

複合コミュニティ空間のための人物動作理解と人物映像加工

Human behavior understanding and human image processing for MIXED COMMUNITY SPACE

中村 裕一[†]

北原 格[†]

坂上 勝彦[‡]

大田 友一[†]

Yuichi NAKAMURA Itaru KITAHARA Katsuhiko SAKAUE Yuichi OHTA

[†] 筑波大学 機能工学系

[‡] 電子技術総合研究所

[†] IEMS University of Tsukuba

[‡] Electrotechnical Laboratory

1 はじめに

「複合コミュニティ空間」とは、現実世界と仮想世界が融合した複合現実の感覚を複数の人間が共有するための、コンピュータによって形成される人工空間である。すなわち、複合コミュニティ空間では、複数の人間同士が現実世界で行っている視覚的な情報交換と現実世界には無い新たな視覚情報の両者を同時に共有する。我々の研究では、このような人工空間を自然なかたちで実現するための新しいパターン認識・理解の枠組みを追求することを目的としている [1]。

2 複合コミュニティ空間の課題

現在の複合現実感技術では、観測・通信・提示の種々の問題により、現実世界で一緒に行動している人間同士のコミュニケーションが阻害される。例えば、頭部装着型ディスプレイ (HMD) を着用すると、一緒に行動している人間同士の視線や表情が読み取りにくくなる。

そのため、人間を中心とした実時間対話環境を実現するためには、複合現実感を実現するために失われる情報、伝わりにくい情報を復元すること、必要に応じてそれらを強調・加工して伝えることが必要となる。そのために我々が取り組んでいる具体的な課題を以下にあげる。

- 互いの顔や動作を視認するための視覚支援
- 複数人による注意や注目の共有のための視覚支援
- 互いを含む周囲の状況を把握するための視覚支援
- 必要な情報を適所で提示するための視覚支援

本研究では、上記の個々の機能を実現するだけでなく、これらの機能を統合した複合コミュニティ空間のプロトタイプシステムを構築し、その評価を行うことを目指している。

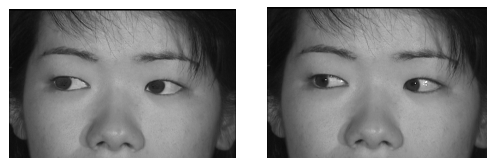
3 個々の視覚支援研究

3.1 顔を視認するための視覚支援

顔は人間にとって最も親しみのある対象の1つであり、特に視線、表情は音声では表現できない情報を伝える。我々の研究では、このような情報伝達形態を阻害しないように、HMD をかけた人物の顔に、HMD をかけていない状態の顔画像を重ね合わせて表示する。

任意視線の顔画像生成: HMD 装着者の視線方向がわかっている場合¹、その方向に対応する眼付近の顔画像を生成し、撮影された映像に重ねる。IBR (Image Based Rendering) の手

¹本研究で、視線検出機能付き HMD を試作した。



(a) 合成画像 (b) 実写画像

図 1: 任意視線の顔画像生成

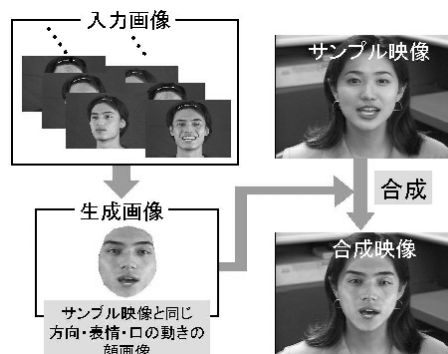


図 2: 任意方向・表情・口形状の顔映像生成

法を用いており、視線の異なる複数枚の顔画像があれば、任意の視線をもった顔画像を生成することができる。図 1 に、生成された顔画像の一部を示す。左が生成した任意視線の顔画像であり、右が実際に同じ方向を見た時の顔画像である。任意方向・表情・口形状の顔画像生成: 任意の方向・表情・口の動きを持った顔映像を生成し、それを現実の映像中の人物像と合成する。これによって、映像中の人物の顔を、同じ姿勢・表情・発話の状態で、他の顔にすり替えることができる。合成結果の一部を図 2 に示す [3]。今後は、視線の動き再現と統合し、照明の効果も加えて、より自然な合成を可能とすること、また表情や口形状の指定を自動化すること、合成をリアルタイム化することを検討していく。

3.2 注意・注目の共有のための視覚支援

複合コミュニティ空間で、注意・注目を共有するために、各々の人間の動作・発話の意図を推定し、それによって各々に対する提示情報を制御する。

現在は、コミュニケーションにおいて特に重要である、指示動作・例示動作を対象とし、以下の項目について研究を進めている。

- 動作と発話の統合処理による、指示・例示動作の認識
- 指示対象や話題の対象になる人物や物体を撮影するためのカメラ制御
- 動作情報を用いた、提示情報の切り替え

そのために、動作の3次元データ(磁気センサから得られる)、発話データ(音声認識システムから得られる)を基に、オンラインで指示動作認識を行うシステムを構築した。これは、我々の研究で構築した人物動作データの撮影・蓄積システム[4]を基にし、動作認識をオンライン化したものである。認識に1~2秒程度の時間遅れがあるものの、簡単な指さし動作などは良好に抽出し、指先を重点的に提示することができる。その例を図3に示す。ここでは指示動作が行われた時に、その指示対象を最も大きく写した映像に切り替えて提示している。このような視覚支援によって、複数人での注目を共有する。

今後は、システムの高速度とともに、指示対象のより一般的な特定方法、指示動作以外の動作の認識と認識結果の利用等について研究を行う予定である。

3.3 状況把握のための視覚支援

複合コミュニティ空間における自分や他人の状況を把握するために、仮想視点からの見え方を正確に生成する。そのために、多数のカメラから得られる映像を用いて被写体の3次元形状を推定し、それを用いて自由視点映像を生成する。

まず、全てのカメラから得られる画像を仮想平面上に2次元射影変換し、それらの写像が重なった領域を用いて、被写体の2次元スライス(断面)形状を取得する。この処理を、平面の高さを変化させながら行うことで、3次元物体の形状を2次元断面形状の集合として得る[5]。

次に、得られた3次元形状を用いて仮想視点から見えるべき自由視点画像を生成する。その際には、3次元的隠れの影響や被写体表面の方向を考慮に入れ、最も適していると思われるテクスチャを選択して自由視点画像の画素値とする。図4に生成された自由視点人物映像を示す。現在、これらの処理の高精度化・高速化について研究を行っている。

3.4 情報提示のための視覚支援

複合コミュニティ空間で行動する人間に対し、適所で適切な情報を提示する。そのために、ユーザ視点位置から撮影された実画像から、視点位置・姿勢を実時間で推定する手法の開発を行っている。

従来、実環境画像からの視点情報の実時間推定手法では、実環境に人工的に配置されたマーカを抽出・追跡し、この情報を用いて視点情報を獲得もしくは補正している。我々の手法では、色ヒストグラムマッチングとテンプレートマッチングを組合せて適用することで、単眼の実環境画像からマーカを用いずに既知の特徴点情報を追跡してユーザの視点情報を獲得し、実環境を撮影した画像と3次元CGの合成画像を実時間で提示する。この手法を用いて3次元CGを合成した画像提示の様子を図5に示す。この例ではコアラのぬいぐるみのCGモデルを、その研究紹介用のパネルに対する付加情報として観察者に提示している。いわば飛び出すポスターである。合成環境を提示する手法にはビデオスルー方式による合成画像提示を採用した。



(a) 指示動作 (b) 指示対象の提示
図3: 指示動作による提示映像の切替

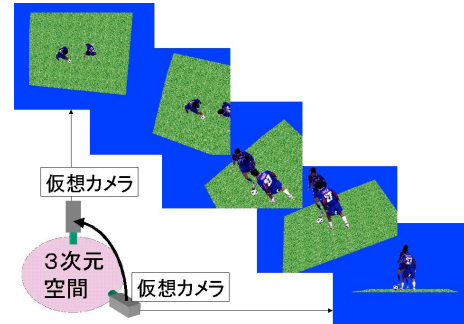


図4: 自由視点人物映像



図5: 情報提示の例(飛び出すポスター)

今後は、追跡特徴点の自動選択、特徴点検出開始の自動化、推定精度の向上・高速化の研究を行い、実世界シーンを実時間で扱うウェアラブル型の視覚情報メディアシステムを開発する予定である。

4 まとめ

複合コミュニティ空間の枠組みとその要素研究について簡単に述べた。今後、個々の要素技術を完成させていくとともに、実際に複合コミュニティ空間として動作可能なプロトタイプシステムの構築を開始する予定である。

参考文献

- [1] 横矢直和, 大田友一: “視覚増強のためのパターン認識・理解”, MIRU2000 論文集, 2000
- [2] 向川康博, 中村裕一, 大田友一: “複数の顔画像の組み合わせによる任意方向・任意表情の顔画像の生成”, 信学論, Vol.J80-D-II, No.6, 1997.
- [3] 大和哲也, 大田友一: “実写画像に基づく顔映像の生成と実映像との合成”, MIRU2000 論文集, 2000
- [4] Y.Nakamura, Y.Kimura, Y.Ye, Y.Ohta: “MMID: Multi-modal Multi-view Integrated Database for Human Behavior Understanding” Int. Conf. on Automatic Face and Gesture Recognition, 1998
- [5] 北原格, 大田友一, 金出武雄: “多視点映像の融合によるスポーツシーンの自由視点映像生成”, MIRU2000 論文集, 2000