

テレマージョン環境における 3次元ビデオアバタの実時間表現

Real time expression of three-dimensional Video Avatar in Tele-Immersion environment

酒井満隆¹⁾, 小木哲朗²⁾
Mitsutaka SAKAI and Tetsuro OGI

- 1) 筑波大学大学院 システム情報工学研究科
(〒305-0006 茨城県つくば市天王台 1-1-1, mitsu@gil.cs.tsukuba.ac.jp)
- 2) 筑波大学 学術情報メディアセンター
(〒305-8577 茨城県つくば市天王台 1-1-1, tetsu@cc.tsukuba.ac.jp)

Abstract: This study aims at constructing the video avatar communication method in the networked immersive projection displays environment. By representing the three-dimensional image of the video avatar in the immersive projection environment, intuitive communication can be realized. Therefore, the three-dimensional video avatar must be constructed in real-time from the video images captured by the multi-camera system. In this study, we constructed the immersive communication system using the three-dimensional video avatar, and the performance of generating the video avatar was examined.

Key Words: three-dimensional video avatar, visual hull, Tele-Immersion.

1. はじめに

近年、遠隔地の人々とコミュニケーションをとる手段としてビデオアバタを用いる手法が研究されている[1]。ビデオアバタとは、ビデオカメラで撮影している映像からリアルタイムで仮想の人物表現を行う技術である。一般的なCGモデリングソフトで作成したキャラクタとしての人物ではなく、あくまで自分自身の実写像での分身であるところにビデオアバタの特徴があると考えられる。このビデオアバタを3次元表現することで、仮想空間での高臨場感表現が可能になることが期待される。

一方、テレマージョン環境として没入型ディスプレイを用いる研究が行われている。没入型ディスプレイとはイリノイ大学のCAVEを代表とする大画面映像投影システムである。例えばスクリーンを立方体状に配置し、その中でユーザが立体映像を見ることで臨場感の高い仮想空間を体験できる。この没入型ディスプレイとビデオアバタを組み合わせることで、実物大の人物映像を没入型3次元仮想空間に合成して表現できることになる。

この没入型ディスプレイを用いたビデオアバタ通信に関しては、従来までにいくつかの研究が行われている[2][3]。これらの研究では2箇所の没入型ディスプレイ間をネットワークで接続し、種々のビデオアバタモデルを用いた通信実験が行われている。特に筆者らはこれまで3地点間の没入型ディスプレイをネットワークで接続し、2次元ビデオ

アバタを用いた遠隔コミュニケーションの実験を行ってきた[4]。しかし、ここで考えなければならない問題として没入型ディスプレイ内でのビデオアバタの表現方法があげられる。1対1通信までならお互いに正面を向いていればよいので、2次元のビデオアバタ表現でも問題は無い。しかし3地点以上の没入型ディスプレイ間におけるコミュニケーションを考えた場合は、お互いの位置関係次第では相手をいろいろな方向から見るが必要になるため、2次元ビデオアバタによる正面映像だけでは不十分になる。

そこで、3次元のビデオアバタ表現手法の確立が要求される。しかし、3次元ビデオアバタの生成はその処理に大きな負荷を要するため、リアルタイムでのコミュニケーションに用いることは困難であった。本研究ではリアルタイムで3次元ビデオアバタを構築し、複数の没入型ディスプレイに遠隔配信することを目的としている。カメラで撮影している人物像あるいは記録された人物像を3次元ビデオとして配信できれば、3次元仮想空間の中での相手に対する直感的な認識が可能となる。本論文では特に、没入型ディスプレイで3次元ビデオアバタの呈示を行う際、いかに実時間での動作が行えるかに焦点をおきパフォーマンスの検討を行った。

2. 3次元ビデオアバタの構築

本研究における3次元ビデオアバタ構築の流れは大きく

以下の4つの工程で行われる。

人物像の撮影 → シルエット像生成 → 3次元形状復元
→ レンダリング

人物を取り囲むように配置したビデオカメラそれぞれの映像からシルエット像を生成し、そのシルエット像をもとに視体積交差法で3次元形状復元を行い、任意視点映像としてレンダリングを行う。

2.1 人物像の撮影

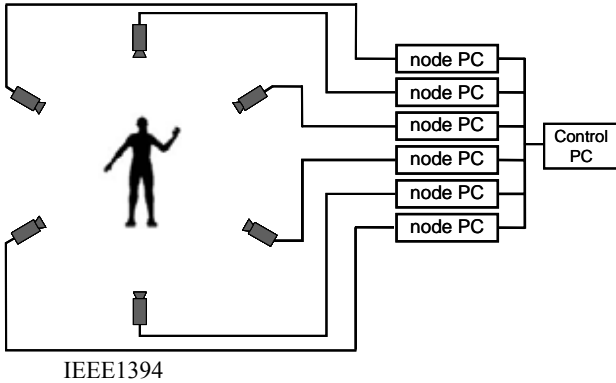


図 1 システム構成図

人物の3次元形状を復元するためには、複数のカメラで人物の全周囲映像を撮影する必要がある。そこで、本研究では図1のような撮影システムを構築した。使用したビデオカメラはIEEE1394カメラ SONY DFW-X710で、解像度1024×768[pixel]、フレームレート15[fps]で1フレーム単位に映像を獲得することができる。各カメラは半径2mの円周上に等間隔に配置し、ノードPCと1対1に接続した。各ノードPCはGigabit EthernetでコントロールPCと接続した。

2.2 シルエット像生成

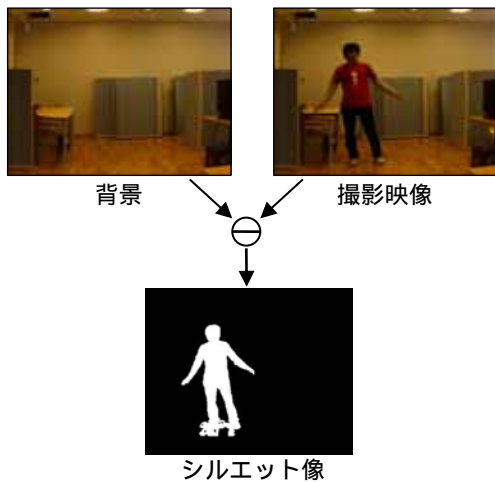


図 2 シルエット生成

視体積交差法で3次元形状を復元するためには、対象のシルエット像を得る必要がある。リアルタイムで対象人物以外の背景領域を除く代表的な手法として、ブルーバックを用いる方法がある。この方法はクロマキーを用いることで、比較的きれいな処理を行うことが可能であるが、背景

条件を特殊な環境に整えなければならない。

そこで、本研究では背景差分法を用いて対象とする人物のシルエット像を生成した。背景差分法とはあらかじめ撮影しておいた背景画像と、同じ環境下で撮影した撮影映像の差分から目的とするシルエット像を抽出する方法である。図2は、1台のカメラ映像でのシルエット像の生成を示した例である。

2.3 3次元形状復元

物体全周の3次元形状を復元する手法として、ステレオマッチングやレーザを用いる手法がある。しかし、ステレオマッチングを用いる手法では物体の色情報を用いるので光学的な環境要因に強く影響を受けてしまう。また、レーザを用いる手法ではレーザ光が照射されている部分の物体形状しか獲得できず、計測にある程度の時間を要する。

そこで、本研究では3次元形状復元に視体積交差法を用いている。その理由としては、複数のカメラで共通に観測される部分のみしか形状復元できない等の制約をもたず、任意台数のカメラで物体全周の近似的な形状を復元できる等の利点があるためである。視体積交差法の基本的な考え方はシルエット制約に基づいている。これは、「対象は任意の視点から撮影して得られる物体の2次元シルエットを実空間に投影して得られる錐体(視体積)の中に含まれる」という制約条件である。この制約を多視点画像で得られたシルエットに拡張すると、「複数の視点に対応する視体積の共通部分の内側に対象が存在する」ということになる。この共通部分は、*visual hull* と呼ばれ、その内部に対象が存在することを意味する。

この*visual hull*で3次元ビデオアバタの形状復元を行う具体的な方法を図3で示す。本研究では3次元オブジェクトをボクセルで表現している。図3では2つの投影中心よりそれぞれの視体積をボクセル空間に投影している。そして、2つの視体積の共通領域(図3では中央太線部分)が*visual hull*となっている。このボクセル空間に投影される視体積が図2のような人物のシルエットであれば、*visual hull*は対象とする人物の概形を表すことができる。また、投影中心、つまりカメラ台数が多いほどより正確な形状を表現できることになる。

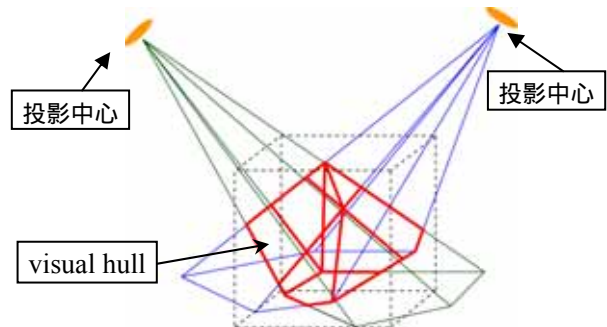


図 3 視体積交差法例

2.4 レンダリング

視体積交差法の処理後、対象とする人物のボクセルモデ

ルをレンダリングした結果を図4に示す。ここでは6台のカメラを用いており、各カメラの視点方向から得たシルエット像による視体積の共通部分が対象とする人物の形状として表現されているのがわかる。



図4 レンダリング結果

また、構築した3次元オブジェクトに色データを付加する方法としてテクスチャマッピングを施したのが図5である。3次元オブジェクトに2次元画像をテクスチャマッピングするには、OpenGLによる投影マッピングを用いた。投影マッピングとはスポットライトを当てるように、物体表面にテクスチャを投影するマッピング方法で、粗いポリゴンにも詳細なテクスチャをはりつけることができる。



図5 テクスチャマッピング

3. パフォーマンス検証実験

本研究では没入型ディスプレイのネットワーク環境で、3次元ビデオアバタを用いた実時間遠隔コミュニケーションを行うことを目的としている。そのため、実時間で動作をするために、使用する複数台のPCの計算負荷をいかに効率的に分散できるかが大きな課題となる。また同時にネットワーク上を流れるデータ量の大きさについても考慮する必要がある。

3.1 実験設備

図1の撮影システムに加えクラスタPCを用いた没入型ディスプレイを使用した。没入型ディスプレイは筑波大学の3面スクリーン構成のCAVE型システムCS Galleryを用いた。図1における6台のノードPCのスペックはIntel Pentium4 3.0 [GHz] CPU、OS: Fedora Core4である。またコントロールPCはIntel Xeon 2.66 [GHz] CPU×2、OS: Red Hat Enterprise Linux WS4 (EM64T)である。CS GalleryのクラスタPCは3台で構成され、それぞれIntel Core2 Duo 2.66 [GHz] CPU、NVIDIA Quadro FX3500、OS: Fedora Core4である。またカメラからの映像は、1024×768 [pixel]の無圧縮

データとして扱った。

3.2 各処理の負荷検証

3次元ビデオアバタ処理の大きな流れは2章で示した。2章におけるの処理は没入型ディスプレイのクラスタPCが担当することになる。複数台のPCで構成されるシステムで処理を行うとき、負荷を分散させるためにはとの処理に着目することになる。そこで、シルエット生成処理および3次元形状復元のためのボクセルデータ生成処理それぞれの処理速度を計測した。

表1 処理速度比較(単位: fps)

シルエット生成	ボクセルデータ生成
24.2	18.3

表1はカメラ6台分のデータをコントロールPC1台に集め、2つの処理を行った際のパフォーマンスを表している。カメラ6台分のデータを扱った場合、各処理は使用しているIEEE1394カメラの映像取得のフレームレート15[fps]を超えており、またシルエット生成よりボクセルデータ生成処理の負荷のほうが高いことがわかった。しかし2つの処理時間を合計するとカメラの映像取得フレームレートを超えてしまうため、2つの処理を計算機間で分けて行うこととした。

3.3 没入型ディスプレイ表現時の検証

次に没入型ディスプレイで3次元ビデオアバタをレンダリングするのに、どの程度のパフォーマンスで動作するかの検証実験を行った。各PCの処理分散は3.2節より次の2通りでの検証とした。

方法1 ノードPCは映像撮影のみ、コントロールPCでとの処理を行う。

方法2 ノードPCは映像撮影との処理まで行い、コントロールPCでの処理を行う。

処理を分けることでネットワークを流れるデータも異なってくる。ノードPCとコントロールPC間については、方法1のときは6台分の映像データの転送のみ。方法2のときは6台からのシルエットデータと選択されたカメラ1台分の映像データの転送が別々に行われる。コントロールPCとクラスタPC間には方法1も方法2も同様にボクセルデータの送信とテクスチャデータの転送が別々に行われる。

3次元ビデオアバタの呈示には構成するボクセルサイズを1辺2[cm]、ボクセル数は120×120×120と設定した。

3.4 実験結果

表2 データ転送速度の検証結果(単位: fps)

ボクセルデータ		テクスチャデータ	
方法1	方法2	方法1	方法2
7.91	12.2	33.2	34.1

表2にコントロールPCからクラスタPCへの通信までを含んだデータ生成速度の検証結果を方法1と方法2について示す。ボクセルデータ転送に関して、方法1は3.2節におけるとの処理が、方法2はの処理を含んだパフォーマンスとなる。

また、表3はノードPC6台からコントロールPCへのデータ転送までを含んだパフォーマンスを示している。

表3 コントロールPCまでのデータ転送(単位: fps)

方法1 映像データ	方法2 シルエットデータ	方法2 テキストチャデータ
12.5	15.0	48.5

表4は3次元ビデオアバタを仮想空間に合成した映像を没入型ディスプレイに描画するときの平均描画フレームレート値である。また、図6は実際に実時間で描画している様子を示したものである。

表4 平均描画フレームレート値(単位: fps)

方法1	方法2
8.38	9.01

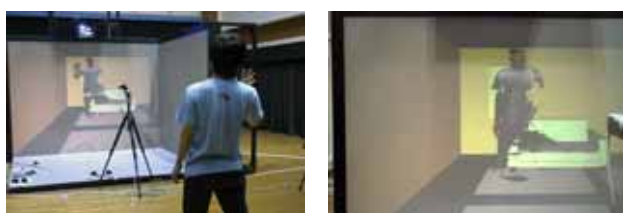


図6 没入型ディスプレイ描画の様子

3.5 考察

表2より、方法2のほうがボクセルデータ送らまでのパフォーマンスがよい結果になった。また、表4よりクラスタPCでの描画パフォーマンスまで含めても方法2のほうがよかった。処理を分けることで差が出る要因としては、3.2節で挙げたようにコントロールPCの負荷が軽くなるからと考えられる。しかし、使用しているIEEE1394カメラの映像獲得フレームレート値が15[fps]であるため、完全な実時間動作を考えると描画フレームレートの結果としてはまだ向上の必要があることがわかった。

また、表3よりノードPCからコントロールPCまでの通信に関しても方法2のほうが実時間性の高い結果となっている。方法1では15[fps]に満たないため、処理の早い段階で実時間性の動作が失われてしまっている。この原因としてはネットワークを流れるデータ量に大きな差があるためと考えられる。

表5 データ転送の必要帯域(単位: Giga bps)

方法1	方法2
1.13	0.755

そこで、表5はノードPCからコントロールPCへのデータ送信に関して方法1と方法2の違いを示した表である。方法1のときは映像データ(解像度1024×768[pixel]、YCbCr形式の無圧縮データ、15[fps])をノードPC6台分通信していることになる。つまり、 $1024 \times 768 \times 2 \times 8 \times 15 \times 6 = 1.13$ [Gbps]のネットワーク帯域が必要となる。それに対し方法2のときはシルエットデータ(解像度1024×768[pixel]、15[fps])6台分とテキストチャデータ(上記映像

データと同様)をノードPC1台分通信している。つまり、 $1024 \times 768 \times 8 \times 6 \times 15 + 1024 \times 768 \times 2 \times 8 \times 15 = 0.755$ [Gbps]のネットワーク帯域ですむことになる。

表5で方法2のときにデータ量が少ない大きな要因はテキストチャデータの扱い方にある。方法1のときは通信している映像データがそのままテキストチャデータとなった。しかし、方法2のときは色情報の無いシルエットデータを通信しているため、テキストチャデータ通信は別に行わなければならない。本研究ではテキストチャデータを没入型ディスプレイのユーザ視点方向と、撮影しているカメラ方向が最も近い映像データとしている。そのため、選択されたノードPC1台分のデータを扱うだけでよいことになり、通信データ量を抑えることが可能になる。しかし、どのカメラ映像をテキストチャデータとして使用するかはコントロールPCを間にはさんで情報を共有するため、映像に遅延が発生してしまう問題を考慮しなければならない。また、色情報がカメラ台数に依存してしまうため、カメラ間の映像を補間することも考えなければならない。

4. 結論

本研究ではテレマージョン環境において実時間で3次元ビデオアバタを表現することを目的とし、3次元ビデオアバタを表現する処理をどう分散すべきか検討を行った。また、没入型ディスプレイを用いてローカルなネットワーク環境で3次元ビデオアバタの実時間動作確認を行えた。

今後は正確なシルエット像の生成、正確な色情報の付加、各処理の速度向上を行う必要がある。そして、今回ローカルで動作確認を行った3次元ビデオアバタ表現を、多地点のテレマージョン環境で動作させることが課題である。

謝辞

本研究は、情報通信研究機構の民間基礎技術研究促進制度に係る研究開発の一部をしておこなわれた。

参考文献

- [1] 小木, 山田, 栗田, 服部, 廣瀬: 仮想空間共有のためのビデオアバタ技術とその利用法, 日本バーチャルリアリティ学会論文誌, Vol.8, No.1, pp.37-46, 2003.
- [2] Tetsuro Ogi, Toshio Yamada, Ken Tamagawa, Makoto Kano, Michitaka Hirose: Immersive Telecommunication Using Stereo Video Avatar, Proceedings Virtual Reality 2001 Conference, pp.45-51, 2001.3
- [3] E.Lamboray, S.Würmlin, M.Gross: Data Streaming in Telepresence Environments, IEEE Transactions on Visualization and Computer Graphics, Special Issue on Haptics, Virtual and Augmented Reality, IEEE Computer Society Press, 2005.
- [4] 小木, 酒井, 立山, 江原, 宮地: ビデオアバタを用いた多地点間CAVEコミュニケーション, 第39回ヒューマンインタフェース学会研究会「人工現実感」, 2006.6.