

環境知能機構グループ - 環境知能機構の研究 - 成果報告

研究組織:

環境知能学研究室	萩田紀博教授	(平成21年度～平成25年度)
視覚情報メディア研究室	横矢直和教授	(平成21年度～平成25年度)
像情報処理学研究室	千原國宏教授	(平成21年度～平成22年度)
音情報処理学研究室	鹿野清宏教授	(平成21年度～平成24年度)
知能コミュニケーション研究室	中村哲教授	(平成23年度～平成25年度)

研究成果の概要 (和文):

生活空間に溶け込んだ IT によって利用者が意識することなくサポートされる「アンビエント環境知能」においては、利用者を含む環境の状況理解を実現し、ユーザに対する適切なサポートを行うことが重要となる。本研究グループでは、環境内に固定配置された大量のカメラ・センサ群および移動体搭載センサから得られる情報を用いて、広域での移動物体（人物、ロボット、車両など）の連続的な位置計測、環境中で発生するイベントの自動検出、時空間データベース化および、騒音環境下などモバイル環境での音声認識に基づく情報検索を実現する手法を開発する。また、効果的なユーザ支援方法について検討する。本事業では、(1)ロボットと人のインタラクションにおける親しみやすさの検証、(2)画像群を用いたシーンの三次元計測手法の開発、(3)拡張現実感技術を用いたユーザ支援システムの作成、(4)音声認識技術を用いた音声対話インタフェースによる親しみやすさの検証、について、研究を実施した。具体的には、(1)について、ロボットに適したタスクの調査および人物歩行軌跡の推定による歩行行動の調査および関連技術の開発を行い、(2)について、空撮動画撮影時におけるカメラ位置推定手法の開発、GPS または空撮画像と地上撮影動画画像を併用するカメラ位置・姿勢推定手法の開発、全方位動画像からの三次元情報の推定を行った。また、(3)について、拡張現実感システムによる三次元領域可視化手法、紙面レイアウト整形支援手法および障がい者支援手法を開発し、(4)について、音声認識・合成に関する基礎研究と、音声インタフェースによる自然な対話システムの開発を行った。

研究成果の概要 (英文):

In order to realize ‘ambient intelligence’ that supports users without user’s interactions and attentions, understanding of the real scene is one of the most important problems. The main goal of our research group is development of a robust and accurate method for scene understanding. For this purpose, we have developed a wide range of techniques including position estimation of moving objects (e.g. users, robots and vehicles), event detection in the target environment, database construction of the target scene and information search by voice recognition under noisy environment. We have studied four topics: (1) survey of affinity of robots under robot-human interaction, (2) development of 3D scene reconstruction method by omnidirectional vision, (3) research for user support system with augmented reality, (4) development of voice based amiable interface. Concretely, for (1), we have investigated for tasks, that are suitable for robots regarding affinity, and behavior of humans by estimating route of walking humans. For (2), we have tested for capturing omnidirectional images from the sky using an airship, and we have developed a method for estimating 3D information from video images. Camera parameter estimation method that uses GPS data/aerial images and vision data is also examined. For (3), user support system based on augmented reality technique is tested with applications of 3D region navigation, disabled person support, and user support for document layout organization. For (4), in addition to fundamental researches for voice recognition and synthesis for improving responses of voice navigation system, we have developed voice based interaction system for achieve natural conversation between human and robot.

キーワード: 人位置計測、全方位動画画像解析、空撮、拡張テレプレゼンス、拡張現実感、インタフェース、センサフュージョン、カメラ位置・姿勢推定、三次元復元・計測、ユーザ支援、障がい者支援、音声認識、音声合成、音声対話、音声インタフェース、機械学習、ロボットインタラクション

1. 研究目的

生活空間に溶け込んだ IT によって利用者が意識することなくサポートされる「アンビエント環境知能」においては、利用者を含む環境の状況理解を実現することが重要となる。本研究グループでは、一般環境下における頑健かつ高精度なシーンの状況理解と効果的なユーザ支援を実現することを目的とし、環境内に固定配置された大量のカメラ・センサ群および移動体搭載センサから得られる情報を用いて、広域での移動物体（人物、ロボット、車両など）の連続的な位置計測、環境中で発生するイベントの自動検出、時空間データベース化および、騒音環境下などモバイル環境での音声認識に基づく情報検索を実現する手法を開発する。また、拡張現実に基づくユーザ支援方法について検討する。

2. 研究方法

上記の研究目的を達成するため、以下の 4 つの研究項目に取り組んだ。

- (1) ヒューマン・ロボット・インタラクションに関する研究
- (2) シーンの画像データベース化と三次元環境の画像計測に関する研究
- (3) 拡張現実感によるユーザ支援に関する研究
- (4) 音声認識・音声合成および音声インタフェースに関する研究

以下、各研究項目について詳細を述べる。

3. 研究内容・成果

3.1. ヒューマン・ロボット・インタラクションに関する研究

(街中におけるロボットに向いているタスクの調査研究)

街中で、コミュニケーションを必要とするが、あまり人がやりたがらない傾向のあるタスクを用意し、ロボット、着ぐるみ、人間がタスクを行っている様子を撮影し、それを見た人々がどのように感じるかという調査を試みた。これにより、人間の印象に基づいたロボットに適したタスクの選定を行うことが可能となる。“うんざりするタスク”、“強いストレスを感じるタスク”、“汚い作業を行うタスク”の 3 つのタスクをロボット、着ぐるみ、人間のそれぞれが行っている様子を撮影し、本研究の評価実験で提示するコンテンツ素材を取得した(図 15 参照)。それらを募集した被験者に見せ、印象評価実験を行った結果、3 つのタスク全てにおいて、ロボットが行う方が良いと考えている事が分かった。



図 15 評価実験に用いたコンテンツ素材

(人位置計測システムを利用した親しみやすい歩行行動の研究)

ラインスキャナ式距離センサを利用した高精度人位置推定システムを使用し、時間毎の人位置から歩行者の歩行軌跡を推定し、ロボットが歩行者に親しみを与えるような歩行行動を調査する実験を試みた。ラインスキャナ式距離センサを利用した高精度人位置推定システムを使用し、ロボットがその歩行者に親しみを与えるような歩行行動を行う歩行モデルの調査研究を行った。2秒間の人の動きから直線で歩行者の歩行軌跡を推定し、人間からモデル化した”親しみやすい歩き方”を行う(図 16 参照)。実際の街中で被験者に対し歩行動作実験を実施し、他の歩行行動と比較した結果、もっとも親しみを感じる歩行行動であるという評価を得た。

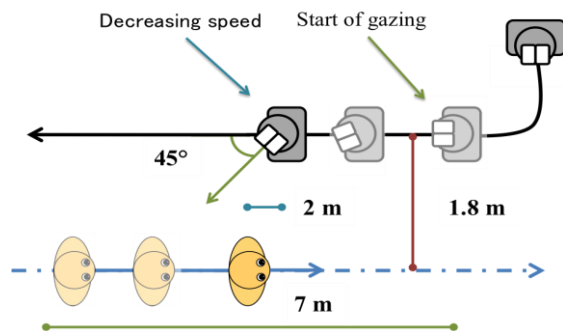


図 16 歩行軌跡の推定

(不可視マーカを利用した位置計測システム)

不可視マーカを用いた位置推定システムを広域屋内環境に導入し、人やロボットの高精度位置姿勢推定が可能な環境を実現した。不可視マーカは、壁紙に近い色の再帰性反射材を固有の ID を表現可能な幾何パターンにとして蒸着させることで作成した。不可視マーカは、壁紙に近い色の再帰性反射材を固有の ID を表現可能な幾何パターンにとして蒸着させることで作成した。図 17 に示す実験環境では、システム天井に不可視マーカが含まれたパネルを設置し、広域における位置姿勢推定を実現した。図 18 に、位置推定システムの例を示す。左図は、バーチャルリアリティなどに用いられるユーザの視点位置依存したディスプレイの例でユーザ位置に実時間計測を行っている。右図はロボットの位置姿勢の推定による、ロボットと人間の協調作業を行っている様子である。また、固有の ID パターンを保有するマーカを再帰性反射材で連続的に印刷する技術の特許出願を行った。



図 17 不可視マーカを設置した様子(左)、フラッシュ撮影した不可視マーカ(右)

(インタラクション機能を持つ掃除ロボットを用いた人とロボットの協調作業)

自律型掃除ロボットが普及しつつあるが、椅子などの障害物がありロボット単体では掃除できない場所があることや、ユーザにとってロボットの掃除した領域が不明であるといった問題点がある。そこで本研究では、人やロボットの位置を実時間計測し、それらの情報を利用可能な環境知能基盤を利用して、インタラクション機能を持つ掃除ロボットが人と協調作業を行うことで、これらの問題を試みた。人やロボットの行動、ロボットが掃除した領域や障害物の位置を実時間で更新する環境マップを作成することで、環境にある障害物の除去を人に依頼することで協調作業を実現した。図 18 に示すような環境を構築し、システムの有効性を示すために実験を行った結果、協調作業が実現できることを明らかにした。

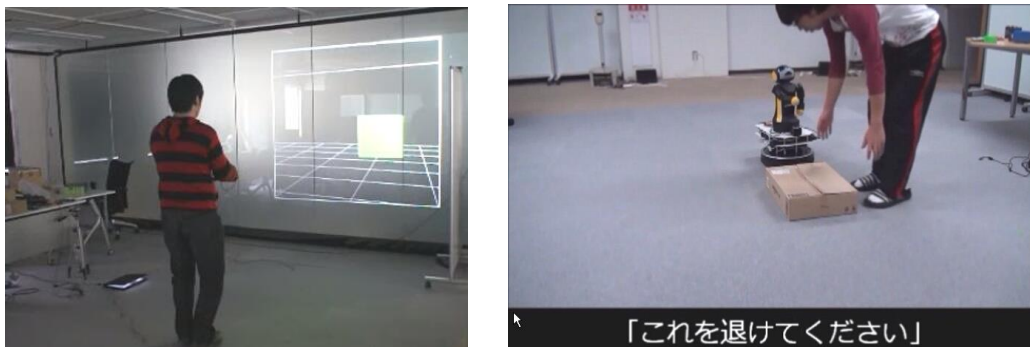


図 18 視点位置依存大型ディスプレイ(左) , ロボットと人間の協調作業の例(右)

(高齢者の発話機会の増加を目的としたソーシャルメディア仲介ロボット)

高齢者の発話機会の増加を目的として、図 19 に示す、リアルタイム性の高いソーシャルメディアと高齢者を、身体性のあるロボットが仲介するシステムの開発し、検証実験を実施した。ソーシャルメディア上のコメントをロボットの発話文として用い、さらに、高齢者の発言を提案システムが収集・認識し、ソーシャルメディアへ投稿することで、社会的孤立が問題となっている高齢者が、面倒なキーボード入力などに煩わされることなくソーシャルメディアに参加することを可能とし、高齢者の発話機会を増加させる。

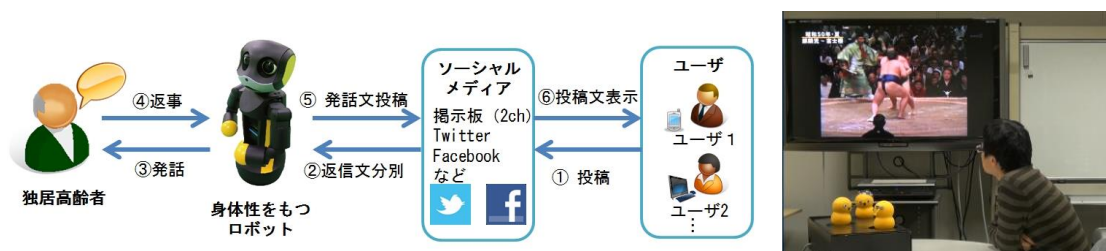


図 19 SNS を利用したロボットに人間の対話の概要 (左) と環境の例 (右)

提案システムの有効性を検証するために、提案システムを用いたテレビ視聴の検証実験を行い、人が一人、人が二人、人とロボットでテレビ視聴した場合の被験者の発話量の比較を行った結果、発話量の増加に一人よりも提案システムを用いることが妥当であることを明らかとなった。

次に、単体または複数台のロボットによりソーシャルメディアを仲介した場合を比較し、人対単体・複数ロボット対話が高齢者の発話機会に与える影響を明らかにした。これにより、高齢者のデジタルデバイドと、ロボットの自然な文脈の発話文生成が困難である従来の問題を解消し、利用者の発話機会の増加が可能となる。

(セルフモニタリング継続のための時空間的な食事行動情報の共有による食事記録支援)

健康管理のために自己の行動を記録・評価するセルフモニタリングは自発的な生活行動改善を促すことで知られている。ネットワークに接続可能なセンシング機器の普及により、体重情報や運動情報は容易に取得可能になってきている。しかし、健康管理に重要な情報の一つである摂取カロリーに関しては、食事情報を自分自身で記録する必要があるため、その煩雑さから継続的に記録することは困難である。本研究では、セルフモニタリング継続のために、食事の場所や時刻といった時空間的な食事行動情報を共有することで、食事行動の推定を実現し、簡便な食事記録を可能にすることよりユーザにかかる負担の軽減を実現した。図 20 に、セルフモニタリングのためのスマートフォンアプリの画面の例を示す。



図 20 食事行動予測によるセルフモニタリングアプリの画面例

3.2. シーンの画像データベース化と三次元環境の画像計測に関する研究

(カメラと GPS からなる全方位撮影系の構築と移動撮影)

全方位マルチカメラシステムと高精度な汎地球測位装置(リアルタイムキネマティック GPS : RTK-GPS)からなる撮影系を車両に搭載する地上移動撮影システムと電動飛行船に搭載する空撮システムを構築した。具体的な撮影場所として、奈良先端大キャンパス、生駒市北大和地区、平城宮跡等を設定した。

全方位マルチカメラシステムと RTK-GPS、無線機器等からなる撮影系を構成し、車両と飛行船に搭載した移動撮影システムを構築した。撮影予定地域での地上移動撮影と空撮を実施し、

本研究で使用するコンテンツ素材を取得した(図 21,22 参照)。また、コンテンツ素材を用いた応用として、取得した全方位動画像上に過去の建物を CG で重畳合成する、拡張現実感技術を用いたテレプレゼンス(拡張テレプレゼンス)システムを開発した。このシステムを、平成 22 年に開催された平城遷都 1300 年祭・なりきり体験館で一般公開し、延べ 1000 人以上が拡張テレプレゼンスシステムを体験した(図 23 参照)。



図 21 飛行船による空撮の様子 図 22 取得した空撮全方位画像 図 23 平成遷都 1300 年祭でのシステム一般公開の様子

(動画像と離散的な位置情報からの密なカメラ位置・姿勢情報の復元)

1Hz で位置情報を取得可能な RTK-GPS の測位情報と structure-from-motion 技法を用いて、測位誤差と再投影誤差最小化の枠組により全方位動画像の全フレームにおいて撮影時のカメラの位置・姿勢を推定する手法を開発した。

内部パラメータを較正済みのカメラシステムと RTK-GPS からなる全方位撮影系を用いて撮影されたビデオ映像に対して、時間的に粗な GPS 測位値(1Hz)を手掛かりにしたビデオ映像撮影時のカメラの位置・姿勢推定法を開発した(図 24 参照)。提案手法では、動画像を対象とした structure-from-motion 技法を用いる。GPS の測位信頼度を考慮した測位誤差のモデル化を行い、GPS 測位値に関する誤差と特徴点の画像平面への再投影誤差を同時に最小化することによって、ビデオの全画像系列についてバンドル調整を行う手法を開発した。

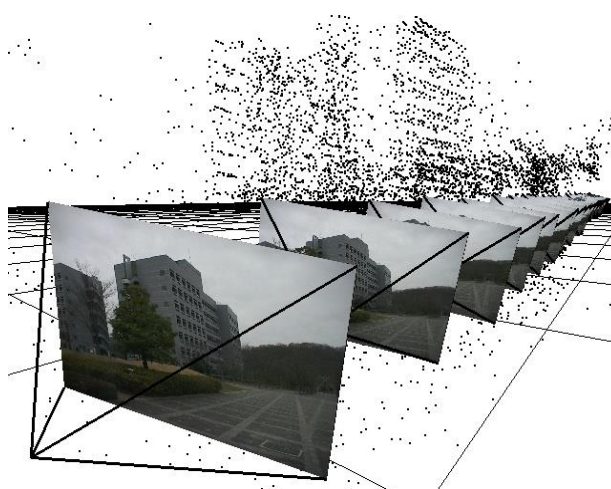


図 24 GPS と動画像によるカメラ位置の推定結果

(全方位動画からのシーンの3次元復元)

全方位動画と structure from motion 法により推定された各フレーム撮影時のカメラ位置・姿勢情報を用いて、大規模環境に対応したマルチベースラインステレオの枠組による密な奥行き情報の復元に取り組んだ。シーンを移動撮影した全方位カメラの位置・姿勢情報つき全方位ビデオ映像を用いて、ビデオの各フレームにおけるシーンの奥行き情報を抽出する、特徴点の数え上げに基づく多視点全方位ステレオ法を開発した。本手法を用いれば、高速かつ頑健に、屋外環境下において、密な三次元シーン情報を獲得することが可能となった(図 25 参照)。

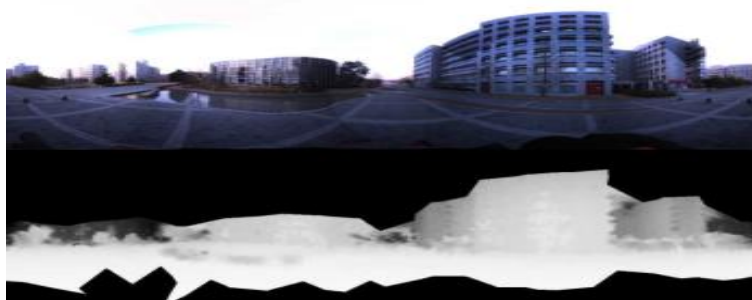


図 25 推定された全方位奥行き画像

(撮影間隔が疎なストリートビュー画像系列からの3次元復元)

Google Street View などに代表される全方位ストリートビュー画像を用い、周囲の環境の三次元点群とカメラの撮影位置・姿勢を復元する手法を開発した。一般に、ストリートビュー画像は数十 m 間隔で画像が提供されており、復元対象となる画像系列上においては、画像上で同一の特徴点の見え方が大きく変化する。これにより、一般的な手法を用いた場合には、図 26 上に示すように、多数の誤対応が生じ、正しく三次元復元を行うことができない。本研究では、特徴点の固有スケールおよび固有回転方向を利用し、この整合性を検証することで、正しい対応点を抽出する手法を開発した。加えて、正しいと考えられる対応点周辺に於いて、ガイドマッチングと呼ばれる手法により、追加での対応点探索を行うことで、対応点を増加させ、カメラ位置・姿勢推定の頑健性を向上させることに成功した。

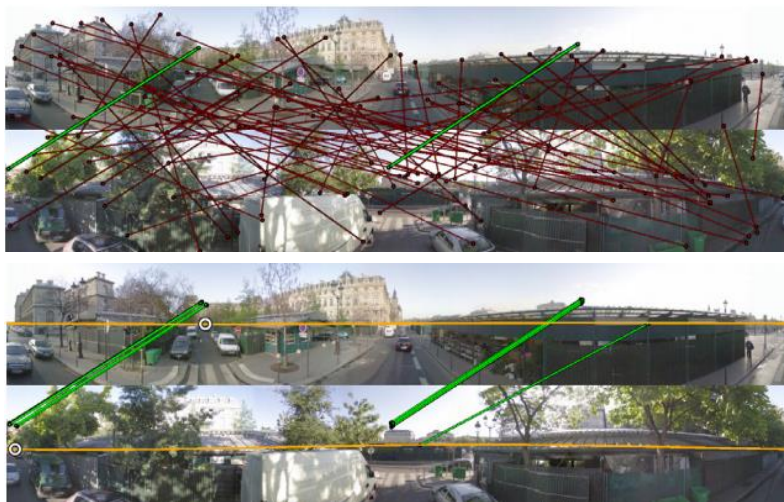


図 26 ストリートビュー画像に対する対応点の例(赤: 謝った対応点, 緑: 正しい対応点)

(空撮画像を用いたカメラ位置姿勢推定)

空撮画像を用いることで、地上においてスマートフォン等で撮影した画像の撮影位置・姿勢を推定する手法を開発した。本研究では、地上撮影画像中の特徴点と空撮画像上の特徴点を頑健に対応づけることで、カメラの位置・姿勢推定を実現する。ただし、一般に、空撮画像上において利用可能な特徴点の多くは、周期的なテクスチャパターン上に存在し、また、道路標識のように、全く同一のパターンが多数存在する場合がある(図 27 参照)。このようなシーンについて、従来の対応点探索手法を適用すると、多数の誤対応が発生し、正しくカメラ位置を推定することができない。本研究では、図 28 に示すように、多数の誤対応が含まれる特徴点の初期対応付け結果から、局所および大局的な特徴点对応の整合性を考慮することで、正しい対応点のみを自動で抽出し、カメラ位置を推定する手法を開発した。

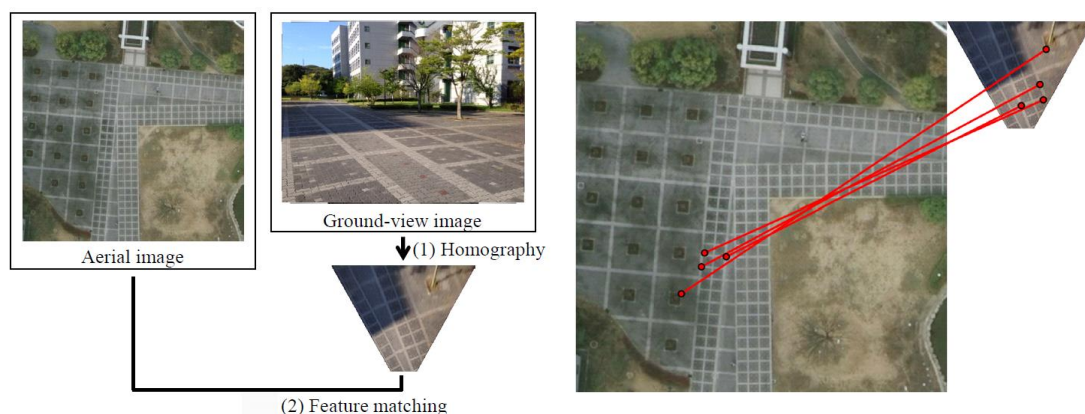


図 27 地上撮影画像と空撮画像の対応付け問題 (左) 地上撮影画像に対する射影変換 (右) 提案手法により得られた正しい対応点の例

3.3. 拡張現実感によるユーザ支援に関する研究

(事前生成型拡張現実感提示システム)

事前にシーンの全方位画像を撮影し、これに CG により描かれる仮想物体を合成しておくことで、iPad 等のモバイル端末を用いた頑健な拡張現実感システムを実現する手法を開発した。本手法は、従来一般的に行われてきたオンラインでの実シーンと仮想シーンの位置合わせを明示的に行わないことで、一般的な拡張現実感で問題となるジッタの問題を解消した。また、事前生成画像中から人などの移動物体を画像処理によりあらかじめ除去しておくことで、提示映像中の仮想物体が遮られることのない、臨場感の高い拡張現実感システムを実現した。本研究は、東大寺および慶應義塾大学大学院との共同研究として実施し、東大寺での 2 回の実証実験を通して手法の有効性を検証した(図 28 参照)。本公開実験用のシステムの作成にあたり、東大寺監修の下、慶應義塾大学のチームが提示用コンテンツを作成し、奈良先端大のチームが上記の拡張現実感提示技術を開発した。延べ 200 名を超える体験者によるアンケート結果から、本システムが拡張現実感技術を用いたユーザ支援システムとして、高い有用性を有していることが明らかとなった。ただし、1 回目を実施した公開実験においては、事前に撮影された景観と、実際

に見る景観の差異(天候や照明条件の違い)が気になるとの意見が得られ、これに対応する手法の改良も行った。改良した手法を用いた 2 回目の公開実験においては、光源環境の変化に対応する提示手法の効果について被験者実験を行い、有効性を確認した。

なお、本公開実験は、環境知能センサ網グループおよび環境知能対話グループと合同で実施し、環境知能センサ網グループは東大寺におけるネットワーク環境の構築および検証を、環境知能対話グループは新たに開発した透視型拡張現実インタフェースの有効性の調査を担当した。



図 28 東大寺における事前生成型拡張現実感システムの実証実験の様子

(未知環境における三次元領域可視化および強調のための拡張現実システム)

タッチパネル機能を有するディスプレイ、およびその背面に取り付けられたカメラとコンピュータを用い、未知環境において三次元領域を可視化および強調提示するためのシステムを開発した。未知である環境の三次元形状とユーザの持つカメラ位置姿勢を同時に推定し、推定した三次元形状の形状特徴と、ユーザがタッチパネルを用いて指示した強調すべき位置に応じて、強調すべき領域を自動的に推定して提示する。

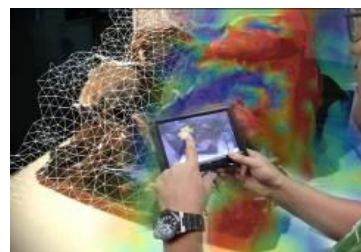


図 29 プロトタイプの外観

図 29 に示すプロトタイプと土砂災害を模したジオラマを用い、正しく指示すべき領域をユーザに提示できるかを確認するための実験を行った。ジオラマは土砂と 3 つの岩石を模した構造が存在し、岩石を強調表示させたいユーザが対象をカメラで観測しつつタッチパネルで指示することで、未知の環境であるジオラマの中から岩石を正しく強調表示させることが可能であることを確認した。

(紙面のレイアウト整形支援のための補助線投影システム)

本研究では、紙面の背面から紙面形状に応じた補助線を提示することにより、紙面上で文字を書いたり図を描いたりする創作作業を支援するシステムを提案した。本システムでは、カメラによりテーブル上の作業対象の画像が取得され、紙面の形状、姿勢が検出され、これに基づき紙面の各辺に平行な補助線がプロジェクタにより投影される。補助線投影システム(図 30 参照)を用い、作業対象を一枚の紙面に限定した状態



図 30 補助線投影システム

での実験を行った。紙面検出性能を確認する実験においては、はがきやノート、メモ用紙などを用いた場合、すべての紙面が検出されることを確認した。また、補助線の提示位置の精度評価を目的とした実験においては、2mmの精度で補助線を提示できることを確認した。さらに被験者実験を行い、提示情報が有意に効果を示していることが確認された。

(視覚障がい者支援のための衣類の色および模様の提示システム)

本研究では、視覚障がい者に向けた適切な情報提示のため、衣類の色と模様を自動で認識するシステムを開発した。カメラで撮影した画像を、PCCS表色系およびカテゴリカルカラーの概念を導入してクラスタリングを行うことにより、衣類の色および模様を簡潔にユーザに提示することで、身に着ける衣類の色や模様を視覚障がい者が自ら認識できるようなシステムを構成した。提案したシステムを用いて成人男性17名、および女性3名の計20名に対し評価実験を行い、システムが正しい色および模様を認識しているかを評価した。色認識においてはクラスタリングに改善の余地がみられるものの、模様認識においては概ね良好な評価を得られ、本システムの有効性を確認できた。

(重ね合わせ可能な半透明二次元カラーマーカ)

カメラ位置姿勢推定のために環境中に設置されるマーカにおいて、高さ方向への直感的な重ね合わせ操作を実現するためのマーカを提案した。提案するマーカは、IDパターン領域に加え、マーカの重なり状態および重なり順序を推定するための領域が配置されている。1枚のマーカは減色混色法における三原色から1色を選択して作成する。重ね合わせられたマーカの重なり状態および重なり順序推定領域を観測することで、最大3枚までの重なり状態および重なり順序を推定でき、さらにそれぞれのマーカのIDパターンを識別できる。

提案したマーカを試作して、実際に重ね合わせ操作が可能であるか実験を行った(図31参照)。重ね合わせた際の、2枚のマーカの許容位置ずれおよび回転ずれについて調査を行い、これらの許容ずれ量が実用可能な値であることを確認した。また、これらのマーカを用いてテーブルトップデザインシミュレータを試作し、仮想物体の重ね合わせが可能であることを確認した。



図 31 半透明カラーマーカ

3.4. 音声認識・音声合成および音声インタフェースに関する研究

(自然な音声対話システム)

自然なヒューマンインタフェースのためのシステム改良を目指し、音声対話システム「たけまるくん」に対し、高品質かつ発話様式制御に優れた隠れマルコフモデル(HMM)に基づく音声合成技術を導入するとともに、データベースの半自動拡張技術を構築した。また、ユーザが親しみやすい音声情報案内システムとしてこれらを実装し、実証実験を試みた。

システムからの出力音声の生成処理として、自然な声質と柔軟な発話様式制御が可能となるHMM音声合成手法を、音声対話システム「たけまるくん」に導入した。また、合成音声の品質をさらに改善させるために、二次モーメント制約を考慮したHMM学習アルゴリズムを提案し、実験的にその有効性を確認した。さらに、合成音声の声質を直感的に制御するた



図 32 「遷都たけまるくん」全景



図 33 子供ユーザの利用光景

めに、知覚的な声質表現語スコアと音響的特徴量の対応付けを可能とする重回帰HMMによる音声合成技術を導入し、より少量の学習データを用いても頑健に動作するモデリング技術を構築した。これらの音声合成技術と入力音声に対する大人と子供の自動識別結果とを組み合わせることで、大人と子供に対して、それぞれ適した文表現および声質・発話様式による応答を行えるようにした。

この技術を導入したシステムを平成22年に開催された「平城遷都1300年祭」平城宮跡会場の「なりきり体験館」に出展し、「遷都たけまるくん」として三ヶ月半にわたり運用した。遷都祭出展にあたり、既存の質問用例データベースをもとに、キーワード置換を行うなどして、遷都祭向けの質問応答ができるように拡張する技術も導入し、その有効性を実験的に示した。本システムは訪問者の好評を博し、設置期間を通して、本システムの総入力数は**131,702**発話に達した（図32,33参照）。

(機械学習の導入による高精度化)

音声対話システム「たけまるくん」の情報案内タスクにおける応答性能を改善するため、機械学習によるトピック分類を導入した。また、音声認識や音声合成における発音推定処理の精度を改善するために、過学習や雑音データに対して頑健な識別的学習アルゴリズムを考案した。さらに、WEBの情報を活用して未知語や読みの変化に対応する技術も開発した。

音声情報案内システムの応答生成性能を改善するために、機械学習によるトピック分類処理を導入した。SVM（サポートベクターマシン）、ME（最大エントロピーモデル）による識別結果のスタッキングにより、小規模セットによる予備的な結果ながら大人音声で90パーセントの分類性能を得た。また、無効入力の棄却処理にも応用し、システムの頑健性を高めた。

音声認識や音声合成における発音推定処理の精度を改善するために、線形識別器による発音推定処理を実装し、識別的オンライン学習を導入した。二値分類等の比較的単純な識別問題に対してその有効性が確認されている学習アルゴリズムをベースとして、発音推定処理のような構造学習問題への適用を可能とし、かつ、過学習や雑音データに対して頑健に動作する学習アルゴリズム（構造化AROW、構造化NAROW）を考案した。様々なデータベースに対する実験的評価の結果から、現存する発音推定技術において世界最高レベルの精度を達成できることを示した。

情報案内システムにおいて、対象となる用語およびその読み方は年月とともに変化していくため、WEBの情報を活用してそれらの変化に対応する技術も開発した。文字列と読みを表す音素列との対応づけ（アライメント）を行う際に、任意の長さの文字列と音素列との対応を考慮することができる確率的アライメント法を考案し、未知語に対するアライメント精度を改善できることを実験的に示した。また、本アライメント法を用いて、WEBから単語に対する読みを検索してくることで、未知語に対する読み推定処理を可能とし、先に述べた高精度化した発音推定処理を組み合わせることで、発音推定処理の性能を大幅に改善することに成功した。

(ミュージカルノイズフリー雑音抑圧技術)

音声インタフェース使用時における外部雑音の影響を低減させるために、雑音抑圧技術の開発を行った。従来の雑音抑圧処理において生じるミュージカルノイズと呼ばれる人工的な音色成分に対し、高次統計量を用いた理論解析を行い、ミュージカルノイズを発生させない雑音抑圧アルゴリズムを考案した。また、外部雑音除去に加え、残響除去も行う技術の構築に取り組んだ。

従来よく用いられているスペクトル減算処理に基づく雑音抑圧処理において、高次統計量であるカートシスを用いたミュージカルノイズ解析を行い、ミュージカルノイズを発生させないパラメータ設定が存在することを理論的に明らかにした。この理論解析結果の妥当性を実験的にも示した。この解析結果に基づき、従来の反復型スペクトル減算処理における自動パラメータ制御を実現し、ミュージカルノイズフリーな雑音抑圧処理を実現した。また、同様の解析を、他の雑音抑圧技術であるウィナーフィルタ法や最小平均二乗誤差規範短時間振幅スペクトル推定法（MMSE STSA法）でも行い、高次統計量に基づくミュージカルノイズ理論の有効性を明らかにした。さらに、マイクロフォンアレー信号処理においても、本理論を適用することで、高い雑音抑圧性能と聴覚的処理ひずみの低減を両立させた雑音抑圧技術を実現した。また、外部雑音抑圧処理と残響抑圧処理を同時に最適化する技術を考案し、その有効性を示した。

(体内伝導音声強調システム)

秘匿性の高い発話による通話を可能とする技術として、専用の体表密着型マイクロフォン（NAMマイクロフォン）により収録される周囲に聞こえないほど小さなささやき声（NAM）を、聞き取りやすい声へと変換する統計的音声変換技術を開発した。また、通話への応用を可能とするため、リアルタイム変換アルゴリズムを開発した。さらに、計算機環境に依らず本技術の使用を可能とするために、計算量を大幅に低減することで、携帯性の高いデバイスであるDSP上での動作を可能とした。

NAMマイクロフォンにより収録されるNAMは、周囲に聞こえないほど小さな声による発声を可能とするものの、体内伝導収録の影響により、その明瞭性および自然性は大きく劣化する。そこで、音声合成分野において古くから研究されてきた統計的音声変換技術を応用し、NAMを聞き取りやすい声へと変換する統計的音声変換技術として、NAMから空気伝導通常音声への変換技術と、NAMから空気伝導ささやき声への変換技術を構築した。実験的評価の結果から、どちらの変換技術もNAMの明瞭性を大幅に改善することができることを示し

た（NAMの単語正解率45.9%を、通常音声への変換により71.8%に、ささやき声への変換により75.9%に改善）。また、両変換技術を比較すると、通常音声への変換よりもささやき声への変換の方が、明瞭性および自然性において優れていることを明らかにした。本技術を発展させ、NAMのみでなく、体内伝導ささやき声、体内伝導小声、体内伝導通常音声に対しても適用可能な統計的音声変換技術を構築し、その有効性を示した。

NAM変換技術を人対人の音声コミュニケーションで使用するには、処理遅延の少ないシステム構築が本質的に重要となる。そこで、従来の発話単位バッチ変換アルゴリズムに対して、カルマンフィルタによる近似処理を導入することで、同程度の変換性能を保ちつつ、50～70 ms 程度の処理遅延で動作可能なリアルタイム変換アルゴリズムを考案し、その高い有効性を示した。さらに、計算量の大幅な低減を可能とする近似処理を導入することで、携帯性の高いデバイスであるDSP（不動小数点型、TI社 TMS320C6748、375 MHz）上で動作する変換処理プログラムを構築した。本技術を2012年9月に開催されたイノベーション・ジャパンにて発表し、当時の文部科学大臣による見学等も含め、好評を博した。

(体内伝導音声認識システム)

秘匿性の高い発話による音声入力を実現するために、非可聴つぶやきを認識して文字に変換するシステムを開発した。また、大量に入手可能な自然音声も利用することで、NAM用音響モデルを改善する学習アルゴリズムも開発した。また、発話者の動作に伴い、体表密着型マイクロフォンと皮膚の摩擦により生じる雑音に対処するため、二つの体表密着型マイクロフォンを用いたブラインド雑音抑圧アルゴリズムを開発した。

サイレント音声入力を実現する技術として、NAM認識システムを開発した。NAMの音響的特徴をモデル化するために、数十人規模のNAM発話データを収録し、NAM用音響モデルの学習を行った。また、大量に入手可能な自然音声データも活用するために、自然音声からNAMへの変換処理を統合したNAM用音響モデル学習アルゴリズムを考案し、その有効性を実験的に示した。

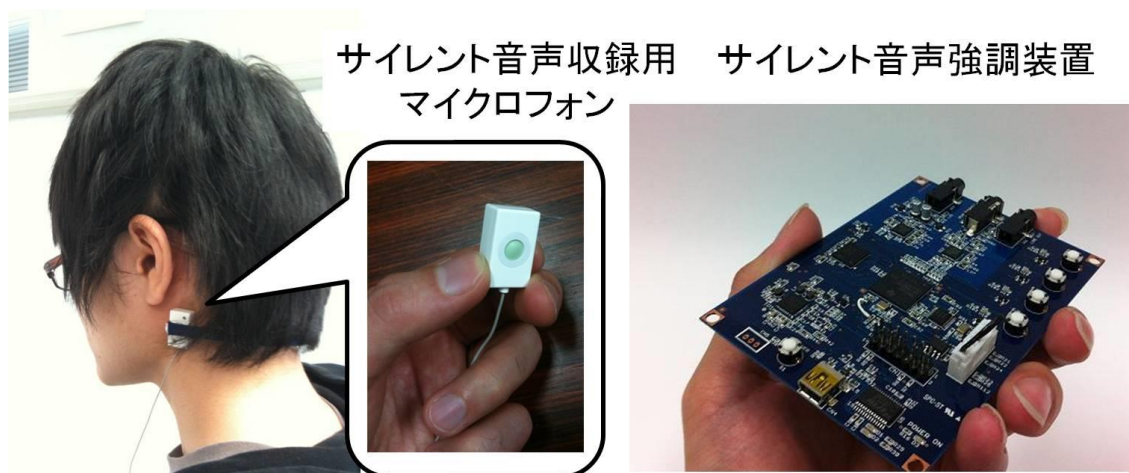


図 34 サイレント音声通話システム

NAMマイクは体表密着型マイクロフォンであり、極めてパワーの小さな音声を収録するために感度を向上させている(図34参照)。その結果、発話者が動くと、NAMマイクロフォン

と皮膚との摩擦により大きな雑音が生じる。この動作雑音はNAM認識精度を大幅に低下させる要因となるため、二つのNAMマイクロフォンを用いた動作雑音抑圧アルゴリズムを考案した。ブラインド音源分離技術を応用することで、NAMマイクロフォンの設置位置や体内伝達関数等に依らない動作雑音抑圧処理を実現し、その高い有効性を示した。

(環境知能機構グループ成果取り纏め: 佐藤 智和)