

特別研究報告書

バーチャルスタジオにおける 仮想物体操作の誤認識の検出

指導教員 美濃 導彦 教授

京都大学工学部情報学科

早瀬 直之

平成19年2月9日

バーチャルスタジオにおける仮想物体操作の誤認識の検出

早瀬 直之

内容梗概

バーチャルスタジオでは、CGで作成した仮想セットにカメラで撮影した実写映像中の演者領域を重畳することによって、仮想セット内に演者が存在しているような映像を作成することができる。このバーチャルスタジオを教育用映像コンテンツの作成に利用すれば、教師となる演者がCGで作成された三次元の仮想物体を教材として用いることができる。ただし、三次元物体は視点位置によって見え方が異なるため、教師が教材の任意の面を視聴者に見せられるように、仮想物体を自由に回転・移動できなければならない。

このための従来研究として、演者の手にデータグローブを装着させて演者の手の動きを計測し、その手の動きに仮想物体の動きを追従させることによって仮想物体の回転・移動を実現するというものがある。この研究ではさらに、演者の手にデータグローブを装着することによる見苦しさを防ぐため、データグローブを装着した演者の手をCGで作成した仮想手に置換している。このとき、データグローブによる計測には誤差が含まれるので、演者が仮想物体を把持しているつもりであっても、実際には仮想手と仮想物体との間に位置ずれが生じる。そこで上の研究では、演者が仮想物体を把持している場合には、仮想手が仮想物体を正しく把持した状態になるように両者の位置・姿勢を補正する処理も併せて実現している。

このような補正を行うためには、その前にまず演者が仮想物体を把持しているかどうかを認識する必要がある。上の手法では、演者に依らない一律の評価関数を用いて演者の手の位置・姿勢を評価することによって把持の有無を認識しているが、物体を把持する際の手の位置や向きには個人差があるため、仮想物体の把持の認識に失敗する場合が生じる。

本研究では、演者による仮想物体の把持を認識するための条件をシステムが学習することで、演者による把持操作を正しく認識できるようにすることを考える。このような学習のためには、システムによる把持操作の認識が正しいかどうかを教示信号として利用する必要がある。そこで本研究では、上述の従来研究の方法を用いた演者による仮想物体の把持の有無の認識において、誤りが生じた場合にそれを検出する処理の実現を目標とする。

システムが演者による仮想物体の把持を誤認識した場合，演者はそれに対処するために特徴的な行動をとると考えられる．例えば，演者が仮想物体を把持しようとしていないにもかかわらず，システムの誤認識によって仮想物体が把持されてしまった場合には，演者は仮想物体を把持している状態から直ちに抜け出すため，仮想物体を操作することなくそれを放そうとすると考えられる．また，演者が仮想物体を把持しようとしているにもかかわらず把持できないときや，把持している状態から手を放そうとしているにもかかわらず解放できないときは，それらの操作を実現するために，掴みにいく動きや解放する動きを繰り返すと考えられる．このような演者による特徴的な行動を検出すれば，システムによる誤認識を検出できると期待される．

上のような演者の特徴的な行動を検出するには，システムによる仮想物体把持認識時点よりも未来の時点も含めて演者の行動を観測する必要がある．その一方で，システムによる仮想物体把持認識の処理自体は実時間で実現できなければならない．したがって，仮想物体把持認識と誤認識検出は同一のシステムでは実現できない．

そこで本研究では，仮想物体把持認識や，それに基づく物体操作を実時間で実現するシステム（仮想物体操作システム）に加え，仮想物体操作システムを用いた演者による仮想物体操作を継続的に観測し，仮想物体操作システムによる仮想物体把持認識に誤りが起きたときに，それを非実時間で検出するシステム（誤認識検出システム）を導入する．この誤認識検出システムは，仮想物体操作システムが利用していない未来における演者の行動の情報も利用することによって，上で述べた演者の特徴的な行動を検出し，その直前に演者による仮想物体の把持の認識に誤りが生じたと判断する．

本手法の有効性を確かめるために，教材である三次元仮想物体を掴んで把持しながら説明を行った後，それを解放するというタスクを被験者に行わせ，システムによる誤認識の後にとられる被験者の特徴的な行動に基づいて，仮想物体操作システムによる誤認識を検出できることを確認した．

Error Detection for Recognizing Virtual Object Manipulation in a Virtual Studio

Naoyuki HAYASE

Abstract

Overlaying real image sequences captured by a studio camera with a virtual background image created by CG, a virtual studio can convey to a viewer the feeling that an actor is in a virtual set. The virtual studio can be used for educational purposes: in image sequences transmitted to viewers, the actor (teacher) can use an educational material in the form of a 3D virtual object created by CG. In order to show any portion of the object, the actor needs to manipulate it by his/her hand, for instance, to *grasp*, *rotate*, *translate* them.

In order to realize virtual object manipulation, it is required to measure the position and posture of the actor's hand. It is proposed in a previous work to make the actor wear a 3D position sensor and a data glove for the measurement. In the work the actor's real hand with the data glove is replaced with a virtual hand created by CG to avoid that the data glove appears in the video presented to viewers. In this process, displacement between the virtual hand and the virtual object often occurs due to the error of the measurement by the data glove and the position sensor. The previous work corrects such displacement so that the actor's hand grasps the virtual object properly. The necessity for correcting such errors depends whether the actor is grasping the virtual object or not. In the previous work grasping by the actor is recognized by the function that evaluates the sensory data obtained from the data glove and the position sensor with a set of parameters generally used for all the actors. Since the geometric relation between the virtual object and the actor's hand depends on each individual actor, the evaluation function with parameters universally used for any actor cannot recognize grasping by all the actors completely.

The system needs to adapt each actor by learning the value of each parameter most appropriate for recognizing grasping operation by each actor. If the system's failure to recognize grasping can be detected, it can be used for adaptation to the actor. Therefore, we focus on detecting the error for recognizing grasping.

When a grasping recognition error occurs, the actor should take *characteristic* actions. For example, the actor tries to *stop current manipulation* or *repeatedly retry the manipulation*. By detecting these characteristic actions, the detection of recognition error could be possible. However, to detect the characteristic actions, the system must monitor past and future actions of the actor, which cannot be done in real time. By contrast, grasping recognition needs to be realized in real time. Therefore, grasping recognition and detection of grasping recognition error cannot be realized by the same system.

This work introduces a non-real-time system which monitors grasping recognition. By referring to past and future actions, the monitoring system detects characteristic actions and determines that an error in grasping recognition has occurred just before such a characteristic action is taken.

In order to verify the effectiveness of this approach above, experiments for sequences of 3D object grasping, explanation, and release are performed. The results of these experiments suggest that the (non-real-time) monitoring system can detect the error of grasping recognition.

バーチャルスタジオにおける仮想物体操作の誤認識の検出

目次

第1章	緒論	1
第2章	教育映像撮影のためのバーチャルスタジオ	4
2.1	バーチャルスタジオ	4
2.2	インタラクティブバーチャルスタジオ	5
2.3	仮想物体の把持認識	7
第3章	仮想物体操作と誤認識検出	8
3.1	システム構成	8
3.2	仮想物体操作システム	9
3.2.1	把持認識のための手の把持段階のモデル化	9
3.2.2	各段階の定量的特徴	11
3.2.3	把持認識における状態遷移	15
3.3	誤認識検出システム	16
3.3.1	演者の反応を利用した誤認識検出	16
3.3.2	把持認識における誤りの分類	16
第4章	実験・考察	20
4.1	実験環境	20
4.2	実験結果	21
4.2.1	誤認識検出の成功例	21
4.2.2	誤認識検出の失敗例	23
4.2.3	考察	24
第5章	結論	25
	謝辞	26
	参考文献	26

第1章 緒論

近年のマルチメディア技術の普及により、様々な視覚的效果を取り入れた映像コンテンツの作成が可能となった。特にテレビのニュース番組や気象情報番組などでは、CGで作成した仮想物体のセット（以下、“仮想セット”と呼ぶ）とカメラで獲得される実写映像を合成し、視認性に優れた映像が作成されている。このような映像を作成する環境としてバーチャルスタジオ（VS: Virtual Studio）がある [1]。バーチャルスタジオでは演者をカメラで撮影することで獲得される実写映像から、クロマキー処理によって演者領域を抽出し、その演者領域を仮想セットに重畳することにより、仮想セット内に演者が存在しているような映像が合成される。合成された映像は視聴者に配信される一方、スタジオ内の床上にあるフロアモニタにも映し出され、演者はこの合成映像を確認しながら自らが仮想セットの中に存在しているかのように演技を行う。

本研究では、このバーチャルスタジオを教育用映像コンテンツの作成に利用することを考える。すなわち、教師である演者がバーチャルスタジオ内で仮想セットを教材に用いて講義を行っている映像を撮影し、その映像を教育用コンテンツとして利用する。通常の講義室での講義で用いられる教材は板書やスライドなどの二次元のものに限られるが、バーチャルスタジオをこのように利用することで、CGで作成された三次元の仮想物体を教材に用いた説明映像を作成することができる。

三次元の物体を教材に用いる場合、二次元教材とは異なり、教材の見え方が視点位置や向きに依存するので、教師は説明の際に教材を様々な位置や向きから視聴者に見せるために、必要に応じて教材を自由に回転・移動できる必要がある。このとき、教師によるその説明に迫真性をもたせるためには、教師がポインティングデバイス等を用いて教材を間接的に操作するよりは、教師自らが教材を手で直接操作する方が望ましいといえる。これは通常の講義室における講義において、教師がスライドを教材に用いて説明を行う場合に、スライドの説明のための指示をマウスでポインタを動かしながら行うよりも、指示棒を用いてスクリーン上のスライドを直接指し示す方が説明の迫真性が増すのと同様である。

三次元教材の直接操作を実現するためには、スタジオ内での教師の手の動きを計測する必要がある。しかし、従来のバーチャルスタジオには演者の動きを

計測する機能がないため、演者が仮想物体を直接操作することはできない。この問題を解決するために尾原ら [2] は、演者の手に三次元位置センサとデータグローブを装着することによって演者の手の動きを計測し、演者の手を用いた直接操作による仮想物体の回転・移動を実現している。演者の手にデータグローブを装着すると映像が見苦しいものとなるが、この手法ではデータグローブを装着した演者の手を CG で作成した仮想手に置換することで、この問題の回避を図っている。また、このような環境では実物体と仮想物体を同時に扱うため、拡張現実感 (AR: Augmented Reality) と同様、演者による仮想物体の位置・向きに関する知覚の誤差や、三次元位置センサ・データグローブの計測誤差により、実物体と仮想物体の位置合わせに誤差が生じる。さらに、演者はフロアモニタに表示される合成映像を見ながら操作を行うため、自分の手が仮想物体に正しく一致するように手を動かすことは難しい。これらの影響により、合成映像中の仮想手と仮想物体との間に位置ずれがしばしば生じる。そこで前述の手法では、演者の手を仮想手に置き換えて視聴者に配信するための映像を合成する際に、仮想手が仮想物体を正しく把持しているように両者の位置・姿勢を補正するという処理を行っている。

演者は講義の撮影中、常に仮想物体を把持しているわけではないので、仮想手と仮想物体の補正を行う際には、演者が仮想物体を把持しているときのみ両者の位置・姿勢を補正する必要がある。このためには、演者の手に装着した三次元位置センサやデータグローブの計測データに基づいて、演者が仮想物体を把持しているかどうかを認識する処理 (以下、“把持認識” と呼ぶ) が必要となる。先述の手法では、三次元位置センサやデータグローブの計測データを基に、仮想手の関節角や仮想手と仮想物体の距離などを求め、これらを評価関数によって評価することで把持認識を実現している。しかし、演者が仮想物体を把持するときの手の関節角や、手と仮想物体との距離変化等の定量的な値は、演者によって様々であるにもかかわらず、これらを演者に依らず一律な評価関数で評価しているため、把持認識に誤りが生じることが多い。

把持認識において演者による個人差に対応するには、手の関節角や仮想手と仮想物体の距離等の値を評価する評価関数内の閾値などのパラメータを、演者ごとに調節する必要がある。しかし、これを手作業で行うのは労力が大きい。ここで、システムが演者による把持の有無を誤認識したことを検出できれば、これを教示信号として、把持認識のための評価関数のパラメータを演者に適応さ

せることが可能になると期待できる．そこで本研究では，システムが演者による把持の有無を誤認識した場合に，それを検出することを目標とする．

システムによる把持認識に誤りが生じた場合，演者は仮想物体を思い通りに操作できなくなるので，それに対処するための何らかの行動をとると考えられる．例えば，演者が仮想物体を把持したいのに，システムの誤認識によって把持が成功しない場合には，演者は把持を成功させようとするため，何度も仮想物体を掴みに行くような行動をとると考えられる．このような演者による特徴的な行動を検出することにより，その特徴的な行動が現れる直前にシステムが演者による把持の有無を誤認識したことを検出できると考えられる．しかし，このような演者による特徴的な行動を検出するには，把持認識に誤りが生じた時点よりも未来における演者の行動を観測する必要がある．その一方，演者による仮想物体の操作自体は実時間で実現できなければならない．このことは，仮想物体操作の実行と誤認識の検出は同一のシステムでは実現できないことを意味する．

そこで本研究では，演者による仮想物体の把持の有無を実時間で認識し，演者による仮想物体操作を実現するシステム（以下，“仮想物体操作システム”と呼ぶ）に加えて，仮想物体操作システムを用いた演者の仮想物体操作の様子を継続的に観測し，仮想物体操作システムによる把持認識の誤りを非実時間で検出するシステム（以下，“誤認識検出システム”と呼ぶ）を導入する．この誤認識検出システムは，仮想物体操作システムが実時間で演者による把持の有無を認識する場合には利用できない未来での演者の行動も利用することにより，上述のような演者の特徴的な行動を検出し，その直前に把持認識に誤りが生じたと判断する．

本論文ではまず続く第2章で，教材映像を作成するためのバーチャルスタジオの使用法について述べる．第3章では，本研究での提案手法について述べ，第4章では，この手法の有効性を確認するために実施した実験について述べる．最後の第5章で本研究の意義についてまとめると共に，今後の課題について議論する．

第2章 教育映像撮影のためのバーチャルスタジオ

2.1 バーチャルスタジオ

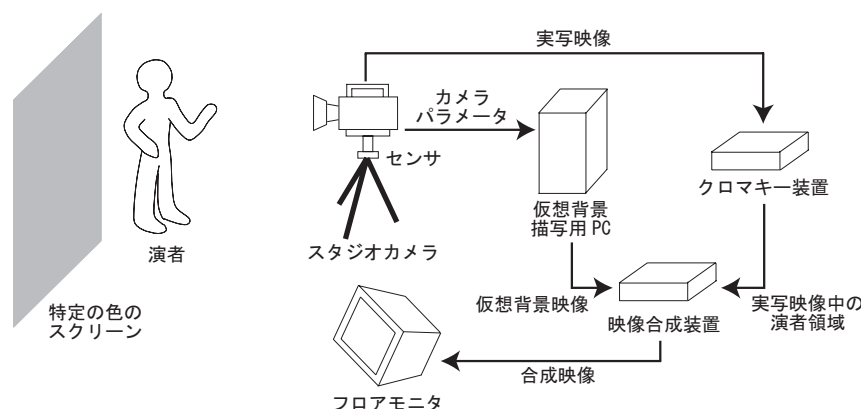


図1: バーチャルスタジオのシステム構成図

バーチャルスタジオとは、仮想セットの背景映像（以下、“仮想背景映像”と呼ぶ）と実写映像中の演者領域を合成することのできる撮影環境である。バーチャルスタジオのシステム構成図を図1に示す。

スタジオの床と壁は特定の色で覆われており、センサを搭載したスタジオカメラでスタジオ内の演者を撮影する。それによって得られた実写映像からは、クロマキー装置で映像中の特定の色領域が取り除かれることによって映像中の演者領域が抽出される。一方、センサによって獲得されるスタジオカメラの位置、方向、ズーム値などのカメラパラメータが仮想背景描写用PCに送られ、獲得されたパラメータを基に仮想背景映像が合成される。そして、実写映像中の演者領域と仮想背景映像を映像合成装置で合成することによって、視聴者に配信するための仮想セットを背景とした演者の合成映像が生成される。生成された合成映像は、スタジオ内のフロアモニタにも表示され、演者はフロアモニタ上の合成映像を確認しながら、演者自身が仮想世界の中に存在しているかのように演技を行う。このような処理により、バーチャルスタジオでは実世界と仮想世界を融合し、実物体と仮想物体の両方を区別なく視聴者に提示することができる。

本研究ではこのようなバーチャルスタジオを教育用映像コンテンツの作成に利用することを考える。近年、e-learningの分野では、教育用の映像コンテンツ

として講義をアーカイブした映像を利用する試みが盛んとなりつつある [3] . しかし , 通常の講義室で提示される教材は板書やスライドなどの二次元のものに限られるため , それをアーカイブ化した映像で用いられる教材も当然二次元のものとなる . これに対してバーチャルスタジオを利用すれば , CG で作成された三次元の仮想物体や仮想セットを教材として提示できるため , 黒板やスライドなどによる二次元の表現では説明が困難であったような説明対象を三次元表現を用いて視覚的にわかりやすく提示できる . また , このとき仮想物体に対して説明箇所のハイライト表示など , CG による様々な特殊効果を加えることにより , 視認性を高めることも可能である . また , 大学教育用の映像をスタジオで撮影する場合 , 同一科目の映像を複数日にわたって撮影しなければならない場合が多く , 同時期に複数科目のスケジュールを調整しながら撮影する必要が生じることも多い . このような場合には , それぞれの科目の背景セットを撮影の度に何度も入れ替えることが必要となるが , バーチャルスタジオでは背景セットの切り替えが容易なため , コストと時間の削減が図れる .

第 1 章でも述べたようにバーチャルスタジオにおいて , 上述のように三次元仮想物体を教材に用いた説明映像を撮影する際には , 三次元物体は二次元教材とは異なり , 視点の位置や向きによって見え方が依存するため , 物体の任意の面を見せるために , 演者による仮想物体の直接操作が必要となる . このためには演者の手の動きを計測する機能が要求されるが , バーチャルスタジオにはその機能がないため , もしそのような映像を作成するならば , 以下の二つの方法のいずれかをとる必要がある . 一つは , 振舞いが事前に決定されている仮想物体を用意し , その振舞いに合わせて演者が仮想物体を直接操作する演技を行うという方法であり , もう一つは , 演者が仮想物体を直接操作する演技に合わせて , スタジオ技術者が仮想物体の振舞いを生成するという方法である . 前者の方法では , 演者が仮想物体を直接操作する演技を , 撮影の前に練習する必要がある . また , 後者の方法では , 演者の演技に合わせて仮想物体の振舞いを生成するために , スタジオ技術者の協力を必要とする . しかし , このような労力を一般の教員やスタジオ技術者に求めるのは望ましいことではない .

2.2 インタラクティブバーチャルスタジオ

2.1 で述べたような問題に対し , 尾原ら [2] は演者の手の動きを計測することによって , 演者による仮想物体の直接操作を実現することが可能なバーチャル

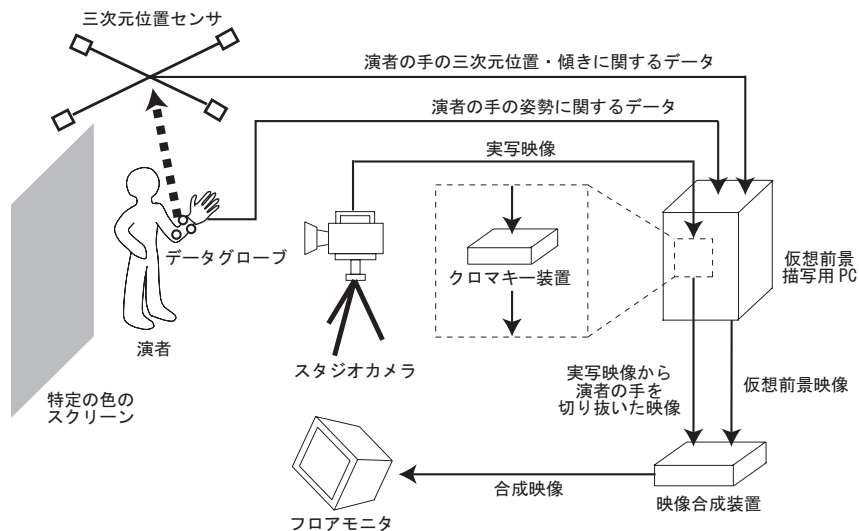


図 2: インタラクティブバーチャルスタジオのシステム構成図

スタジオ（インタラクティブバーチャルスタジオ）を提案している．このインタラクティブバーチャルスタジオのシステム構成図を図 2 に示す．

インタラクティブバーチャルスタジオでは，演者の手の動きをデータグローブと三次元位置センサで計測し，これを基に演者による仮想物体操作を実現する．このとき，演者の手にデータグローブを装着することによる見苦しさを防ぐため，演者の手を次のように CG で作成した仮想手に置換する．

まず，演者の手首と前腕に 3 つの三次元位置センサを装着させ，演者の手にデータグローブとその上から緑色の手袋を被せる．スタジオカメラで演者を撮影した実写映像は仮想前景描写用 PC に送られ，クロマキー処理によって緑色の部分を抽出することで，実写映像から演者の手を切り抜いた映像が作成される．また，三次元位置センサとデータグローブから演者の手の三次元位置・傾き，および演者の手の姿勢に関するデータをそれぞれ獲得する．獲得されたデータは仮想前景描写用 PC に送られ，それらのデータを基に，仮想手と教材である仮想物体からなる仮想前景映像が作成される．最後に，実写映像から演者の手を切り抜いた映像と仮想前景映像を映像合成装置で合成することにより，演者が自らの手で仮想物体を直接操作しているような映像が得られる．

しかし上のようなシステムでは，三次元位置センサとデータグローブによる計測誤差やカメラのキャリブレーション誤差などが原因となって，実物体と仮想物体の位置合わせに関する誤差が生じる．そのため，演者が自らの手で仮想物体を把持しているつもりであっても，仮想手と仮想物体との間に位置ずれが

生じる(図3)。また、演者はフロアモニタを見ながら仮想物体を操作するので、自分の手が仮想物体に一致するように手を動かすこと自体が難しい。これらの問題に対処するために、上の手法では、仮想手が仮想物体を正しく把持するように、仮想手と仮想物体の位置と姿勢を補正する処理も実現している。

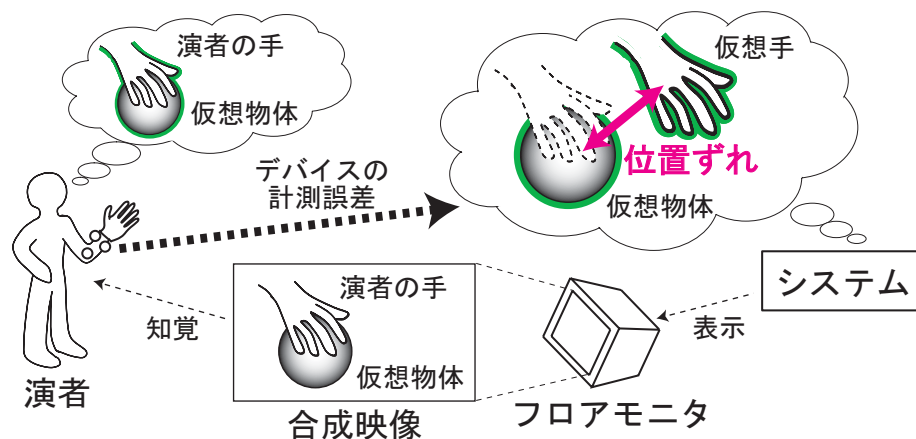


図3: 仮想手と仮想物体との位置ずれ

2.3 仮想物体の把持認識

仮想手と仮想物体の補正においては、仮想手が仮想物体を把持しているときだけ位置ずれの補正を行うようにする必要があるため、仮想手が仮想物体を把持しているかどうかを認識しなければならない。上の手法ではこのために、仮想手と仮想物体の定量的な位置関係を単純な評価関数で評価することによって把持認識を実現している。しかし、仮想物体の把持の有無を判断する基準となる仮想手と仮想物体の定量的な位置関係は演者によって様々に異なる。この個人差の影響により、把持認識を異なる演者に対して一律な評価関数で実現すると、把持認識に誤りが生じる。そこで本研究では、仮想物体操作システムによる把持認識の誤りを誤認識検出システムで検出し、それを教示信号として把持認識のための評価関数を学習することで、仮想物体操作システムを演者へ適応させることを図る。続く第3章でこのような仮想物体操作システムと誤認識検出システムの詳細について述べる。

第3章 仮想物体操作と誤認識検出

3.1 システム構成

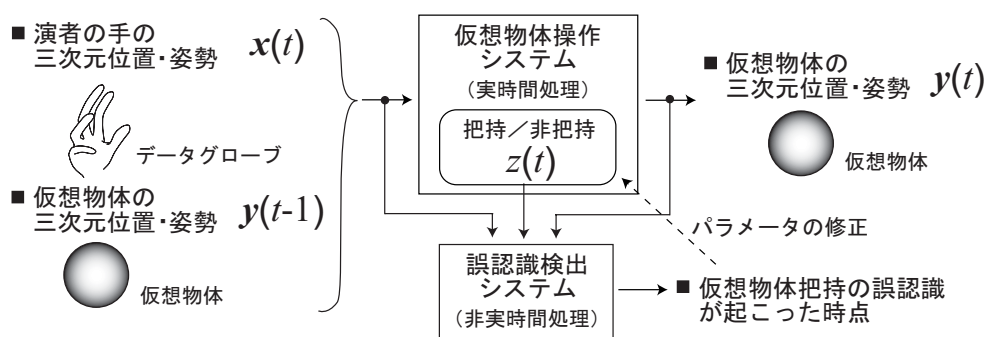


図4: 仮想物体操作システムと誤認識検出システムの関係

インタラクティブバーチャルスタジオにおいて、演者が仮想物体を直接操作できるためには、演者の手の動きに基づいて仮想物体の動きを実時間で生成するシステム（仮想物体操作システム）がまず必要となる。いま、三次元位置センサとデータグローブから得られる、時刻 t での演者の手の三次元位置・姿勢を $x(t)$ で表し、そのときに表示される仮想物体の三次元位置・姿勢を $y(t)$ で表すと、仮想物体操作システムは $x(t)$ と $y(t-1)$ を入力として $y(t)$ を実時間で出力するシステムとして定式化できる（図4）。この仮想物体操作システムの入出力関係は、演者が仮想物体を把持しているときと把持していないときで次のように異なる。

$$y(t) = G(x(t), y(t-1)) \quad (1)$$

$$= \begin{cases} y(t-1) + (x(t) - x(t-1)) & \text{(把持のとき)} \\ y(t-1) & \text{(非把持のとき)} \end{cases} \quad (2)$$

このため、仮想物体操作システムは各時刻で演者による仮想物体の把持の有無を認識する必要がある。そこで時刻 t においてシステムが認識した把持・非把持の区別を以下のように $z(t)$ で表す。

$$z(t) = \begin{cases} Grasp & \text{(把持のとき)} \\ Free & \text{(非把持のとき)} \end{cases}$$

$z(t)$ は $x(t)$, $y(t-1)$ を評価関数によって評価することで定まる。この評価関

数を F とおき, F において $\boldsymbol{x}(t)$, $\boldsymbol{y}(t-1)$ を評価するための閾値等のパラメータをベクトル \boldsymbol{w} で表すと, $z(t)$ は次のように表現できる.

$$z(t) = F(\boldsymbol{x}(t), \boldsymbol{y}(t-1); \boldsymbol{w})$$

(2) 式により, この $z(t)$ によって $\boldsymbol{y}(t)$ は次のように定まることになる.

$$\boldsymbol{y}(t) = \begin{cases} \boldsymbol{y}(t-1) + (\boldsymbol{x}(t) - \boldsymbol{x}(t-1)) & (z(t) = \textit{Grasp} \text{ のとき}) \\ \boldsymbol{y}(t-1) & (z(t) = \textit{Free} \text{ のとき}) \end{cases}$$

$z(t)$ を正しく認識するための \boldsymbol{w} の最適な値は演者によって様々であるため, \boldsymbol{w} を演者に依らない一律の値に設定すると, F によって定まる $z(t)$ の値が演者の意図したものとは異なってしまふことがある. これが仮想物体把持の誤認識である.

そこで, 個々の演者に合わせて \boldsymbol{w} の値を適切なものに修正することにより, 仮想物体操作システムを個人適応させることを目標として, そのために必要となる教示信号を生成するためのシステムとして, 仮想物体操作システムによる仮想物体把持の誤認識を検出するシステム (誤認識検出システム) を導入することを考える (図4). この誤認識検出システムは $\boldsymbol{x}(t)$, $\boldsymbol{y}(t)$, $z(t)$ の時系列データを継続的に観測し, 仮想物体操作システムが誤認識を起こした際に, それに対応するために演者がとる特徴的な行動に基づいて誤認識を検出する. この結果を仮想物体操作システムへの教示信号として利用し, 仮想物体操作システムを漸進的に学習させることでパラメータ \boldsymbol{w} を調整する. これにより, 把持の誤認識なく演者が仮想物体操作システムを用いて思い通りに仮想物体を操作できるように, 仮想物体操作システム内の評価関数 F を演者へ適応させることができる. 本研究では, このうち誤認識検出システムの実現を目標としており, その具体的な実現方法について以下で述べる.

3.2 仮想物体操作システム

3.2.1 把持認識のための手の把持段階のモデル化

$z(t)$ が *Grasp*, *Free* のどちらであるのかを認識する把持認識処理において *Free* は *Grasp* ではないことを意味しているにすぎないため, *Grasp* の認識が実現できれば十分である. この *Grasp* とは, 演者が仮想物体を把持している状況に相当するが, いま演者は 2.2 で述べたような誤差のある状況で仮想物体把

持を行うため、演者の手が仮想物体に触れていなくても *Grasp* と認識する必要がある場合がしばしば生じる。そこで、単にその時点での手と仮想物体の位置関係を静的に評価して *Grasp* を認識するのではなく、把持に至るまでの手の動きも考慮して *Grasp* を認識する必要がある。

従来から、人間による実演に基づいてロボットに作業教示を行うことを目的として、人間による把持行動を認識する研究がある [4],[5]。その中でも川崎ら [4] は、ロボットにタスクを教示するための人間の把持行動として、仮想環境における人間の仮想物体操作を利用し、このときの把持行動を認識する手法を提案している。この手法では、仮想物体を把持した後に搬送して放すというタスク（以下、“ピックアンドプレースタスク”と呼ぶ）を対象とし、そのタスクを以下に示す 6 段階に分割した上で、それらの連続的な過程として人間による把持行動をモデル化している（図 5）。この 6 段階とは、仮想物体に向かって手を移動させる段階である“move”，仮想物体の近傍で物体形状に合わせて指を開いている段階である“approach”，手を閉じて仮想物体を把持する段階である“grasp”，仮想物体を把持したまま搬送する段階である“translate”，仮想物体を所定の位置に置く段階である“place”，仮想物体から手を放している段階である“release”であり、この 6 段階の後、再び move に戻ることになる。

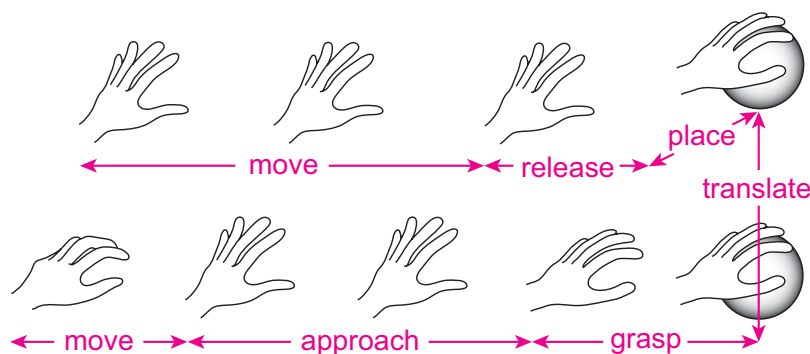


図 5: ピックアンドプレースタスクにおける手の動きのモデル化 [4]

本研究でもこれに従って把持行動をモデル化する。ただし本研究における把持認識の目的は上のようなロボットへのタスク教示ではなく、三次元仮想物体を教材に用いた教育映像撮影である。このことから、教師である演者が教材である仮想物体を所定の位置に置くということを想定する必要はなく、演者は説明のために仮想物体を把持・操作した後そのまま放すことになる。すなわち、

place という段階を考える必要はない．このとき，translate から place に遷移する過程は grasp に戻ることになる．したがって，本研究で扱う仮想物体把持の段階としては move, approach, grasp, translate の後，再び grasp を経て，release に至る 6 段階となる．そこで次に，これら 6 つの段階を仮想物体操作システムが認識するための特徴量について議論する．

3.2.2 各段階の定量的特徴

以下では，時刻 t で仮想物体操作システムが認識した把持の段階を $\omega(t)$ で表す ($\omega(t) \in \{\text{free, approach, grasp, translate, release}\}$) ．

(I) move の特徴

川崎ら [4] の研究におけるピックアンドプレースタスクでは，人間は一貫して仮想物体の組み立て作業を行っているため，move は仮想物体操作のための一連の 6 段階が完了して次の 6 段階を始めるための段階となる．このため，この研究では move を仮想物体や所定の位置に向かって手を移動させる段階と定義している．これに対して，本研究で想定する教育映像撮影においては演者は常に仮想物体を操作しているわけではなく，仮想物体を操作せずに口頭のみで説明を行っている段階も存在し，このときの手の動きは仮想物体操作とは関係ない．このことから本研究では，move を手が仮想物体を掴みに行くことのない自由な段階と定義し直す．このように move を定義すると，move 自体を定量的に特徴付けることは困難である．そこで $\omega(t)$ が他の 5 つのどの段階にも当てはまらない場合に， $\omega(t) = \text{move}$ と認識することにする．

(II) approach の特徴

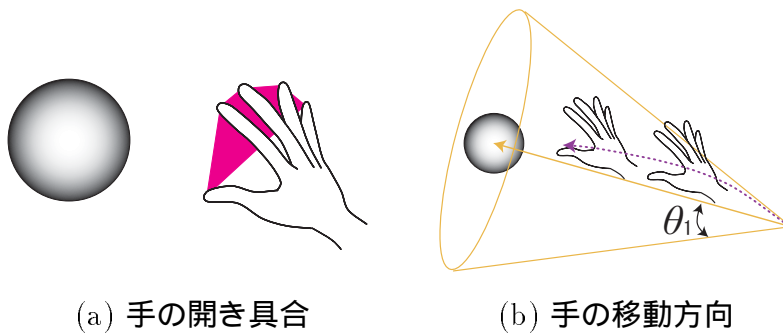


図 6: バーチャルスタジオでの approach の特徴

川崎ら [4] の研究において (I) で述べたように move を手が仮想物体へ近づ

く段階と定義しているため，*approach* は仮想物体の近傍で物体形状に合わせて指を開いている段階と定義しており，手の移動方向については定義していない．一方，本研究では *move* を手が自由な段階と定義したため，*approach* の条件に手が仮想物体のある方向に移動していることも含める．すなわち本研究では，*approach* を仮想物体の近傍で物体形状に合わせて指を開いており，かつ手が仮想物体のある方向に移動している段階と定義する．以上を踏まえて， $\omega(t) = \text{approach}$ と認識するための条件を定義する．

ここで人間の手が物体を掴みに行くときの手の開き方に関する特徴量として，Kang ら [5] は FPA (Fingertip Polygon Area) という特徴量を定義している．これは手の人間の把持行動を認識する際に手の広がり具合を定量的に評価するための特徴量であり，母指，示指，中指の先端を頂点とする三角形の面積 A_1 ，母指，中指，環指の先端を頂点とする三角形の面積 A_2 ，母指，環指，小指の先端を頂点とする三角形の面積 A_3 の総和で定義される (図 7) ．そこで本研究においても *approach* の認識の際に，手の広がり具合を評価するための特徴量として FPA を用いる．

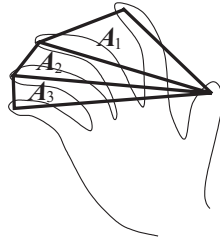


図 7: 手の広がり具合に関する特徴量

時刻 t において $\boldsymbol{x}(t), \boldsymbol{y}(t)$ から得られる，掌の重心の三次元位置を $\boldsymbol{P}_{palm}(t)$ ，掌から仮想物体へ向かうベクトルを $\boldsymbol{D}(t)$ ，FPA を $FPA(t)$ ， $\omega(t) = \text{free}$ である認識された最新の時刻を t_{fr}^E で表すと， $\omega(t)$ を *approach* と認識するための条件は (3) 式，(4) 式のように表現できる．

$$FPA(t) \geq k_1 \quad (k_1 : \text{定数}) \quad (3)$$

$$\frac{(\boldsymbol{P}_{palm}(t) - \boldsymbol{P}_{palm}(t_{fr}^E)) \cdot \boldsymbol{D}(t_{fr}^E)}{\|\boldsymbol{P}_{palm}(t) - \boldsymbol{P}_{palm}(t_{fr}^E)\| \|\boldsymbol{D}(t_{fr}^E)\|} \geq \cos \theta_1 \quad (\theta_1 : \text{定数}) \quad (4)$$

(III) grasp の特徴



図 8: バーチャルスタジオでの grasp の特徴

川崎ら [4] の研究において, grasp は手を閉じて仮想物体を把持する段階と定義されている. これは把持のため手が閉じると考えているためである. この定義は仮想物体として手よりも小さい物体を想定した場合には妥当であり, 実際この研究で人間が操作する物体は手よりも小さいものが想定されている. これに対し, 本研究では仮想物体は教材であるため, 仮想物体は視聴者が見やすいように十分大きいことが望ましい. このような場合には, 仮想物体は手より大きくなるので演者が仮想物体を把持するときには手は開いた後ほとんど閉じないと考えられる. 一方, 演者が仮想物体を把持・操作しているときには当然, 手は仮想物体に触れているはずである. そこで本研究では, grasp の条件として, 手が閉じるという条件の代わりに, 手が仮想物体に触れているという条件を導入する.

ただし, 3.2.1 で述べたように実際には誤差が存在するため, 仮想物体は理想的には演者の手と接触していなければならないものの, 演者の手が物体表面上に存在しない可能性もある. したがって, 各指の先端・掌が仮想物体の表面付近に位置していることを, grasp の条件とする (図 8). すなわち手の親指, 示指, 中指, 環指, 小指をそれぞれ第 1~5 番目の指と考えると, 時刻 t での第 i 指と仮想物体の表面の最短距離をそれぞれ $d_i(t)$ で表し, さらに掌と仮想物体の表面の最短距離を $d_0(t)$ で表すと, $\omega(t) = \text{grasp}$ と認識するための条件は (5) 式で定義される.

$$\forall i \in \{0, 1, \dots, 5\}, d_i(t) \leq \varepsilon \quad (\varepsilon: \text{正定数}) \quad (5)$$

(IV) translate の特徴

川崎ら [4] の研究では, translate を仮想物体を把持したまま搬送する段階と定義している. これに対して本研究においては仮想物体は三次元教材であるので, 演者は説明の際に仮想物体を様々な位置や向きから視聴者に見せるため, これを単に移動させるだけでなく, 必要に応じて教材を回転させる可能性も高い. そこで本研究では translate を仮想物体を把持したまま回転・移動している段階

と定義する．すなわち，手が仮想物体を把持した後に剛体運動していることを `translate` の特徴とする．

ここでは簡単のため，手が剛体運動しているかどうかを手が十分移動しているかどうかで評価する．すなわち， $\omega(t) = \text{translate}$ と認識するための条件は (6) 式のように表現できる．

$$\| \mathbf{P}_{palm}(t) - \mathbf{P}_{palm}(t-1) \| \geq k_2 \quad (k_2 : \text{定数}) \quad (6)$$

このとき，時刻 $t-1$ から時刻 t にかけての手の移動量・回転量を時刻 $t-1$ での仮想物体の位置・姿勢データに加えることによって仮想物体を手の動きに追従させ，演者による仮想物体の移動・回転を実現する．

(V) `release` の特徴

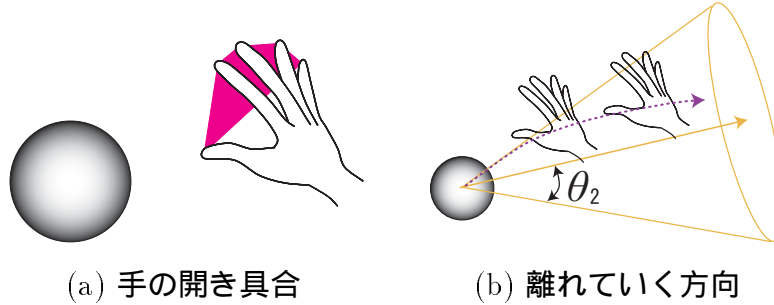


図 9: バーチャルスタジオでの `release` の特徴

川崎ら [4] の研究では，`release` を仮想物体から手を放している段階と定義している．本研究においても `release` を同様に定義する．手が仮想物体から離れているときには手は開いており (図 9(a))，さらに手は物体中心から離れている (図 9(b)) ことが予想されることから， $\omega(t) = \text{release}$ と認識するための条件を次のように定義する．

`grasp` の最初の時刻を t_{gr}^S で表すと， $\omega(t)$ を `release` と認識するための条件は (7) 式，(8) 式のように表現できる．

$$FPA(t) \geq k_3 \times FPA(t_{gr}^S) \quad (k_3 : \text{定数}) \quad (7)$$

$$\frac{(\mathbf{P}_{palm}(t) - \mathbf{P}_{palm}(t_{gr}^S)) \cdot \mathbf{D}(t_{gr}^S)}{\| \mathbf{P}_{palm}(t) - \mathbf{P}_{palm}(t_{gr}^S) \| \| \mathbf{D}(t_{gr}^S) \|} \geq \cos \theta_2 \quad (\theta_2 : \text{定数}) \quad (8)$$

教材である仮想物体が大きいことが原因で，`approach` を定式化する際に用い

た閾値 k_1 で手が開いているかどうかを評価すると, grasp のときに既に手が開いていることが満たされるため, release のために手が開いているという条件が grasp のときから既に満たされる可能性がある. そこで release の場合には k_1 とは別の閾値 k_3 によって, release のために手が開いていることを評価する.

3.2.3 把持認識における状態遷移

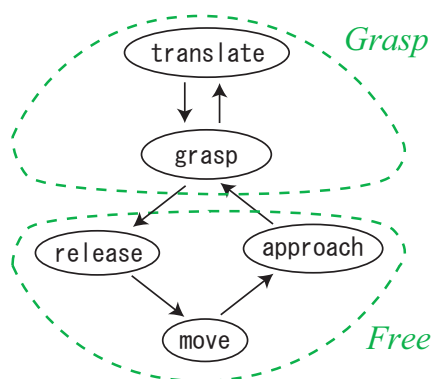


図 10: 仮想物体把持の段階の遷移制約

以上のように演者の手の動きをモデル化し, これに基づいて $\omega(t)$ を定めると, この $\omega(t)$ の値に基づいて, $z(t)$ が (9) 式のように定まる. このとき, $z(t)$ が *Grasp* のときには演者が手を動かすと仮想物体がその手の動きに追従するが, $z(t)$ が *Free* のときには演者が手を動かしても仮想物体は追従せず手だけが独立に動くことになる.

$$z(t) = \begin{cases} Grasp & (\omega(t) \in \{\text{grasp, manipulate}\}) \\ Free & (\omega(t) \in \{\text{free, approach, release}\}) \end{cases}$$

人間の把持動作が 3.2.1 で述べたモデル通りに実行される場合, approach, grasp, translate, grasp, release, free の各段階がこの順序で生じるはずである. すなわち, 図 10 の状態遷移図で示したような段階 (以下, “状態” と呼ぶ) の遷移が生じることになる. ただし, 3.2.2 で定義した特徴量は手と仮想物体の定量的な関係を閾値によって評価するものであるため, その閾値が不適切である場合には把持認識に誤りが生じる可能性がある. 次節ではこのような把持認識における誤りについて議論する.

3.3 誤認識検出システム

3.3.1 演者の反応を利用した誤認識検出

3.2.2 で述べた状態認識処理において誤りが生じ、 $z(t)$ が *Grasp* であるか *Free* であるかが演者の意図とは異なったものとなる時、演者はそれに対して何らかの反応をすると考えられる。例えば、演者が仮想物体を把持しようとした ($z(t) = \textit{Grasp}$ であるべき状況) にもかかわらず、仮想物体操作システムが $z(t) = \textit{Free}$ と認識したときには、演者は手を動かしても仮想物体が追従しないことから把持に失敗したことを知覚し、再び仮想物体を把持しようとするはずである。逆に、演者が仮想物体を把持しているときに放そうとした ($z(t) = \textit{Free}$ であるべき状況) にもかかわらず、仮想物体操作システムが $z(t) = \textit{Grasp}$ と認識し続けたときには、演者は手を動かしたときに仮想物体が追従して動くことから解放に失敗したことを知覚し、再び仮想物体を放そうとするはずである。本研究ではこのような誤認識後に生じる演者の反応を利用して、把持認識における誤りを検出することを考える。

3.3.2 把持認識における誤りの分類

(1) $z(t) = \textit{Free}$ に留まるべきときに $z(t) = \textit{Grasp}$ に遷移する誤り

$\omega(t-1)$ が approach で、時刻 t において演者は仮想物体を把持するつもりはないが、演者の手の掌と指が仮想物体の表面付近に位置するという条件が満たされてしまったために、仮想物体操作システムが $\omega(t)$ を grasp と認識した場合を考える (図 11)。このときには $z(t)$ が *Free* であるべきところを、仮想物体操作システムでは $z(t) = \textit{Grasp}$ へと遷移する誤りが生じる。この場合の誤認識を“誤認識 1”と呼ぶことにする。このとき演者は元々仮想物体を把持するつもりがなかったので、仮想物体をすぐに放そうとするはずである。その演者の反応によって、状態は grasp からさらに release, move と順に遷移する。全体として、「approach → grasp → release → move」という状態遷移のパスが生じるため、誤認識検出システムはこの状態遷移のパスを検出することにより、誤認識 1 を検出することができる。

(2) $z(t) = \textit{Grasp}$ に留まるべきときに $z(t) = \textit{Free}$ に遷移する誤り

$\omega(t-1)$ が grasp で、時刻 t において演者は仮想物体を放すつもりはないが、手の開き具合が grasp の最初の時刻におけるそれよりも大きく、仮想物体の中心から手が退くという条件が満たされてしまったために、仮想物体操作システム

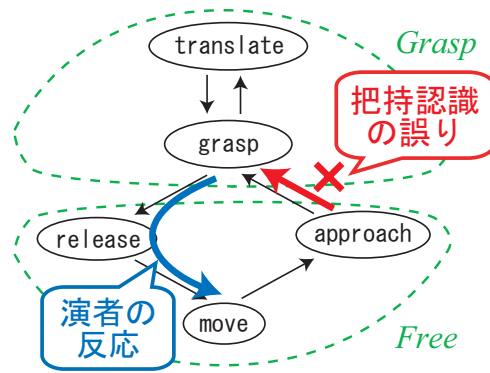


図 11: 誤認識 1 が生じた後の演者の反応

が $\omega(t)$ を release と認識した場合を考える (図 12) . このときには $z(t) = Grasp$ であるべきところを, 仮想物体操作システムでは $z(t) = Free$ へと遷移する誤りが生じる . この場合の誤認識を “誤認識 2” と呼ぶことにする . このとき演者は元々仮想物体を放すつもりがなかったので, 仮想物体をすぐに把持し直そうとするはずである . その演者の反応によって, 状態は release からさらに move を経由することなく approach, grasp と順に遷移する . 全体として, 「grasp \rightarrow release \rightarrow approach \rightarrow grasp」という状態遷移のループが生じるため, 誤認識検出システムはこの状態遷移のループを検出することにより, 誤認識 2 を検出することができる .

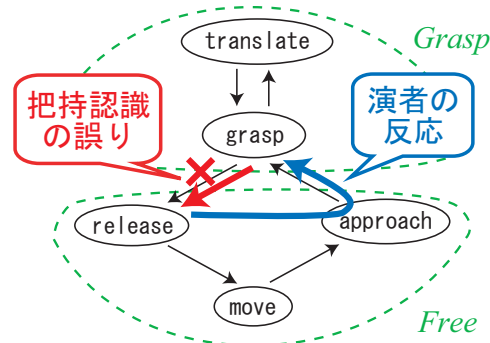


図 12: 誤認識 2 が生じた後の演者の反応

(3) $z(t) = Grasp$ に遷移すべきときに $z(t) = Free$ に留まる誤り

$\omega(t-1)$ が approach で, 時刻 t において演者は仮想物体を把持しようとしているが, 演者の手の掌と指が仮想物体の表面付近に位置するという条件が満たされないために, 仮想物体操作システムが $\omega(t)$ を grasp ではないと認識した場

合を考える (図 14) . このときには $z(t) = Grasp$ に遷移すべきところを , 仮想物体操作システムでは $z(t) = Free$ に留まっているという誤りが生じる . この場合の誤認識を “誤認識 3” と呼ぶことにする . このとき演者は元々仮想物体を把持するつもりであったので , 仮想物体を把持し直そうとするはずである . その演者の反応によって , 状態が *move* 又は *approach* になりながら手が何度も開閉することになる (図 13) . すなわち , 状態が *move* 又は *approach* になりながら (3) 式が何度も満たされたり満たされなかったりする .



図 13: 把持し直す際の手の開閉

そこで誤認識検出システムは , 状態が *free* 又は *approach* のときに手が何度も開閉することを検出することにより , 誤認識 3 を検出することができる .

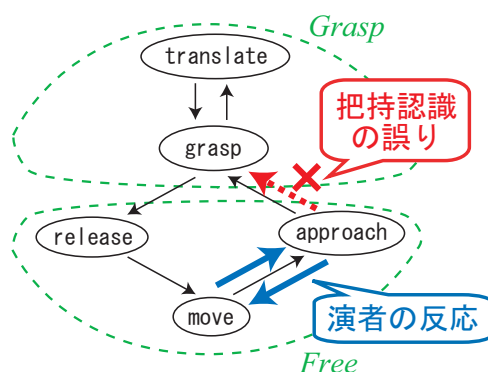


図 14: 誤認識 3 が生じた後の演者の反応

(4) $z(t) = Free$ に遷移すべきときに $z(t) = Grasp$ に留まる誤り

$\omega(t-1)$ が *grasp* で , 時刻 t において演者は仮想物体を放そうとしているが , 手の開き具合が *grasp* の最初の時刻におけるそれよりも大きく , 仮想物体の中心から手が退くという条件が満たされないために , 仮想物体操作システムが $\omega(t)$ を *release* ではないと誤認識した場合を考える (図 16) . このときには $z(t) = Free$ に遷移すべきところを , 仮想物体操作システムでは $z(t) = Grasp$ に留まってい

るという誤りが生じる．この場合の誤認識を“誤認識 4”と呼ぶことにする．このとき演者は元々仮想物体を放すつもりであったので，仮想物体を放し直そうとするはずである．この演者の反応によって，状態が grasp 又は translate になりながら手が何度も開閉することになる（図 15）．すなわち，状態が grasp 又は translate になりながら (7) 式が何度も満たされたり満たされなかったりする．

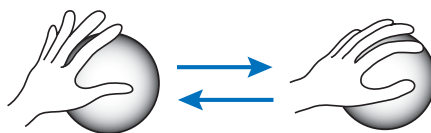


図 15: 放し直す際の手の開閉

そこで誤認識検出システムは，状態が grasp 又は translate のときに手が何度も開閉するという特徴によって，誤認識 4 を検出することができる．

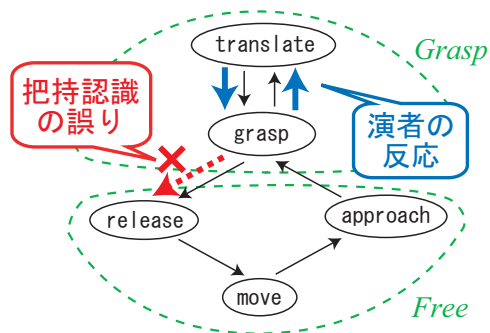


図 16: 誤認識 4 が生じた後の演者の反応

第4章 実験・考察

4.1 実験環境

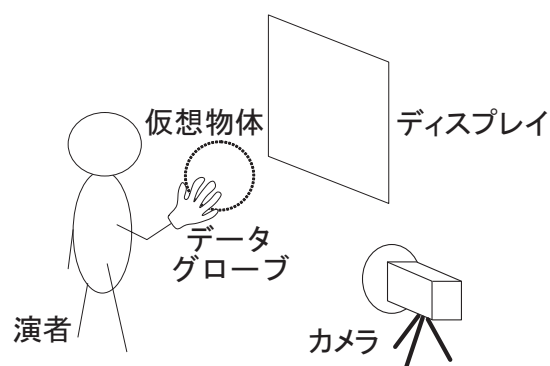


図 17: 実験環境

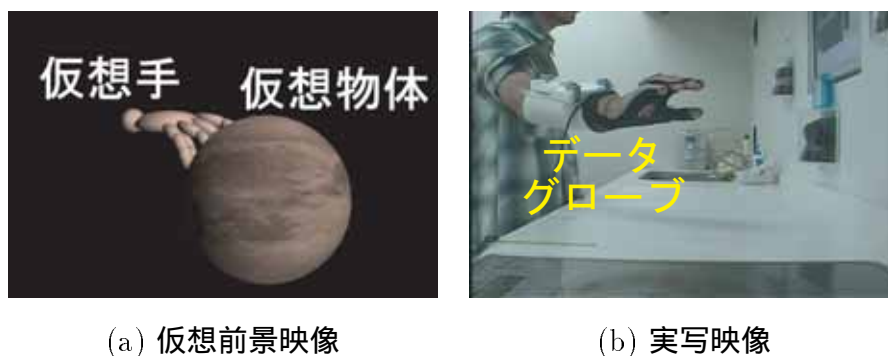


図 18: 実験で獲得される映像

実際に演者が仮想物体に対して操作を行っているときに発生する把持認識の誤りを、第3章で述べたような提案手法によって検出できるかどうかを確認する実験を行った。演者にデータグローブ (Immersion 社製) を装着させ、そのデータグローブで計測される演者の手の動きを表現する。仮想手と、演者が操作する対象となる仮想物体からなる仮想前景映像 (図 18(a)) と実写映像 (図 18(b)) との合成映像がディスプレイ上に表示される環境 (図 17) を構築し、演者にその合成映像を見ながら仮想物体操作タスクを行ってもらった。仮想物体操作タスクとしては、教材である仮想物体を把持してから仮想物体の説明を行った後に放すというタスクを扱い、所要時間は約 37.5 秒 (150 フレーム, フレーム間

隔 0.25 秒)であった．このとき仮想物体は半径 12.5cm の球とし，把持認識に用いる閾値を表 1 のように設定した．実験においては，手を広げながら仮想物体に手を近づけると把持でき，手を広げて仮想物体から手を放すと解放できるということを被験者 3 名（全員右利き）に伝えた後，被験者に上述のタスクをそれぞれ 3 回ずつ行わせた．

表 1: 閾値パラメータの値

パラメータ	値	パラメータ	値
k_1	50.0 [cm ²]	$\cos \theta_1$	0.15
k_2	5.0 [cm]	$\cos \theta_2$	0.15
k_3	1.15	ε	5.0 [cm]

4.2 実験結果

4.2.1 誤認識検出の成功例

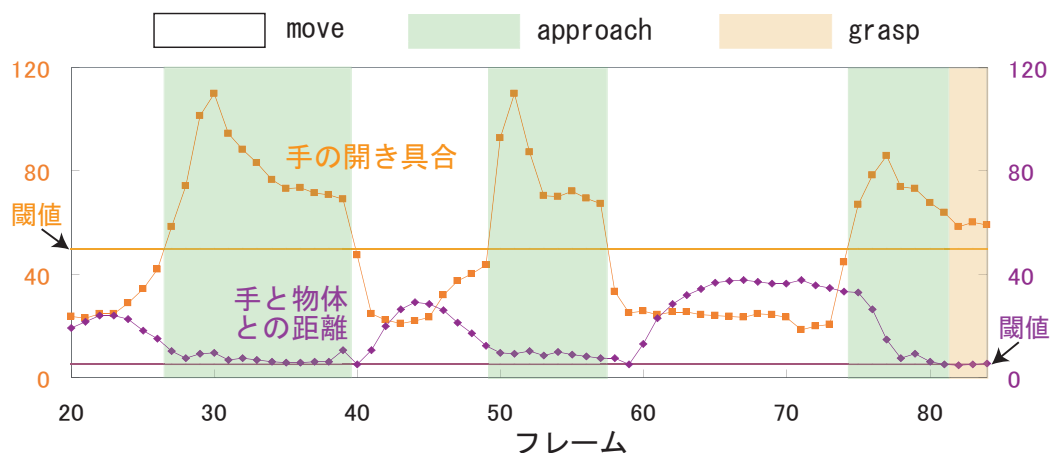


図 19: 仮想物体を把持できないときの演者の反応（被験者 A）

図 19 は，被験者 A が仮想物体を把持しようとしているが，仮想物体操作システムによる把持認識に誤りがあるために把持できず，再び仮想物体を把持しに行っている様子を示している．被験者 A は 20 フレーム目から 39 フレーム目までの間，仮想物体を把持しようとして手を広げながら手を仮想物体に接近させている．しかし手と仮想物体との距離が閾値以下にならないため，仮想物体操

作システムが grasp と認識せず，被験者 A は仮想物体を把持できていない．そこで被験者 A は 40 フレーム目から 57 フレーム目までの間に，再び仮想物体を把持しようとして手を広げながら手を仮想物体に接近させている．先と同様の理由で被験者 A は仮想物体を把持できないので，被験者 A は 58 フレーム目から 81 フレーム目まで再び仮想物体を把持しに行っている．最終的に手と仮想物体との距離が閾値以下になったため仮想物体操作システムが状態を grasp と認識し，82 フレーム以降で被験者 A は仮想物体を把持できている．結局，20 フレーム目から 81 フレーム目までの間に被験者 A は 3 回も仮想物体を把持しに行く行動をとっており，このことから誤認識検出システムは仮想物体操作システムによる把持認識に誤りがあったことを検出できる．

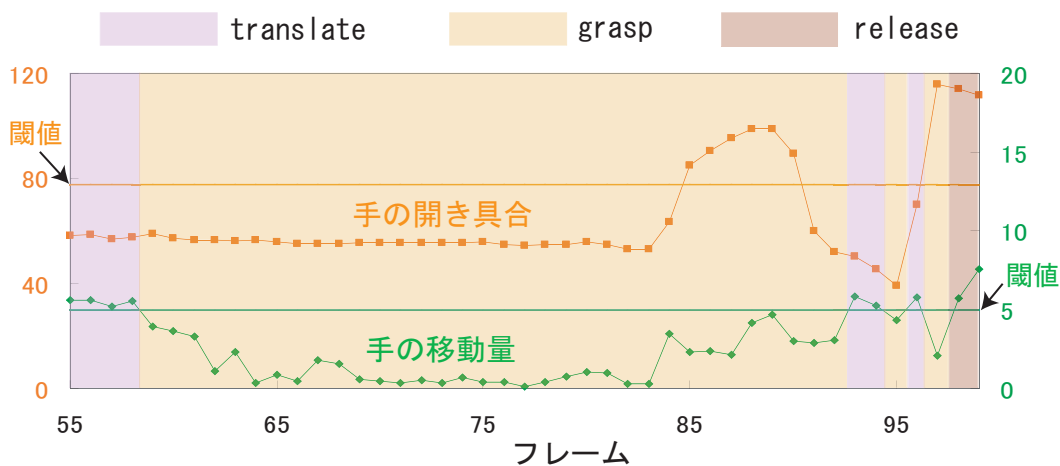


図 20: 仮想物体を放せないときの演者の反応 (被験者 A)

図 20 は被験者 A が仮想物体を放そうとしているが，仮想物体操作システムによる把持認識に誤りがあるために放せず，再び仮想物体を放しに行っている様子を示している．被験者 A は 83 フレーム目から 89 フレーム目までの間，仮想物体を放そうとして手を広げながら手を仮想物体から退けている．しかし手の移動量が閾値以上にならないため，仮想物体操作システムは手が動いていないと解釈して状態を引き続き grasp と認識している．そこで被験者 A は再び仮想物体を放すために，90 フレーム目から 95 フレーム目にかけて手を閉じている．そして 96 フレーム以降で再び仮想物体を放そうとして手を広げながら手を仮想物体から退けている．最終的に，手の開き具合と手の移動量が共に閾値以上になったため仮想物体操作システムが状態を release と認識し，98 フレーム目

以降で被験者 A は仮想物体を放すことに成功している．結局，83 フレーム目から 97 フレーム目までの間に被験者 A は 2 回仮想物体を放しに行く行動をとっており，このことから誤認識検出システムは仮想物体操作システムによる把持認識に誤りがあったことを検出できる．

4.2.2 誤認識検出の失敗例

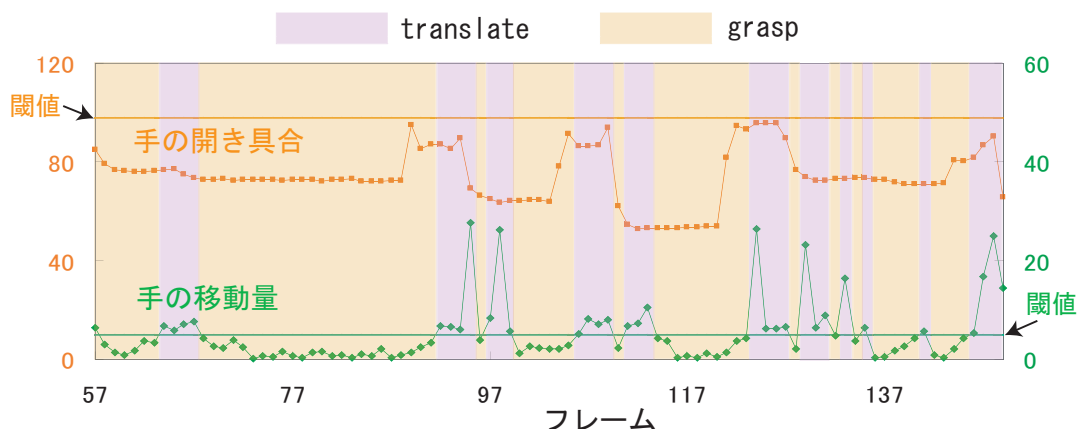


図 21: 仮想物体を放せないときの演者の反応 (被験者 B)

図 21 は被験者 B が仮想物体を放そうとしているが，仮想物体操作システムによる把持認識に誤りがあるために放せず，再び仮想物体を放しに行っている様子を示している．被験者 B は 87 フレーム目から 93 フレーム目までの間，仮想物体を放そうとして手を広げながら手を仮想物体から退けている．しかし手の開き具合が閾値以上にならないため，仮想物体操作システムは手が仮想物体を放そうとしないと解釈して状態を `translate` と認識している．そこで被験者 A は再び仮想物体を放すため，103 フレーム目から 109 フレーム目，120 フレーム目から 126 フレーム目，143 フレーム目から 148 フレーム目の間に手を広げながら手を仮想物体から退ける動きをしているが，仮想物体操作システムは状態を `release` とは認識せず，被験者 B は仮想物体を放すことができていない．これは `grasp` の最初の時刻 (57 フレーム目) における FPA の値が大きすぎたために，続く 58 フレーム以降において (7) 式が満たされないことに依る．このような場合，被験者 B は仮想物体を放せないので何度も放しに行く行動をとっているが，57 フレームから 149 フレームにおいて手の開き具合が閾値を一度も越えることがないために，その間被験者 B は仮想物体を放すために手を開いてはいないと

解釈され、誤認識検出システムは仮想物体操作システムによる把持認識の誤りを検出することができなかった。

4.2.3 考察

仮想物体操作システムは手と仮想物体のデータを定量的に評価するため、状態認識に個人差による影響が生じやすいのに対して、誤認識検出システムでは状態が妥当でない遷移の仕方をしている、又は手が何度も開閉しているなどの個人差の影響を受けにくい定性的な評価をすることによって、仮想物体操作システムによる把持認識の誤りを検出できることが確認できた。ただし、本手法で提案した把持認識のための手の動きのモデルや、それに基づく誤り検出のための特徴量は演者の行う可能な全ての操作を漏れなくとらえているわけではないため、4.2.1で示したように誤認識を正しく検出できる場合だけでなく、4.2.2で示したように把持認識に誤りがあるにもかかわらずそれを検出できない場合も生じた。今後はこのような演者の動きの多様性を考慮してよりロバストな誤認識検出のための特徴量について検討しなければならない。

第5章 結論

本研究では、インタラクティブバーチャルスタジオにおいて演者が仮想物体操作を行う際、演者による仮想物体把持を評価するために用いる閾値等のパラメータを、演者ごとに適応させることを目標として、そのために必要な仮想物体操作システムへの教示信号を獲得するために、仮想物体操作システムが演者による把持操作を誤認識したことを検出する誤認識検出システムの実現手法を提案した。仮想物体操作システムは実時間で動作することが必要であるのに対し、誤認識検出システムは非実時間で動作することにより、把持認識の時点よりも未来における演者の行動を観測することができるため、仮想物体操作システムが演者による仮想物体把持を誤認識した後に演者がとる特徴的な行動を検出することにより、誤認識を検出することができる。

本手法の有効性を検証するために、演者が仮想物体を操作するような状況をタスクとして、被験者3人に仮想物体を把持・操作した後に放すというタスクを3回ずつ行わせた。その結果、実際に誤認識検出システムは何度も仮想物体を掴みに行く、又は何度も仮想物体を放し直すといった特徴的な行動を検出し、仮想物体操作システムの誤認識を検出することができた。しかし、被験者によっては不適切な閾値のため、仮想物体を放し直す際、手が解放のために開いているとは見なされず、それによって何度も放し直すことを検出できない場合も見られた。今後はこのような状況にも対応できるように、状態の判定条件を改善する必要がある。また、本研究ではそれぞれの誤認識によって演者の反応が一意に決まるとして誤認識検出を考えたが、演者によっては本研究で扱った反応とは別の反応を示す可能性がある。これについては、今後実験を重ねて誤認識後に生じる演者の反応にどのような種類があるのかを観測する必要がある。

今後の課題としては、先述のように誤認識検出が検出した仮想物体操作システムの誤認識を教示信号として、仮想物体操作システムを漸進的に学習させることで、仮想物体把持の条件を演者に適応させる処理の実現が挙げられる。第二に、本実験において誤認識後に生じる演者の反応を検出できないことがあったので、誤認識検出に誤りのないよう状態の判定条件を改善することが考えられる。第三に、本研究では演者が三次元仮想物体の任意の面を視聴者に提示できるようにするため、仮想物体操作としては、把持・剛体変換・解放を扱ったが、さらに拡大・縮小操作を導入することで、演者が仮想物体の説明を行うと

きに仮想物体を拡大し，それ以外の時には縮小して目立たなくするといったことを実現することも考えられる．最後に，仮想物体を回転させる際には，仮想手に仮想物体が完全にくっついていることが原因となって，仮想物体の演者側の面を視聴者に提示しづらいといった問題も起きる．そこで，仮想物体の演者側の面を視聴者に提示しやすくなるように，仮想物体を把持したまま指の動きで仮想物体の姿勢を行えるようにすることも考えられる．

謝辞

本研究を進めるにあたり熱心なご指導を賜りました美濃導彦教授，角所考助教授に深謝いたします．英文梗概のチェックならびに校正をして下さいました Anca Ralescu 客員教授に感謝いたします．また，グループミーティングでアドバイスをして下さいましたモデルグループの皆様，ならびに美濃研究室の皆様にご感謝の意を表します．

参考文献

- [1] Gibbs, S., Arapis, C., Breiteneder, C., Lalioti, V., Mostafawy, S. and Speier, J. : “Virtual Studios: An Overview”, *IEEE Multimedia*, Vol. 5, No. 1, pp. 18–35 (Jan.–Mar. 1998).
- [2] 尾原秀登, 角所考, 美濃導彦: “バーチャルスタジオにおける演者の手を用いた仮想物体の直接操作の実現”, 電子情報通信学会技術報告 MVE, Vol. 106, No. 91, pp. 103–109 (2006).
- [3] 椋木雅之, 西口敏司, 美濃導彦: “複数台カメラによる講義の観測と撮像”, 情報処理学会研究報告 CVIM, Vol. 2002, No. 2, pp. 77–84 (2002).
- [4] 川崎晴久, 中山寛治, グラハム パーカー: “VR 環境での人間の動作意図に基づくロボット教示”, 日本バーチャルリアリティ学会論文誌, Vol. 5, No. 2, pp. 899–906 (2000).
- [5] Kang, S. B., Ikeuchi, K. : “Toward Automatic Robot Instruction from Perception – Temporal Segmentation of Tasks from Human Hand Motion”, *IEEE Transactions on Robotics and Automation*, Vol. 11, No. 5, pp. 670–681 (Oct. 1995).