：特定領域研究

研究期間：2006～2010

課題番号：18049016

研究課題名（和文） セマンティック・ヒューマノイドインタラクション

研究課題名（英文） Semantic Humanoid Interaction

研究代表者

國吉 康夫 (KUNIYOSHI YASUO)

東京大学・大学院情報理工学系研究科・教授

研究者番号：10333444

研究成果の概要（和文）：

本研究では人間とロボットが実世界で臨機応変に意味レベルのインタラクションを行う技術の構築を目指し、人間の振る舞いや周囲環境の様子を認識・理解し、その情報をライブログやサイバースペース等から得られる膨大な知識情報と統合・解釈し、状況に即した対話や支援などを直観的な形で人間に提供する機能の研究を行った。これらを統合した代表的なシステムとして、日常環境中を自律的に動き回り、人々の興味を喚起する出来事を発見、認識、理解し、インタビューなども行って、ニュースブログのような分かり易い形にまとめて情報発信する“ジャーナリストロボット”を、世界で初めて提案し実現した。

研究成果の概要（英文）：

In this research, we focus on the development of the semantic interaction between the human and the robot in the real world according to circumstances. Our research consists of 1) the recognition of human behavior and environment, 2) the knowledge acquisition from large-scale information in the cyber space and the lilelog, and 3) the intuitive interaction and support for human according to circumstances. As the representative example by integrating those functions, we proposed and developed the world's first "Journalist Robot" that explores the daily-life environment autonomously, finds and understands news-like events, and publishes them to the web by summarizing the events.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2006年度	20,900,000	0	20,900,000
2007年度	31,400,000	0	31,400,000
2008年度	24,300,000	0	24,300,000
2009年度	32,650,000	0	32,650,000
2010年度	17,400,000	0	17,400,000
総計	126,650,000	0	126,650,000

研究分野：知能システム情報学

科研費の分科・細目：情報学・知能情報学

キーワード：ロボット，行動，意味，認識，学習，自然言語

1. 研究開始当初の背景

情報システムの多機能化，高機能化によってもたらされる複雑性を克服し，爆発する情報量を有効に活用し，安心・快適な情報環境

を構築するには，人間と情報システムとの間に存在する様々なレベルでのギャップを，ロボットによる身体インタラクションを含めたマルチモーダル・インタラクションによつ

て解消することが必要である。

2. 研究の目的

本研究では、人間とロボットが、実世界の中で臨機応変に意味レベルのインタラクションを達成することを目指す。すなわち、日常生活や仕事場の環境の中で、人間が様々な物を扱いつつ行動するとき、人間の振る舞いや周囲環境の様子を認識・理解しつつ、その情報をライブログやサイバースペース等から得られる膨大な知識情報と統合・解釈し、状況に即した手助け行動や対話など、直観的かつ実効的な形で人間への知的支援を行う機能の構築を目指す。これによって、実世界事象と知識情報の双方向融合を実現するための理論基盤を構築するとともに、爆発する膨大な情報の蓄積を、極めて具体的かつ効果的に人間生活に役立て、多様な場面で人間を知的にサポートするロボットに向けた技術を開発する。

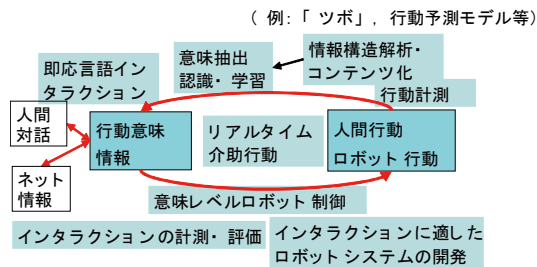


図1: セマンティック・ヒューマノイドインタラクションの構造

3. 研究の方法

本研究課題の研究項目を以下にあげる。

- (1) 基礎データとしての人間行動の計測と情報構造解析ならびにコンテンツ化と蓄積
- (2) 人間行動および環境中の物体・事象からの意味情報抽出
- (3) 実世界意味情報と事前に蓄積された知識情報との統合解釈
- (4) 統合解釈を実世界状況に反映するための意味レベル行動生成
- (5) 意味情報抽出、蓄積情報との統合解釈、意味レベル行動生成を融合、一体化した意味レベル相互作用の理論とシステム
- (6) 意味レベル相互作用を知的人間支援行動として具現化するロボットシステムや言語インタラクション

本研究の全体構成と上記各課題の関係を図1に示した。研究期間の前半 (H18~H19) では、前述の課題のうち要素的部分である(1)~(4)の研究を推進し、中盤 (H20~H21) ではそれらの高度化と共に意味レベル相互作用 (課題(5))、最終年度は統合化・システ

ム化 (課題(6)) に取り組んだ。当初計画の想定を超えた新たな知見と展開も得られ、これに伴いさらに強力な成果を得るべく研究を加速して行った。

4. 研究成果

(1) 基礎データとしての人間行動の計測と情報構造解析ならびにコンテンツ化と蓄積

人間支援応用と人間行動モデル構築のために重要な (マクロとミクロの) 二つの側面として、家庭内で家具や日用品を扱いながら行う日常行動全体の計測技術・データ化と、熟練作業技能の精密な計測・解析・モデル化がある。前者に関して、複数TVカメラによるマーカーレス実時間動作認識手法を高度化して家具による隠ぺいや照明変化に影響されない性能を達成し、開発した動作認識手法と統合して日常生活における人間行動の意味理解機能が達成した。また、後者に関しては、触覚センサと運動計測を統合したウェアラブルモーションキャプチャを独自開発し、単なる運動ではなく身体と物体の接触を統合して行動に意味づけを行う手法を構築した。また、打撃等を含む高速な技能動作のコツの解析も進めた。

ロボットによる生活支援を行う際に有用な情報である人体の姿勢を追跡する手法として、多視点動画像からのボクセルベース姿勢推定に取り組んだ。人体は多リンクの複雑な構造からなり、基本的に常に動いているため、多視点で撮影して得られるボクセルデータは時空間的に膨大な情報となる。複数の人の行動を把握しようとする、他の人や家具などに重なることで頻繁に隠れてしまうため、グラフィカルモデルにもとづく高次時系列フィルタリング、時空間ラベリングによるマルチターゲット追跡を中核とした手法を実装し対応した。このアプローチの初期プロセスにあたる人体形状復元では、背景差分によるシルエット抽出と視体積交差を用いる。人のシルエット抽出にグラフカットアルゴリズムによるエネルギー最小化を用いることで、複雑背景に起因するシルエットの欠損をおさえる、影の影響に対して頑健である、人以外の動的な物体 (ロボット等) が存在しても人領域のみを復元推定できるという特長を得ることができた。さらにこの方法を拡張し、カメラ隠蔽や背景変化等に脆弱である問題に対処するため、背景差分によるシルエット抽出に加えてカメラ画像間のステレオマッチングも考慮する三次元グラフカットアルゴリズムを定式化した。その結果、特に家具などによる隠蔽が発生する状況や照明変化に対して優れた性能を発揮することを

確認した。図 2 に複数カメラから得られた人の姿勢推定結果を示す。

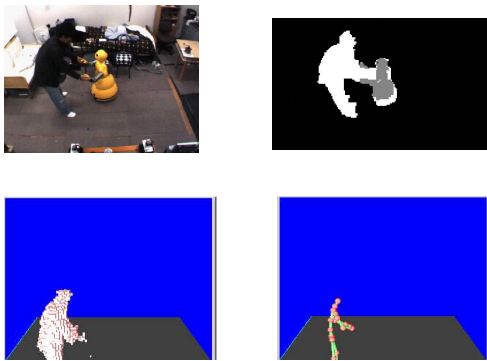


図 2: 複数カメラによる人の姿勢推定

さらに高速運動時に無視できなくなる身体の内部自由度の運動を調べるために加速度を面としてとらえ全体の特徴を求める「加速度センサーアレイ」を開発し、和太鼓の打撃においてプロ演奏者を含む熟練者と中級者の打撃を、片腕に 12 個の 3 軸加速度センサを装着して測定を行った (図 3 参照)。打撃後 200ms の腕の表面の振動を主成分分析によって分析したところ、上腕において熟練者の方が振動モードが多かった。これは、中級者の腕が緊張している一方で熟練者では脱力が起きていることが示唆された。

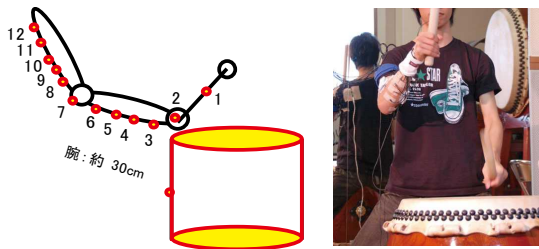


図 3: 和太鼓の打撃の解析

(2) 人間行動および環境中の物体・事象からの意味情報抽出

日々の生活の中で触れる実世界情報を常時記録し、自動的に解析し活用する機能は、ロボットの実世界進出に必要不可欠な技術である。ロボット以外にも、高齢者の記憶支援など幅広い産業応用上の成果を得ることが期待される。

その中核的技術として実世界画像中の事物を単語と対応づける、画像アノテーション・リトリバルがある。しかし任意の画像を扱うのは極めて難しく、既存技術は精度・速度の両面において難を抱える。本研究では、高次局所自己相関特徴と確率的正準相関分析の組み合わせにより、画像・単語間の概念を効率よく学習することで、2007 年開発当時、従来の世界最高精度の手法に比し、精度・速

度の両面で圧倒的な向上を実現した。

この手法を基礎として、視覚情報から環境中の事物を柔軟に認識・検索を行える AI ゴーグルを提案した (図 4 参照)。カメラを備えたゴーグルとタブレット型 PC からなるシステムである。装着者の視界と同期したカメラ画像を取得し、独自の超高速画像アノテーション手法によりリアルタイム処理し、その結果を HMD (Head Mount Display) に出力する。同時に、この結果と取得画像を合わせて視覚ログに随時保存する。また、検索の要求があった場合、視覚ログから該当する画像を選び、HMD へ出力する。これにより限定的な計算資源においてもリアルタイムかつ高精度に実世界シンボル化が行えることを示した。

さらに、移動体に装備されたカメラ画像から、通常の状態からの逸脱として定義される異常の有無を判断し、異常領域を特定することに取り組んだ。



図 4: AI Goggles システム

(3) 実世界意味情報と事前に蓄積された知識情報との統合解釈

汎用的な一般物体認識の実現のためには、膨大な数の対象と画像のアプリアランスを学習する必要がある。人手によって学習過程を管理することは難しい。このため、Web 上の大量の画像を用い自律的に画像知識の獲得を行う方法が近年検討されている。そこで、Web からの完全自律的な画像知識獲得へ向けた、高速画像アノテーション・リトリバル手法を提案した。本手法は、複数ラベルによる画像のコンテキストを用い、高速に学習・認識を行うことが可能である。実験では、270 万枚の Flickr 画像から学習を行い、本手法の有効性を検証した。この膨大な情報からの自律的画像知識の獲得と柔軟な認識能力の実現は、ジャーナリストロボットの実世界記述機能やライフログの言語要約、視覚記憶の言葉による検索などの将来の人間支援につながるものである。

(4)統合解釈を実世界状況に反映するための意味レベル行動生成

本課題は課題(6)に統合された。具体的には、統合ロボットシステムがニュース取材行動として課題(2)の異常検出手法に基づいて対象に接近する行動や、インタビューのために現場付近の人間を発見し接近し話しかける行動など、状況に応じて行動生成する機能として構築した。

(5)意味情報抽出、蓄積情報との統合解釈、意味レベル行動生成を融合、一体化した意味レベル相互作用の理論とシステム

ヒトやヒューマノイドロボットにおける全身の身体動作を表現する際、関節数の分の次元数をもつベクトルの時系列情報が必要となり膨大な情報量となる。これを基本的な動作の合成として表現することで、情報量の圧縮が可能となる。従来手法では内挿による合成が主なアプローチであったが、本研究では内挿・外挿の双方を対象とした、基本動作の合成による新規動作の生成、および、未知動作の基本動作への分解、を可能とするアルゴリズムの開発を行った。具体的には身体動作をHMMで抽象化し、各動作のHMM間の距離をバタチャリア距離によって計測し、距離情報を多次元尺度法によって処理することで空間内の配置を決定する。この空間表現により、ロボットが人間の動作提示に基づいて動作を模倣学習する際に必要なパラメータ表現が少なく済むようになった。この他にも、人間の初心者動作を習得する際に、ロボットが必要となる動作の箇所を強調するというアプリケーションの構築が可能となり、人間とヒューマノイドロボットの双方にとって、有効な身体動作表現となったことを確認した。

ヒューマノイドロボットとユーザが言語インタラクション、および身体的物理インタラクションを行う系の設計をスムーズに評価・設計するための支援ツールの構築を行った。ユーザの日常生活活動を支援するためのロボットの状況判断・行動計画システムを設計・評価するためには、実機ロボットにおけるユーザとのインタラクションを大量に行う必要があり、評価フェーズの負荷が高くなる。そこで、身体的物理シミュレーションと社会的コミュニケーションのシミュレーションを統合可能な仮想環境を開発し、その中にロボットを実装することで、効率的に人間との共同作業を評価するシステムを実現した。具体的には、図5に示すように、人間とロボットが協調して料理を行うタスクを

対象とし、タスク遂行の所要時間を通じてロボットの行動認識・行動計画モデルの妥当性を評価した。



図5:人間とロボットとのインタラクションのシミュレータ

(6)意味レベル相互作用を知的人間支援行動として具現化するロボットシステムや言語インタラクション

インターネットの普及により、世界中で起きている事象に対していつでもどこでも容易にその情報を得ることが可能となったが、自分の周りに生じている事象は、最も身近なことであるにも関わらず、情報を得るのが困難であり見過ごされる場合が多い。そこで、実世界で生じるイベントを自律的に発見しニュースブログのような分かり易い形にまとめて情報発信するジャーナリストロボットを提案し実現した。はじめに、実世界の事象・行動の認識と大規模知識情報とを統合して知的対人インタラクションを生成する機能の構築に向け、ジャーナリストロボットとして共通プラットフォーム上でのシステム化に取り組んだ(図6参照)。その上で、意味レベルインタラクション機能として、実世界を自律的に探索し、重要な意味を持つ異変を画像認識と知識情報を統合して認識・解釈し、インタビューも自律的に行い、それらの結果を自然言語処理と統合してニュース記事として生成、配信するシステムを構築した。



図6:ジャーナリストロボットシステム

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 42 件)

- ① 國吉康夫. ロボティック・サイエンスとは何か. 日本ロボット学会誌, Vol. 28, No. 4, pp. 370-374, 2010.
- ② Tatsuya Harada, Hideki Nakayama, Yasuo Kuniyoshi, and Nobuyuki Otsu. Image Annotation and Retrieval for Weakly Labeled Images Using Conceptual Learning. *New Generation Computing*, Vol. 28, No. 3, pp. 227-298, 2010.
- ③ 中山英樹, 原田達也, 國吉康夫. 大規模 Web 画像のための画像アノテーション・リトリバル手法. *電子情報通信学会論文誌*, J93-D, 8, pp. 1267-1280, 2010.
- ④ Hiroshi Noguchi, Taketoshi Mori, Takashi Matsumoto, Masamichi Shimosaka, and Tomomasa Sato. Multiple-Person Tracking by Multiple Cameras and Laser Range Scanners in Indoor Environments. *Journal of Robotics and Mechatronics*, Vol. 22, No. 2, pp. 221-229, 2010.
- ⑤ 稲邑哲也, 柴田智広. 動作パターンとシンボルを相互変換する原始シンボル空間における動作パターンの内挿・外挿. *日本ロボット学会誌*, Vol. 28, No. 4, pp. 512-521, 2010.
- ⑥ Saifuddin Md. Tareeq, and Tetsunari Inamura. Rapid Behavior Adaptation for Human-centered Robots in Dynamic Environment based on Integration of Primitive Confidences on Multi-sensor Elements. *Journal of Artificial Life and Robotics*, Vol. 15, No. 4, pp. 515-521, 2010.
- ⑦ Kohei Matsumura, Tomoyuki Yamamoto, and Tsutomu Fujinami. The role of body movement in learning to play the shaker to a samba rhythm. *Research Studies in Music Education*, to appear, 2011.
- ⑧ Yasuo Kuniyoshi. Autonomous Adaptive Emergent Systems. *Advanced Robotics*. Vol. 23, No. 11, pp. 1481-1485, 2009.
- ⑨ Yasuo Kuniyoshi, and Masayuki Numao. Academic Roadmap in Integrated Information Field. *Advanced Robotics*, Vol. 23, No. 11, pp. 1465-1474, 2009.
- ⑩ 原田達也, 金崎朝子, 國吉康夫. 三次

元環境地図からの物体探索タスク応用を目指したカラー立体高次局所自己相関特徴の開発. *日本ロボット学会誌*, Vol. 27, No. 7, pp. 749-758, 2009.

- ⑪ 森武俊. 人の行動を理解し動作を認識するビジョン技術. *日本ロボット学会誌*, Vol. 27, No. 6, pp. 27-31, 2009.
- ⑫ Tetsunari Inamura, Kei Okada, Satoru Tokutsu, Naotaka Hatao, Masayuki Inaba, and Hirochika Inoue. HRP-2W: A Humanoid Platform for Research on Support Behavior in Daily Life Environments. *Journal of Robotics and Autonomous Systems*, Vol. 57, No. 2, 2009.
- ⑬ 稲邑哲也, 谷江博昭, 中村仁彦. 幾何学的シンボル操作による多様な動作パターンの認識・生成を実現する原始シンボル空間の構成法. *日本ロボット学会誌*. Vol. 27, No. 5, pp. 84-94, 2009.
- ⑭ Tomoyuki Yamamoto, Tsutomu Fujinami. Hierarchical organization of the coordinative structure of the skill of clay kneading. *Human Movement Science*, Vol. 27, No. 5, pp. 812-822, 2008.
- ⑮ Katsuhiko Ishiguro, Nobuyuki Otsu, Max Lungarella, and Yasuo Kuniyoshi. Detecting direction of causal interactions between dynamically coupled signals. *Physical Review E*, Vol. 77, No. 2, 026216-1 - 026216-6, 2008.
- ⑯ Katsuhiko Ishiguro, Nobuyuki Otsu, Max Lungarella, and Yasuo Kuniyoshi. Comparison of nonlinear Granger causality extensions for low-dimensional systems. *Physical Review E*, Vol. 77, No. 3, 036217-1 - 036217-9, 2008.

他 26 件.

[学会発表] (計 73+128=201 件)

- ① Tatsuya Harada, Hideki Nakayama, and Yasuo Kuniyoshi. Improving Local Descriptors by Embedding Global and Local Spatial Information. *European Conference on Computer Vision (ECCV 2010)*, 2010.
- ② Hideki Nakayama, Tatsuya Harada, and Yasuo Kuniyoshi. Global Gaussian Approach for Scene Categorization Using Information Geometry. *The Twenty-Third IEEE International Conference on Computer Vision and Pattern Recognition (CVPR2010)*, 2010.
- ③ Shigeyuki Odashima, Tomomasa Sato, and Taketoshi Mori. Household object

management via integration of object movement detection from multiple camera. the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010.

- ④ Masamichi Shimosaka, Takahito Ishino, Hiroshi Noguchi, Tomomasa Sato, and Taketoshi Mori. Detecting human activity profiles with Dirichlet enhanced in homogeneous Poisson processes. the Twentieth International Conference on Pattern Recognition, 2010.
- ⑤ Tetsunari Inamura, et al., Simulator platform that enables social interaction simulation --SIGVerse: SocioIntelliGenesis simulator--. IEEE/SICE International Symposium on System Integration, 2010.
- ⑥ Md. Hasanuzzaman, and Tetsunari Inamura. Adaptation to New User Interactively Using Dynamically Calculated Principal Components for User-Specific Human-Robot Interaction. IEEE/SICE International Symposium on System Integration, 2010.
- ⑦ Kohei Matsumura, and Tomoyuki Yamamoto. A Method for Analyzing Fast Rotational Movement using Multiple 6-axial Accelerometers HBS (Human Behavior Sensing), 2010.
- ⑧ Tatsuya Harada, Hideki Nakayama, and Yasuo Kuniyoshi. Image Annotation and Retrieval Based on Efficient Learning of Contextual Latent Space. 2009 IEEE International Conference on Multimedia and Expo, 2009.
- ⑨ Takashi Shibuya, Tatsuya Harada, and Yasuo Kuniyoshi. Causality Quantification and Its Applications: Structuring and Modeling of Multivariate Time Series. The 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2009.

他 192 件.

[図書] (計 7 件)

- ① 浅田稔, 國吉康夫. ロボットインテリジェンス, 岩波書店, 2006.

他 6 件.

[産業財産権]

○出願状況 (計 8 件)

名称: 特徴量生成装置、特徴量生成方法および特徴量生成プログラム、ならびにクラス判

別装置、クラス判別方法およびクラス判別プログラム

発明者: 原田達也, 中山英樹, 國吉康夫

権利者: 原田達也, 中山英樹, 國吉康夫

種類: 特許

番号: 特願 2009-121244

出願年月日: 2009 年 5 月 19 日

国内外の別: 国内

他 7 件.

○取得状況 (計 0 件)

名称:

発明者:

権利者:

種類:

番号:

取得年月日:

国内外の別:

[その他]

ホームページ等

<http://www.isi.imi.i.u-tokyo.ac.jp/>

6. 研究組織

(1) 研究代表者

國吉 康夫 (KUNIYOSHI YASUO)

東京大学・大学院情報理工学系研究科・

教授

研究者番号: 10333444

(2) 研究分担者

森 武俊 (MORI TAKETOSHI)

東京大学・大学院医学系研究科・

特任准教授

研究者番号: 20272586

稲邑 哲也 (INAMURA TETSUNARI)

国立情報学研究所・情報学プリンシプル研

究系・准教授

研究者番号: 20361545

山本 知幸 (YAMAMOTO TOMOYUKI)

大阪大学・大学院基礎工学研究科・特任准

教授

研究者番号: 60345113

原田 達也 (HARADA TATSUYA)

東京大学・大学院情報理工学系研究科・准

教授

研究者番号: 90345653

長久保 晶彦 (NAGAKUBO AKIHIKO)

独立行政法人産業技術総合研究所・研究員

研究者番号: 00357617

(3) 連携研究者

()

研究者番号: