

Real-time Human Proxy : 仮想空間を用いた遠隔コミュニケーション

有田, 大作
九州大学システム情報科学研究院知能システム学部門

吉松, 寿人
九州大学システム情報科学研究院知能システム学部門

羽山, 大介
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎
九州大学システム情報科学研究院知能システム学部門

<https://hdl.handle.net/2324/5870>

出版情報 : 電気学会研究会資料. IIS, 産業システム情報化研究会, pp.17-22, 2005-05. 電気学会
バージョン :
権利関係 :

Real-time Human Proxy: 仮想空間を用いた遠隔コミュニケーション

有田大作*, 吉松寿人, 羽山 大介, 谷口 倫一郎 (九州大学)

Real-time Human Proxy: Distant Communication via Virtual Space
Daisaku Arita, Hisato Yoshimatsu, Daisuke Hayama, Rin-ichiro Taniguchi (Kyushu University)

Abstract

In this paper, we propose Real-time Human Proxy (RHP), which is a new concept for distant communication via a virtual space. Using RHP, information about participants is captured, symbolized, transferred and presented in real-time. And as a prototype system using RHP, we have developed a Virtual Environment for Immersive Distributed Learning(VEIDL) for estimating RHP. Also we present experimental results.

キーワード: 遠隔地コミュニケーション, 仮想空間, アバタ, 動作認識, 動作表現
(distant communication, virtual space, avatar, action recognition, action presentation)

1 はじめに

遠隔地の人同士がコミュニケーションを行うためにこれまでさまざまなツールが開発されてきているが、近年はネットワーク技術の発達もありテレビ電話が普及してきている。これは音声情報だけでなく視覚的な情報も利用することができるため表現力が豊かなコミュニケーションツールとして期待されている。しかし、テレビ電話を多対多のコミュニケーションに用いようとする、誰の発話なのか、誰に向けた発話なのかといったことを把握するのが難しくなってしまう。これは、テレビ電話では参加者間の空間的な関係が欠落してしまっていること、つまり、自分に対して相手がどのような位置に存在しているのかという空間的な配置に関する情報が得られず、その整合性が取れないため、参加者が混乱してしまうことに起因している。

これを解決するために、仮想空間を介したコミュニケーションについての研究が行われている。これは、全ての参加者をアバタとして仮想環境に投影し、参加者は仮想環境中に表示されるアバタを介してコミュニケーションを図るものである。しかしこの手法では、現在の参加者の状態を絶えず仮想環境に反映させなければならない。そのためこの手法は、全ての参加者にそれぞれのアバタを的確に操作し続けなければならないことを強要する。これは、対戦ゲームなどのように操作自体やその技術を競うことが目的である場合を除き、参加者に対して本来のコミュニケーションの目的とは関係のない負担となってしまう。

そこでわれわれは、遠隔地の人同士が仮想空間を介してコミュニケーションを行うための枠組みであるリアルタイムヒューマンプロキシ(RHP)を提案している。RHPとは、参加者を観察することで参加者についての情報を獲得し、その中からコミュニケーションに必要なものを選

択的に参加者間で交換し、交換された情報を基に参加者の代替としてのアバタを操作することで、コミュニケーションを実現するものである。さらに、情報を交換する際は、生の情報(例えば、モーションキャプチャシステムの獲得データ)を交換するのではなく、それを認識処理した結果であるシンボル情報を交換する。これにより、アバタ表現の自由度が上がるだけでなく、動作や音声などのモダリティ間の情報交換も可能となる。本稿では、リアルタイムヒューマンプロキシについて概説し、それを身体動作情報に用いた実験について報告する。

2 リアルタイムヒューマンプロキシ

RHPでは、人物の情報をシンボル化する、またはその逆を行なう。人物の情報とは、例えば手、足の動きや形状、顔の形や表情、声質、発せられた言葉、無意識にしがちな癖など、その人に関する情報一般を指す。シンボルとは、それらのうち1部分の情報についての記述である。すなわち、RHPとは、ある人につわる情報をシンボルという単純な記述の集合へ変換すること、もしくはその逆に、シンボル群からその人の表現をつくり出すことである。

そのとき、人物に関する情報はその全てが常に必要とされているわけではない。さらに、伝達すべき情報とそうでない情報は想定する状況によって決定される。RHPは人物につわる情報を適当と考えられる細かさのシンボルに分解し、必要と思われるものを選択する。そして選択されたシンボルからその人物を表現する。

このように人物情報をシンボル化することの利点としては、以下の2点が挙げられる。

人物表現の自由度 例えば生の動作情報を交換する場合は、それを表現するアバタは、参加者と身体構造や体格がまったく同じである必要がある。また、生の音声情報

を交換する場合は、その音声に合うようなアバタの外見でなければならない。一方シンボル化した情報を交換する場合は、アバタの身体構造や体格にあわせた動作情報をシンボルから生成することによってどのようなアバタでも利用可能となる（例えば、CGで表示されるアバタだけでなく、人型ロボットをアバタとして利用することも可能となる）。同様に音声情報についても、音声合成によってアバタに合うような声を作り出すことができる。このように、実際の参加者とは異なる条件のアバタを利用することが可能となり、人物の表現の自由度が高まる。モダリティー間の情報交換 動作情報、音声情報、表情情報などといった異なるモダリティー間で情報を交換するためには、それぞれの生の情報から交換可能な情報を引き出すために、情報の抽象化が必要となる。この抽象化された情報がシンボルであり、これによりモダリティー間の情報の交換を行うことができる。例えば、発話が強くなるタイミングにあわせて手を振らせたり、楽しそうな表情をしているときには声を明るくするといったことが考えられる。

3 RHPの実現

本節では身体動作情報についてRHPを用いるための、人物情報の入力、シンボル化、表現の各過程について述べる。

3.1 入力：モーションキャプチャシステム

入力が困難な人間の情報に体の動きがある。これはアバタを介したコミュニケーションには欠かせないものであるが、人間のからだの動きが持つ自由度は非常に高いため、キーボードやマウスのような入力装置では満足な入力を確保することが難しい。加えて、体の動きを表現することがインタラクションの目的そのものでない場合も多く、この場合は参加者が煩雑な入力操作によってアバタの動作を決定しなければならないことは大きな負担となるばかりでなく、本来目的とするインタラクションの妨げともなり得る。

RHPでは体の動き情報に関して、実時間モーションキャプチャシステム（以下MCS）を入力装置として使用する。これにより体の動きそのものを処理することが可能となり、普通の仕草をそのまま入力として利用できる。したがって参加者は、入力方法を特別に学習する必要がなく、負担を大きく軽減できる。

3.2 シンボル化：ジェスチャ認識

MCSから得られる人間の動き情報は動作に関するシンボルへと変換される。変換にはジェスチャ認識の手法を利用する。採用するシンボル、すなわち認識されるべき動作はあらかじめ決定しておく。これは想定するインタラクションの性質、および設定によって、必要とされる動作情報の種類が変わってくるためである。

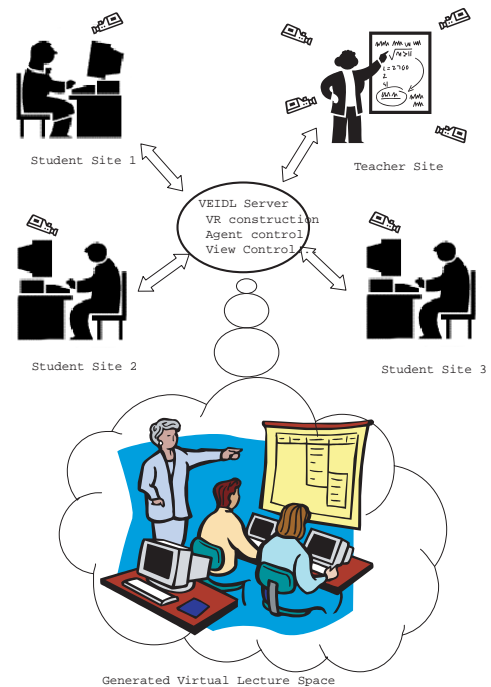


Fig. 1: Virtual classroom by VEIDL

3.3 表現：動き情報生成

シンボルの形で表現された人物の情報を利用者に分かりやすく提示するために、3次元コンピュータグラフィックスを用いてアバタを表示する。あらかじめシンボルに対応づけた動き情報を用意しておき、それを基にアバタの動きを生成し、アバタに振舞わせる。より自然な動作を表現するために、動き情報はMCSを利用して用意することも可能である。

4 VEIDL：RHPのプロトタイプシステム

この節では、RHPのプロトタイプとして構築中の仮想講義環境（Virtual Environment for Immersive Distributed Learning, 以下VEIDL）について述べる。VEIDLを構築する目的は、ある特定のコミュニケーションを想定しそれに必要なシンボルを選出した上で、RHPの概念に基づいたシステムを動作させ、実際にRHPを用いてコミュニケーションを行うことが可能であることを示すこと、およびRHPの評価をすることにある。

4.1 VEIDLの概要

VEIDLとは、参加者が遠隔地から個々に接続することを想定した仮想講義環境（図1）である。参加者は仮想環境にアバタとして参加し、同様に仮想環境に参加する他のアバタを通じて各参加者とコミュニケーションを行う。仮想環境の中では講義が執り行われる。講義の形態は、講師はスライド映像を映す仮想スクリーンと指示棒を用いて解説をし、生徒はそれを受講することを基本とする。

また、参加者が討論を行う際はそれぞれが向かい合い、身ぶりを交えて会話を行うことができるものとする。

4.2 システム構成

VEIDL では、各参加者は入力装置として複数台のカメラとマイク、出力装置としてディスプレイとスピーカをもつ。カメラとディスプレイは非言語情報の獲得と提示に使用する。非言語情報として参加者の体の動きがカメラによって獲得され、得られた情報はシンボル化され、他の各参加者へ配信される。配信された各参加者の側では、受信したシンボルに基づいてアバタをコンピュータグラフィクスによって仮想環境中に描画し、それをディスプレイに表示する。仮想環境は参加者のアバタが存在する位置からの視点で描画される。マイクとスピーカは言語情報の獲得と提示に使用する。マイクから得られる言語情報はそのまま各参加者に配信され、スピーカによって参加者に提示される。これにより参加者は自分のアバタの視点から描画された他の参加者のアバタを通じてインタラクションを行うことが可能となる。

以下に VEIDL における動作情報の入力から表示までの処理について述べる。

4.3 動作情報の獲得

非言語情報である動作に関する入力には、本研究室で構築された実時間非接触型のモーションキャプチャシステム（以下 MCS）を用いる。講義を行う仮想環境への参加者には講師と生徒の二種の形態が存在するため、それぞれに適した MCS を使用する。使用する MCS の一方は全身用 MCS であり、これは講師が使用することを想定している。もう一方の MCS は上半身用 MCS であり、これは生徒が利用することを想定している。

4.3.1 全身用非接触型実時間 MCS

全身用 MCS[1] は、一人の人物について各体部位の 3 次元座標を実時間で獲得するシステムである。獲得には、カメラ映像の画像処理を使用する。このシステムは、対象となる人物に機材を装着せずにその姿勢を獲得することに主眼を置いて構築されたシステムであり、よって対象人物はマーカ等の機材を装着することを必要としない。姿勢の獲得には、肌色抽出、輪郭処理、逆運動学に基づいた推定処理を複合的に使用する。さらに、このシステムは実時間性を重視している。

この MCS は 8 台のカメラと、9 台の PC によって構成される PC クラスタから構成されており、MCS の観測空間内に存在する人物の頭部、両手、両足、両肘、両膝、両肩、腰両側、および胴体中心の 3 次元座標を実時間で獲得することができる（図 2）。また、拡張として色のついたマーカを同時に使用することが可能であり、これによって観測空間内の任意の座標を獲得することもできる。

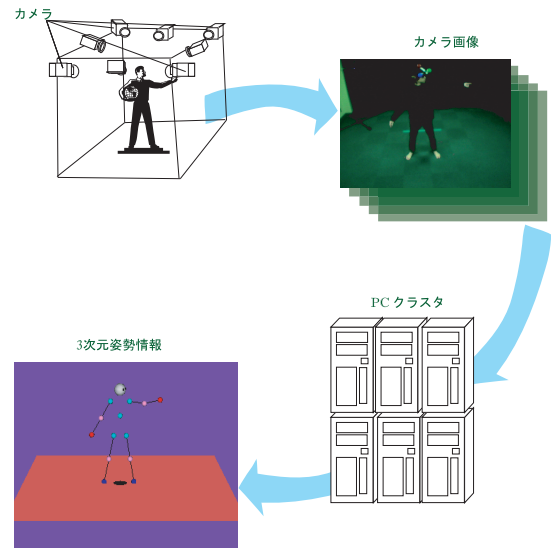


Fig. 2: MCS(whole body)

VEIDL が想定する講義環境において、講師はより動きの自由度の高い参加者である。それに加え、生徒に比べ環境に参加する人数は非常に少ない。このことから、比較的大規模ながら全身の情報が得られるこの MCS を講師の動き情報獲得に使用する。

4.3.2 上半身用非接触型実時間 MCS

上半身用 MCS[2] は、3 次元座標を獲得する対象の体部位を上半身に限定した MCS であり、特徴は上に述べた全身用 MCS と多くを共有する。この MCS は、PC ディスプレイを前方に配置した机上で使用することを想定しており、3 次元情報を入力するためのデスクトップインタフェースとなるものである。そのため要求される機材は全身用 MCS に比べ少なく、比較的構築・運用が容易である。

この MCS は 2 台のカメラと、1 台の PC によって成り立つ。これにより、MCS の観測空間内に存在する人物の頭部、両手の 3 次元座標を身体にマーカを装着することなく実時間で獲得することができる（図 3）。また、拡張として色のついたマーカを同時に使用することが可能である。

4.4 シンボル化

MCS によって獲得された人物の姿勢の時系列は、ジェスチャ認識の手法を用いて認識され、シンボルに変換されるシンボル化されるべき動作は、実現する仮想環境の種類によって変化するため、VEIDL が必要とするシンボルはあらかじめ決定しておく。以下に、採用するシンボルの種類と、動作の認識手法を述べる。

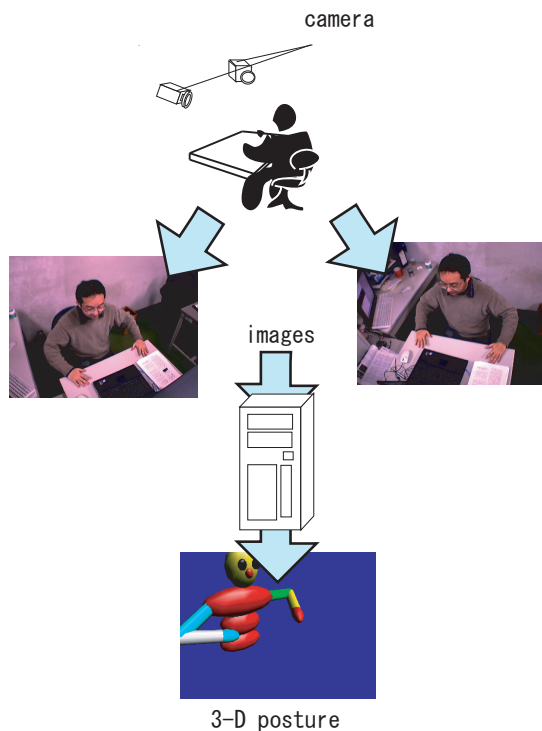


Fig. 3: MCS(upper body)

4.4.1 シンボル

シンボルは動作の種類を表すラベルと、動作の内容を記述するいくつかの値によって成り立つ。ここでは、これを‘シンボル名(値1, 値2, …)’と示す。以下に VEIDL において現在実装しているシンボルを列挙する。

講師側シンボル

- 顔の向き (方向)
講師の頭部の向きを示す。頭部の向きは講師の注意の向きを示す重要な情報であるため、シンボル化する。なお、顔の向きを表す値を伴う。この値は顔の向きのベクトルの場合と、向いている参加者の識別子の場合がある。
- { 右手/左手 } 指さし動作開始 (方向)
講師が指さし動作を開始したことを示す。意図的に空間中の物を示したいときに、指で方向を指示することが考えられる。そのため、シンボルとして採用する。なお、方向は‘顔を向ける’シンボル同様、指で指している方向のベクトルの場合と、指されている物体の識別子の場合がある。
- { 右手/左手 } 指さし動作変更 (方向)
講師が指さし動作を変更したことを示す。新たに指さした方向を示す値を伴う。

- { 右手/左手 } 指さし動作終了
講師が指さし動作を終了したことを示す。
- 指示棒による指示動作 {{ 開始/変更 } (方向) / 終了 }
指さし動作と同様に、講師が指示棒を持っていた場合の指示棒の動きを示す。
- 歩行動作 {{ 開始/変更 } (方向, 速度) / 終了 }
講師は全身の姿勢情報が得られるので、仮想環境内を歩行することが可能である。そのため、歩行の開始、歩行内容の変更、歩行の終了をシンボルとして示す。なお、歩行の開始と変更には、歩行速度および歩行方向を表す値が伴う。
- 身体の向き (方向)
講師のからだの向きは、顔の向き同様、講師の注意の向きを示す重要な情報であるため、シンボルとして採用する。

生徒側シンボル

- 顔の向き (方向)
講師側と同様である。
- { 右手/左手 } 指さし動作 {{ 開始/変更 } (方向) / 終了 }
講師側と同様である。
- { 右手/左手 } 挙手動作 { 開始/終了 }
生徒側が発言するためには、自分が発言しようとしていることを表現する必要がある。講師側は参加人数が少なく、さらに環境の司会者的立場にあるため、このシンボルは採用していない。一方、生徒側は同様な参加者が多数存在するため、このシンボルを採用する。

4.4.2 シンボル化手法

MCS から得られる情報が上に述べたシンボルに該当する動作があればこれを認識する。現段階では、認識すべき動作は限られたものであり、比較的単純な手法を使用することができる。以下にそれぞれの動作に対して用いられる認識手法を説明する。

- 顔の向き
MCS で得られる顔の向き情報をそのまま利用する。顔の向く先にアバタが存在する場合は、そのアバタの方向を向いているものと判断する。
- 指さし動作
身体の中心から手までの距離が一定値以上離れた場合、指さし動作をしていると判断する。この際、身体の中心から手までのベクトルを延長した先に

アバタが存在する場合は、そのアバタを指さしているものと判断する。

- 歩行動作
足が動いている状態で、身体の中心座標が一定値以上移動した場合に歩行していると判断する。
- 身体の向き
顔の向きと同様である。
- 挙手動作
手の座標を頭部の座標と比較し、頭部より高い座標にある場合、挙手をしていると判断する。

なお、これらの手法はシンボルの増加に伴って、より高度なものへと移行することが予定されている。

4.5 動き情報の生成

仮想環境で交換される情報は観測値そのものでなく意味の解釈を経たシンボルの形であり、それを VEIDL 利用者に分かりやすく提示する事が求められる。ここでは人物を表現するシンボルが十分であることを評価する目的で、さらに仮想環境への没入感を高めるため、仮想環境の表示には 3 次元コンピュータグラフィクスを利用する。

シンボルの形で伝えられた動作情報をより自然に分かりやすく 3 次元コンピュータグラフィクスで表現するためには、シンボルをアバタの動きへ変換させなければならぬ。以下では、シンボルを基にアバタの動きを生成する手法を述べる。

4.5.1 行動モデルと動作計画

アバタの次の動作を選択するのが動作計画である。アバタの動作は基本的には送られてくるシンボルによって決定される。しかし、シンボルとして獲得される動作情報はコミュニケーションに必要なもののみであるため、これを表示するだけだと、アバタがまったく動かない時間が長くなり不自然になってしまう。これを避けるためには、コミュニケーションに影響のない動作を自律的に行わせる必要がある。そこで RHP では、行動モデルを使って動作計画を行う。

行動モデルには、アバタが行いうるすべての動作が、標準姿勢（気をつけの姿勢）からある姿勢への動作（往動作）とある姿勢から標準姿勢への動作（復動作）として保持されている。例えば、挙手については、手を挙げる往動作と手を下げる復動作が保持されている。シンボルに応じた動作については、シンボルを受け取ったときだけその動作が選択される。一方、自律的な動作については確率的に選択される。ただし、身体部位の重複により同時には行うことができない動作の組み合わせが存在するので、そのような動作は選択されないようにチェックしている。

4.5.2 動きモデルと動き生成

選択された動作に対応する動き情報（時系列姿勢パラメータ）を生成するのが動き生成である。動きモデルは動作と動き情報の 1 対 1 対応を記述したものであるため、これを引くことによって、動き情報を生成することができる。このとき、異なる身体部位を用いた動作は同時に行うことができるので、そのような動作の動き情報は合成され、同時に表示される。

また、すべての動作は標準姿勢を基準とした往動作と復動作から成り立っているため、それから生成される一連の動きはすべて標準姿勢を経由したものとなってしまう、不自然に感じられる。そこで、同じ部位を用いる連続する復動作と往動作に対応する動きを合成することにより標準姿勢を経由しない動きを生成する [3]。

4.5.3 外観モデルと外観生成

外観モデルはアバタの身体構造、形状、色の情報を持っており、これと生成された動き情報からアバタの外観を生成する。

4.6 表示

仮想環境の情報は全ての参加者の間で共有される。システムは、VEIDL 参加者それぞれに対して 4.5 節に述べた手法によって生成された動き情報を用いて、各アバタを共有されるアバタの存在位置に描画する。なお、この処理は仮想環境に参加する全てのアバタに対して行われるものであり、仮想環境を描画している VEIDL 参加者自身のアバタもまたこれに従う。仮想環境の描画は、該当する VEIDL 参加者のアバタからの視点で行う。これにより、自分のアバタを通じた仮想空間へのインタフェースを実現する。

5 実験

本実験では VEIDL システムを用いて簡単な講義を行い、RHP におけるアバタを用いてインタラクションが可能である事を確認する。

5.1 実験方法

仮想講義に参加する参加者は一人の講師と三人の生徒のみとする。講義では英会話のロールプレイを行う。会話内容は仮想講義室（図 4）に描画されるスライドに従う。まず役割 A が役割 B と役割 C をそれぞれ紹介し、次に役割 B と役割 C が互いに挨拶を交わすという内容である。

実験において使用する動作シンボルは、指さしと挙手、歩行（講師のみ）および頭部回転である。

仮想講義の流れを以下に述べる。まず以下の流れで各生徒の役割を決める。

- 講師が生徒に「役割 A になりたい人は手をあげてください」と伝える。
- 役割 A になりたいと思った生徒が手をあげる。

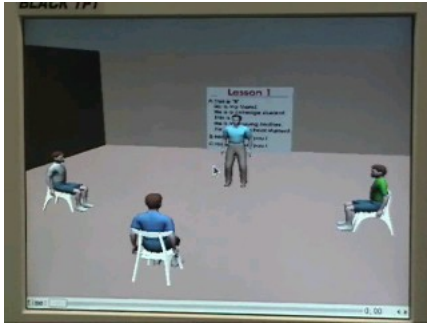


Fig. 4: 仮想環境の概観

- 講師が手をあげている生徒の中から一人を指さして役割Aになる生徒を決める。
- 講師が役割Bになる生徒と役割Cになる生徒を指さして決める。

次にそれぞれの役割になった生徒が以下の流れでロールプレイを行う。

- 役割Aの生徒が役割Bの生徒と役割Cの生徒を紹介する。紹介する際はその生徒を指さしながら行う。紹介されている生徒は手をあげる。
- 役割Bの生徒が挨拶をする。
- 役割Cの生徒が挨拶をする。

5.2 実験結果

各参加者は各自の役割を演じ、前節で述べた流れに沿ってロールプレイを行う事ができた。以下にこのインタラクションを行った感想を列挙する。

- 仮想環境における相手のアバタを指さした場合、特に遅れを感じることなく自分のアバタが相手のアバタを指さす事が確認できた。
- 相手が自分の方を向き、また指さしている事が、アバタを通して伝わり、仮想環境インタラクションがやりやすかった。
- 動きを不自然に感じる事はなかった。
- ゲーム感覚でインタラクションを楽しむことができた。

これらから、RHPによる遠隔地コミュニケーションが可能であることが示された。

6 おわりに

本稿では、仮想空間を介した遠隔地コミュニケーションを実現するための枠組みであるリアルタイムヒューマンプロキシRHPを提案し、その有効性を確かめるためにRHPを用いた遠隔講義のための没入型仮想環境VEIDLによる実験を行った。

今後の課題としては以下が挙げられる。

動作の選択 RHPではコミュニケーションに必要な動作を事前に定義しておく必要があるが、必要な動作を漏れなく定義することは困難な作業である。このため、人体動作を観測し、そのコミュニケーションで行われる動作を自動的にすべて列挙することを目指して現在研究中である。これが実現すると、列挙された動作の中から必要な動作を選択することができるようになる。

動き生成の高度化 現在の動きモデルはアバタの体格に依存したものであるため、アバタごとに動きモデルを生成する必要がある。この問題を解決するために、アバタの体格に依存しないメタな動き情報から動き情報を自動生成する手法の実現を目指す。

謝辞

本研究の一部は、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)」、および、科学研究費補助金若手研究(B)「リアルタイムヒューマンプロキシを用いたアバタベース遠隔地コミュニケーションの研究(16700108)」の補助を受けた。

参考文献

- [1] Naoto Date, Hiromasa Yoshimoto, Daisaku Arita, Satoshi Yonemoto and Rin-ichiro Taniguchi, "Performance Evaluation of Vision-based Real-time Motion Capture", Proc. of Workshop on Parallel and Distributed Computing in Image Processing, Video Processing, and Multimedia, in IPDPS CD-Rom Proceedings, 2003.
- [2] Satoshi Yonemoto and Rin-ichiro Taniguchi, "A Direct Manipulation Interface with Vision-based Human Figure Control", Proc. of HCI International 2003, pp.811-815, 2003.
- [3] Daisuke Hayama, Hisato Yoshimatsu, Hiromasa Yoshimoto, Daisaku Arita, Rin-ichiro Taniguchi, "Avatar generation for Real-time Human Proxy", Proc. of 10th International Conference on Virtual Systems and Multimedia, pp.386-395, 2004.