

## 分散講義のための没入型仮想環境構築における 非言語情報の獲得

有田 大作 谷口 倫一郎

九州大学大学院システム情報科学研究院知能システム学部門

{arita,rin}@limu.is.kyushu-u.ac.jp

**概要** 我々がこれまで研究を行ってきた多視点画像からの実時間モーションキャプチャシステムの応用として、獲得された動作情報により仮想空間中のアバタを動かし、そのアバタを介して人間どうしのコミュニケーションを支援することが考えられる。このとき、人間の動作情報のすべてを利用するのではなく重要なものだけを選択して利用し、それ以外の動作情報は事前知識として与えた情報によって補完し、アバタを自然に振舞わせることを考える。そこで本研究は、このようなアバタを介したコミュニケーションとして没入型遠隔講義を扱うことにする。そして遠隔講義システムの構築を通して、人間の動作情報の獲得と提示の効率的な手法について明らかにすることを目指している。本稿では、本研究の概要を説明し、遠隔講義システム構築に向けた最初の実験の結果を示す。

## Nonverbal information acquisition on immersive virtual environment for distributed classroom

Daisaku Arita and Rin-ichiro Taniguchi

Department of Intelligent Systems, Kyushu University

**abstract** As an application of real-time motion capture system with multiple cameras which we have researched for, such a system can be proposed that supports communication between humans via avatars in a virtual space that behave based on acquired information about humans' motion. For making an avatar behave naturally, we aim that not all information acquired about human motion, but only important information acquired about human motion and information pre-defined as knowledge to compensate other information are used. Therefore, we will develop a virtual environment for immersive distance learning as an example of avatar-based communication system in order to make it clear how to balance between acquired information about human motion and pre-defined information.

### 1 はじめに

我々はこれまで、多視点画像からの実時間モーションキャプチャシステムの研究を行ってきた [1]。これは磁気センサなどのマーカをつけることなく、カメラからの動画像情報のみを利用して人間の動作情報を獲得するものである。観測された動作情報から、例えば実時間で仮想

空間中のアバタに同じ動作をさせることが可能となる。これを利用することにより、アバタを介して仮想空間中のオブジェクトに対して操作を行ったり、ひとつの仮想空間中の複数のアバタを介して人間どうしのコミュニケーションを行ったりすることができる。

しかし、このような応用を考えたとき、獲得された動作情報の誤りによってアバタの動作が

不自然になってしまうという問題が発生する。動作情報に誤りが含まれてしまうことは避けられないので、この問題を解決するためには、人間の動作に関する何らかの知識によって動作情報を補正し、アバタの動作を自然なものにする必要がある。ここで、システムの目的に応じて、画像解析によって獲得する情報と事前に知識として持たせておく情報のバランスを変えることが考えられる。つまり、できるだけ実際の人間の動作に近い動作をアバタにさせる必要がある場合には、画像解析によって獲得する情報をなるべく多くし、事前知識としての情報による補正をなるべく避けることになる。一方、人間の動作とまったく同じであるかどうかにはこだわらず、人間の意図さえアバタが表現することができればいい場合には、獲得した動作情報から人間の意図を表す動作情報のみを抽出し、それ以外の情報は事前知識を利用して補うことによって、情報獲得側から情報提示側への情報転送量を削減しながらも、アバタに人間の意図を表す動作をさせることができる。

このような考えに基づき、本研究では伝えたい情報の用途に応じてこのバランスを変えながら、人間と人間がアバタを介した自然なコミュニケーションを行うことを目指す。ただし、人間の意図を表す動作情報がどんなものであるかについては、その人間がおかれた状況によって異なる。そこで、我々は上記のようなコミュニケーションの状況として遠隔講義を考え、仮想空間中のアバタを介した遠隔講義システムの構築を通して、人間の動作情報の獲得と提示の効率的な手法について研究を行う。

## 2 遠隔講義システム

近年、テレビ講義、インターネット講義などの遠隔講義が盛んに行われるようになってきている。特に、インターネットを利用した遠隔講義は、参加者がどこにいても好きな講義を受けることができるので、教育における地理的障壁を取り除く有力な手段である。つまり、今後重要になるであろう個人適応型学習の枠組みを提供することになる。一方で、学習におけるクラ

スルームやクラスメートの役割も重要であると言える。例えば、受講者は他の受講者から多くのことを学ぶであろうし、グループ学習は有効な学習形態のひとつである。したがって、遠隔講義の枠組みとしては、個人ベースの特徴と集団ベースの特徴の両方を備えたものが望ましい。

また、遠隔講義の最大の問題は、視覚的および聴覚的情報を講義サイトの間で交換しているにもかかわらず、参加者が実際の講義を受けているという雰囲気を感じることができないことである。例えば、参加者は自分が見たいところをいつでも見ることができるわけではない。

ここで、我々の研究の位置づけを明確にするために、他の遠隔講義形態との比較を行う。

**テレビ放送** テレビ放送を利用した遠隔講義は歴史があり最もよく利用されている遠隔講義の枠組みである。テレビさえあれば誰でも講義に参加することができるので、地理的に隔離された参加者が何人でも同時に参加することができるという利点がある。しかし、講師や受講者の間でインタラクションがないということが問題であり、どのような受講者がいるのかということ講師はまったく知ることができず、受講者は講師に見られているという感覚を持つことができない。これにより、現実感に乏しく緊張感に欠ける講義になってしまう。

**双方向型遠隔講義** いくつかの学校や会社では、2ヶ所以上の講義室を音声と映像のネットワークで結び、カメラとプロジェクタによってそれぞれの講義室の状況を映し出す双方向型遠隔講義システムを導入している [2]。これを利用することにより、講師は自分がいる講義室の受講者だけでなく遠隔地の講義室にいる受講者の姿を見ることもできる。そして、遠隔地の講義室にいる受講者も講師とインタラクションを行うことができる。しかし、受講者にとって自分がいない講義室の受講者とインタラクションを行うことは簡単ではない。さらにこのシステムのもうひとつの問題として遠隔講義システムが用意されている講義室まで行かなければならないことが挙げられる。

ビデオ会議 NetMeeting などのビデオ会議システムを利用した小規模の遠隔講義も外国語専門学校などで広く行われている。このシステムの利点は、特別な装置を要求しないので、容易にその講義を受講することができるようになることである。しかし、映像と音声の品質はそれほど高くない。さらに、ディスプレイ上に同時に表示できるウィンドウの数の制限から、参加できる学生の数に限りがあることが問題である。

**VEIDL** 我々が構築する遠隔講義システムは、没入型遠隔講義のための仮想環境 (VEIDL: Virtual Environment for Immersive Distance Learning) である。具体的には、ネットワーク上に仮想講義室を構築し、その中のアバタを介して地理的に隔離された講師と受講者がコミュニケーションを行うことができるシステムである。VEIDL を利用することにより、参加者はまるで実際の講義空間にいるような感覚を持つことができる。これは高解像度映像や高音質音声の利用を意味するのではなく（もちろん、それらも重要な要素ではあるが）、講義空間における幾何学的な一貫性によるものである。例えば、それぞれの参加者が自分の見たいところを見ることができたり、ある参加者が指差したものがなにかをすべての参加者が理解できたりといったことである。これが可能になったのは、講義空間が仮想的に再構成されたものであり、その再構成された情報がそれぞれの参加者に提供されているからである。遠隔地を撮影した画像をそのまま提示する方式の遠隔講義システムでこのような特徴を得ることは不可能である。

### 3 VEIDL の詳細

仮想講義室の中では、講義室のリアルな状況が提供され、参加者はそれぞれが話すことができるだけでなくほかの参加者の様子を見ることがもできる（図 1 参照）。このとき、講師を含むそれぞれの参加者はそれぞれ別々の場所（サイト）にいて、HMD（またはその他の 1 人用 3 次元表示装置）とヘッドホンを用意しているとす。これらの装置を通して、それぞれの参加

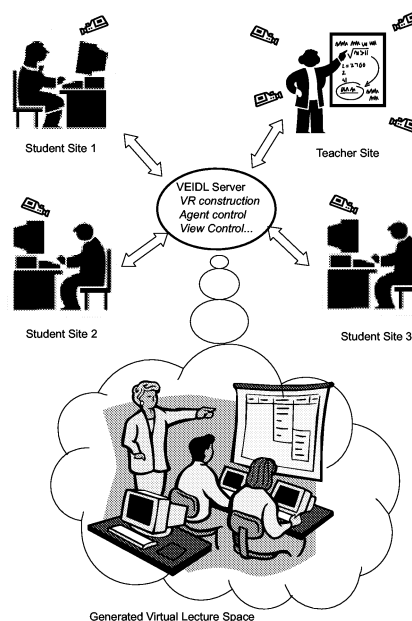


図 1: システム概念図

者は各サイトで構築された仮想講義空間中の参加者を、あたかも現実の講義空間にいるように見聞きすることができる。言い換えれば、現実世界で離れ離れの場所にいるそれぞれの参加者が仮想世界中のひとつの講義空間を共有していることになる。

このとき、1 節で述べたように

- 参加者の情報獲得時には、参加者の意図を提示するための情報など必要な最低限の情報のみを獲得し、
- 講義シーンの提示時には、それぞれの参加者についてのモデルを各参加者のエージェントに持たせ、獲得・転送されてきた以外の情報を補うことにより、仮想講義空間中の参加者のアバタを自然に振舞わせる

ことを目指す。

このようなシステムを実現するためには、以下に挙げる機能が必要である。

#### 3.1 情報の獲得

講師については、複数カメラと PC クラスタを利用した全身用実時間モーションキャプチャ

システムを利用することにより、講師の位置と姿勢の情報を獲得する。その情報を基に、必要な動作情報を抽出する。受講者については、ステレオカメラと1台のPCを利用した上半身用実時間モーションキャプチャシステムを利用し、同様の処理を行う。講師についての情報は通常非常に重要であるため十分なカメラとPCを用意するが、受講者についての情報は通常それほど重要ではなく、またそれぞれの受講者のために高価な装置を利用することは現実的ではないため必要最低限のカメラとPCのみを利用する。

ここで問題になるのはどのような情報を獲得し、どのような情報を事前知識として用意しておくかということである。本研究では情報の用途別に以下のように考えている。

#### 仮想空間の位置あわせのための動作情報

VEIDLの長所として空間の幾何学的な一貫性が挙げられる。これを実現するためには参加者どうしのやり取りにおいて位置関係が重要な役割を果たす動作について、正確な位置情報が必要となる。具体的には、

- 講師の位置
- 参加者の顔の向き
- 指差しの方向

が挙げられる。

#### 参加者の意図を表す動作情報

参加者の意図さえ伝われば正確な動作を再現する必要のない動作情報として

- 歩行
- 指差し
- 拳手

が挙げられる。例えば、歩行に関しては歩いていることさえわかれば、正確な足やひざの位置を再現する必要はない。これらの動作は意図的に行われるので、画像解析によって情報を抽出することは比較的容易であるが、情報抽出の誤りは許されない。

#### 参加者が知りたい情報

参加者、特に講師が知りたい情報として

- 眠い
- 注意力散漫
- 落ち着きがない
- 理解できていない

といったものが挙げられる。これらの情報は意図的に行われるわけではないので、画像解析によって情報を抽出することは難しいが、役に立つ情報なのでアバタの動作に反映させることが望ましい。意図を表す情報と比較して情報の抽出自体に誤りがあっても許される場合が多い。

なお、音声に関しては、マイクロホンから獲得した音声情報をそのまま利用し、音声認識は行わない。

### 3.2 情報の転送

獲得された参加者に関する情報は、共通の仮想空間を構築するためにサイト間で交換される必要がある。この情報交換は情報管理サーバと呼ぶサーバを経由して行われる。したがって、システムのネットワーク構造は情報管理サーバを中心とした星型になる。各サイトで獲得された参加者に関する情報は情報管理サーバに集められる。そして、集められた情報は情報管理サーバから各サイトへ転送される。この情報管理サーバから各サイトへの情報転送において、データ転送量を減らすために情報の選択を行う。動作情報については、各サイトの参加者の視線の向きと情報の用途に応じて選択を行う。音声情報については、仮想空間中での参加者間の距離に応じて選択を行う。

### 3.3 情報の提示

講義シーンは、獲得された参加者についての情報と、あらかじめ用意されている教材に起案する情報を基に、各サイトのPC上で実時間で構築される。各参加者はHMDとヘッドホンを通して講義シーンを見聞きすることができる。このとき、参加者の姿勢に応じて提示される情報が実時間で変化する。

このとき事前知識としての情報を利用してアバタを自然に動作させるために各参加者のエージェントを用意しておく。参加者エージェントは

- その参加者の色形状モデル
- 表情モデル
- 参加者共通の動作モデル

を事前知識として持っておき、届けられた動作情報を基にアバタを動かす。このようにしてアバタを動かすということは、サイトによってアバタの動作が異なる可能性があるということの意味する。しかし、参加者の動作の中で重要な情報はアバタの動作に確実に反映されているので、それほど大きな問題にはならない。

## 4 実験

本稿では、情報の獲得と提示についての最初の段階の実験を行ったので、その結果を示す。実験の内容は、1人の受講者についての動作情報を獲得し、その動きをアバタによって実時間で再現するものである。扱う動作は拳手と頬杖（それぞれ右手と左手）であり、それらを参加者の意図を表す情報とした。今回は1人の動作を1ヶ所で再現するだけなので、情報管理サーバを利用せず、情報獲得部と情報提示部が直接通信を行っている。

情報獲得部では、ステレオカメラによって撮影された画像から受講者の動作を解析し、拳手および頬杖の動作が行われたかどうかを判定している。動作が行われた時点、および動作が終了した時点（拳手をやめた時点と頬杖をやめた時点）に、それぞれの動作情報を情報提示部に転送する。情報提示部では、届いた動作情報を基に、あらかじめ用意されていたアバタの動作を表示している。

図2から図7にステレオカメラからの入力画像とアバタの表示画像を示す。各図中の上の2枚の画像がステレオカメラからの入力画像、下の画像がアバタの表示画像である。参加者が意図を表す動作を行ったときはアバタもその通りに動作し、それ以外の動作を行った場合はアバタは動いていない。

## 5 おわりに

本研究では、没入型遠隔講義のための仮想環境 VEIDL を構築することにより、仮想空間中のアバタを介した人間どうしのコミュニケーションにおける人間の動作情報の獲得・提示の効率的な手法を見出すことを目的としている。本稿では、この研究の枠組みを説明し、その最初の実験結果を示した。今後は、VEIDL システムを構築するとともに、時間的な一貫性の問題について検討する必要がある。時間的な一貫性とは、イベントが発生した順序と同じ順序で参加者に対してイベントが提示されることを意味する。さらに、このようなイベントレベルの時間的な一貫性に加えて、信号情報の時間的な一貫性も、自然な講義空間構築のためには必要である。これは、動作情報と音声情報の時間的な一貫性を保つこと、代表的なものとしては口の動きと音声の同期をとることである。VEIDL のような分散型システムにおいてはこのような時間的な一貫性を保つことは最も難しい問題のひとつである。

## 謝辞

本研究は科学研究費補助金学術創成研究(2)「人間同士の自然なコミュニケーションを支援する知能メディア技術(13GS0003)」の補助を受けた。

## 参考文献

- [1] R. Taniguchi, S. Yonemoto, D. Arita, Real-time human motion analysis for human-machine interface, Proc. of the Working Conference on Advanced Visual Interface, pp.195-202, 2002.
- [2] Y. Kameda, K. Ishizuka, M. Minoh, A Study for Distance Learning Service - TIDE Project -, Proc. of IEEE International Conference on Multimedia and Expo, vol.3, pp.1237-1240, 2000.

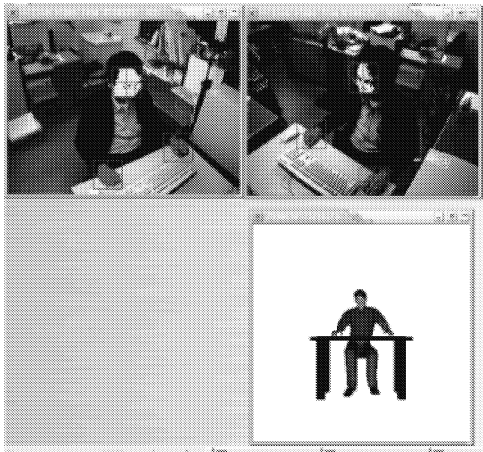


図 2: 標準姿勢

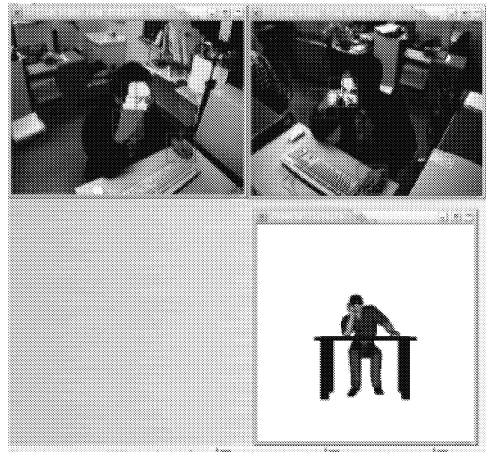


図 5: 頬杖 (右手)



図 3: 拳手 (右手)

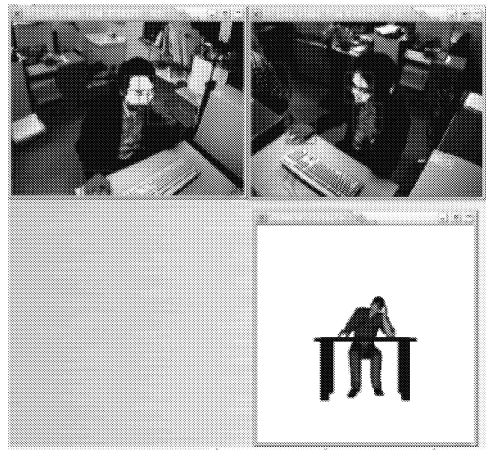


図 6: 頬杖 (左手)

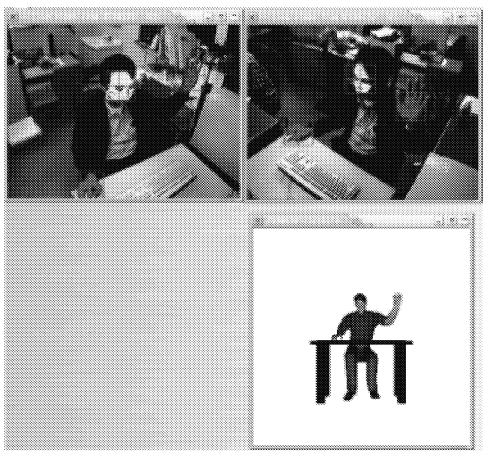


図 4: 拳手 (左手)

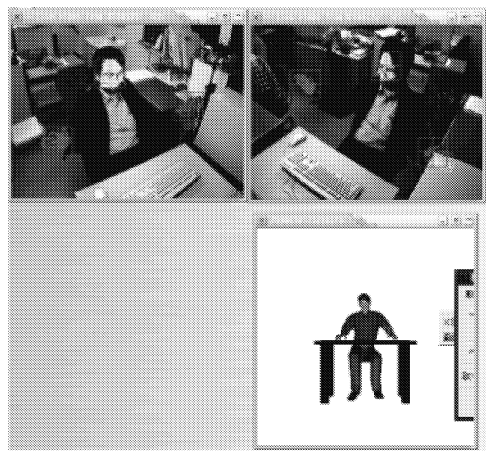


図 7: 不要な動作