

多視点カメラ画像からの高速な ポリゴン形状復元手法を用いたインタラクティブシステム

濱崎 省吾 吉田 裕之 重永 信一

松下電器産業株式会社 九州マルチメディアシステム研究所

Email: hamasaki@qrl.mei.co.jp

あらまし： 最近、多視点シルエット画像から物体形状をボクセルで再構成する視体積交差法が注目され、PC クラスタによる実時間処理システムが開発されている。我々はこれまでに、視体積交差法の高速化手法である境界ボクセル法に MarchingCubes 法を適用し、シルエット画像からポリゴン形状を生成する高速手法を提案している。本報では仮想世界でインタラクションを行うシステムを試作し、境界ボクセル法の有効性を示す。本システムでは多視点カメラ画像から境界ボクセル法で人物動作をポリゴン形状として入力する。入力したポリゴンデータは仮想世界でアバタとして表現され、仮想物体とのインタラクションを行う。4 台のカメラと 2 台の PC で構成するシステムで 25[mm]精度で形状入力した実験では、毎秒 16 フレームのスループットを確認した。

キーワード： 視体積交差法、ボクセル、ポリゴン、形状復元、仮想現実感、インタラクション

An Interactive System using High Speed Polygon Generation from Multiple Camera-Images

Shogo Hamasaki Hiroyuki Yoshida Shinichi Shigenaga

Kyushu Multimedia Systems Research Laboratory, Matsushita Electric Industrial Co., Ltd.

Email: hamasaki@qrl.mei.co.jp

Abstract： View volume intersection method, which perform 3D-Voxel reconstruction from a multiple-silhouettes, receives much attention recently. Before, we proposed a high speed 3D-polygon-shape reconstruction algorithm "Boundary Voxel testing Method"(BVM) . In this paper, we implemented the VR interactive system to show the usefulness of VBM. In this system, the human motion is inputted as polygons from multiple camera-images by using BVM. The polygon data is expressed as ABATOR that interacts with virtual objects in the virtual world. In our system that consists of 4-cameras and 2-PCs, with a rate of 16 frames per second, the system can input the form in the 25mm spatial accuracy.

Key words： View volume intersection method, Voxel, Polygon, Shape reconstruction, Virtual reality, Interaction

1 はじめに

3次元形状復元技術は、CADモデリング、CG制作、医療画像処理、バイオメカニクス等の幅広い分野でニーズが高まっている。従来は接触式センサやレンジファインダに代表される高価な専用ハードウェアによる方式が主流であり、計測範囲も狭かった。近年、Computer Visionの分野では、汎用のCCDカメラで撮影した画像から形状復元するアプローチが盛んである。

一方、我々は人の普段の動作をインタラクティブに計測するシステムの構築を目指している。このシステムでは、人の動作を仮想世界に取り込み仮想物体とのインタラクションを行う機能と、動作を計測する機能とが必要となる。このうち前者のインタラクション機能の実現には、以下の2つの両立が課題となる。

・高精度な3次元形状モデリング

人体の滑らかな表面形状を表現するには、ポリゴンによる高精度なモデリング手法が求められる。これにより、ポリゴンベースで構築された仮想世界への組込みも容易となる。

・実時間処理を実現する高速処理

インタラクティブ性の実現には、高いスループットと低いレイテンシ（処理遅延）が重要である。

近年、PCクラスタを使って視体積交差法[1]を並列処理し、実時間で3次元形状復元を行うシステムが開発されている[2][3]。[2]はポリゴンデータで形状復元を行っているが、実時間処理できているのはボクセルによる形状復元までであり、ポリゴンデータによる形状復元の実時間性については言及されていない。ボクセルを立方体として描画しただけでは滑らかな曲面を表現できず、高精度な人体形状モデリングはできない。[3]は、ボクセルデータに胴体及び四肢に見たてた楕円体をフィッティングさせることで、人物動作を実時間でCG表現している。しかし、楕円体近似では手首や肘など比較的細かい形状の復元には向いていない。楕円体の数を増やすとフィッティングの探索空間が膨大となり解の収束に時

間を要したり、近接する楕円のフィッティング精度が低下するなどの問題がある。

我々はこれまでに、視体積交差法の高速処理手法である境界ボクセル法[4]を提案している。さらに、境界ボクセル法の間接データにMarchingCubes法[5]（以下、MC法）を適用し実時間でポリゴンデータを生成する手法を提案してきた[6]。この手法では25[mm]の分解能で毎秒16フレームの性能である。

本報では、多視点カメラ画像から境界ボクセル法で人物形状を実時間モデリングし、仮想世界アパタとして仮想物体とインタラクションを行う試作システムについて述べる。

2 システム概要

2.1 システム構成

我々が開発したインタラクティブシステムのハードウェア構成を図1に示す。

計測空間の人物を撮影する4台のカメラと、画像処理および仮想世界を制御する2台のPCと、仮想世界の映像をフィードバックするディスプレイで構成されている。

カメラとしては、IEEE1394-based Digital Camera Specification[7]に準拠したデジタルカメラを用いており、撮影した画像は全画素読み出しで非圧縮のままパケットに分割して送信する。PCはAT互換機でそれぞれ2つのCPUを搭載している。うち1台のPCでカメラ画像をキャプチャしている。

ネットワークは、PC間を接続する100MbpsのEthernetと、カメラPC間を接続する400MbpsのIEEE1394を使用している。4台のカメラに接続されたIEEE1394ケーブルはHubで1本に束ねられPCに接続されている。

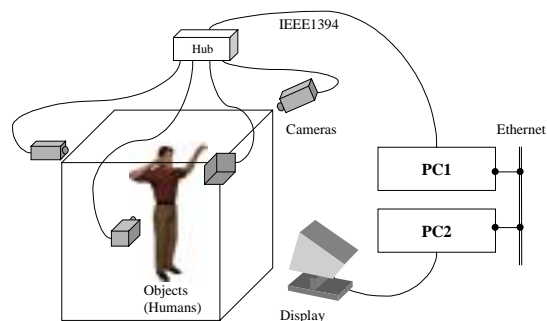


図1 ハードウェア構成

2.2 処理フロー

システムの基本的な処理フローを以下に示す。

1. 同期撮影と画像キャプチャ

計測空間を囲むように配置したカメラで対象となる人物を同時に撮影し、コンピュータに取込む。

2. シルエット抽出

あらかじめ撮影しておいた背景画像との差分から人物のシルエットのみを生成する。

3. ボクセル再構成

視体積交差法により多視点シルエット画像から人物形状をボクセルで再構成する。

4. ポリゴン形状復元

MC法を用いてボクセルデータからポリゴンデータを生成し、人物を形状復元する。

5. 仮想世界制御

ポリゴンデータは仮想世界のシーングラフに追加されアバタとして表現される。また、アバタと仮想物体との衝突判定を行う。仮想世界の映像はディスプレイによりフィードバックする。

以上の一連の処理を実現するモジュール構成を図2に示す。各モジュールについて次の章で詳しく説明する。

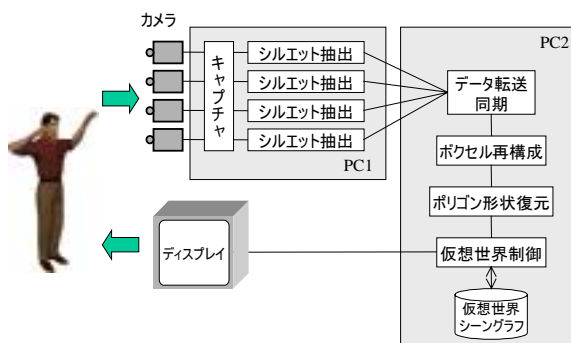


図2 モジュール構成

3 アルゴリズム詳細

3.1 同期撮影と画像キャプチャ

多視点動画像を実時間でキャプチャするためには、大量の画像データを確実に伝送することが重要となる。IEEE1394はデータ伝送の実時間性を重視したアイソクロナス転送モ

ードを持ち、帯域 400Mbps のうち最大 320Mbps をアイソクロナス転送に使用可能で、外部トリガで同期を取って複数台カメラで撮影した動画像データのキャプチャ環境を容易に実現できる[8]。

しかし、IEEE1394-based Digital Camera Specification V1.2 ではキャプチャ画像に対して撮影したカメラや撮影時刻を特定するラベル情報は定義されていない。このため、ボクセル再構成処理前に必須なデータ転送同期（同時刻のフレームが揃ったか否かの判定）の実現は難しい。そこで我々は、下記の方法で画像毎にカメラIDとフレームIDを決定し、キャプチャ時に添付する機能を持ったライブラリを開発することで、この問題を解決した。
カメラID: ドライバ初期化時に1394バス上に接続されている全カメラの台数とカメラが持つデバイス識別子を調べ、ユニークなカメラID(=0,1,2...)を決める。

フレームID: アイソクロナス転送時のパケットのヘッダ部にある同期用ビットSY（各画像データの最初のパケットのみSY=1、それ以外ではSY=0）を監視し、SY=1のパケットをカウントし、フレームIDとする。

3.2 シルエット抽出と同期処理

シルエット抽出の問題は、雑多な背景から高速かつ正確にシルエットを抜き出すことである。あらかじめ人物が含まれていない計測空間を撮影した背景画像と実行時の入力画像との差分からシルエットを抽出するが、画素毎の単純差分ではノイズの影響が大きく、正確なシルエットの抽出は難しい。

本システムでは、ptracker[9]で使用されている方式で、統計的な背景モデルと入力画像とのマハラビノス距離からシルエットを抽出するアルゴリズムを実装した。その処理手順を以下に示す。

1. 複数枚の背景画像から分散共分散行列CVと平均背景画像BFを背景モデルとして作成。
2. 入力画像の各画素 $I_{i,j}$ について次を繰り返す
マハラビノス距離 $dm_{i,j}$ を式1で計算
if $dm_{i,j} < \text{閾値}$ then 画素 $I_{i,j}$ は背景
else 画素 $I_{i,j}$ はシルエット
endif

$$dm_{i,j} = (y - y_0 \quad u - u_0 \quad v - v_0) CV^{-1} \begin{pmatrix} y - y_0 \\ u - u_0 \\ v - v_0 \end{pmatrix} \quad (1)$$

$$\cdot \text{分散共分散行列 } CV = \begin{pmatrix} YY & YU & YV \\ YU & UU & UV \\ YV & UV & VV \end{pmatrix}$$

・平均背景画像 $BF_{i,j}(y_0, u_0, v_0)$

・入力画像 $IF_{i,j}(y, u, v)$

この手法により、単純差分の場合と比較して良好なシルエット画像を得ることができた。

また、カメラ毎にシルエット抽出処理を並列して実行することで高速処理を実現している。各モジュールがそれぞれPC2へシルエット画像をデータ転送しているが、PC間でデータ転送の遅れやデータ落ちが発生する可能性がある。そこで、フレームに添付しているカメラIDおよびフレームIDを使って、データ転送同期処理を行うことで確実な処理を実現している。

3.3 境界ボクセル法によるボクセル再構成

多視点のシルエット画像から物体形状をボクセル再構成する視体積交差法の原理を説明する。(図3)

計測空間をX, Y, Z方向の小さな単位格子(ボクセル)に分割する。ボクセルが空間的に対象物体の内部に位置する場合を占有状態、外部にある場合を空状態と定義する。図Xに示すように、ボクセルをカメラ画像面に射影してシルエットとの重なりを調べる。全てのカメラ画像面でシルエット内部に位置すればそのボクセルは占有状態、それ以外の場合は空状態と判定する。全ボクセルについて判定処理を繰り返すことで、対象物体をボクセルの集合体として再構成する。

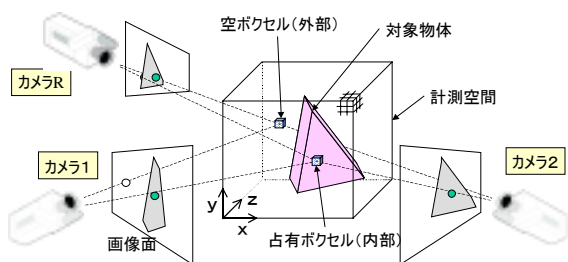


図3 視体積交差法の原理図

視体積交差法では、全ボクセルを処理対象とするため計算に時間がかかり、実時間処理の実現は難しかった。

近年、計測空間を粗い部分空間に再帰的に分割し、空の部分空間を排除することで視体積交差法を高速化する手法が Kanade らによる SPOT[3]、我々による境界ボクセル法[4]として、ほぼ同じ時期に提案されている。これら手法では、部分空間に対象物体が含まれているか否かの状態判定を、正確かつ高速に行うことが重要となる。

SPOT は部分空間を各カメラ平面に射影した凸型領域について、その領域に含まれる画素を一様にサンプリングして物体シルエットに含まれている画素数をカウントし、閾値以上の場合は対象物体と部分空間とが重なっていると判定する手法である。3次元的な重なり判定を、投影した複数の2次元平面上で行っているため、原理的に誤判断する可能性がある。ここで重要となる閾値の設定では、用途に合わせて許容できる誤判断率以下となる閾値を選んでいるに過ぎず、状態判定の正確さには課題が残る。

一方、境界ボクセル法では図4に示す様に、あらかじめ対象物体を既知サイズの関節物体としてモデル化し、最も粗い部分空間の大きさ L_1 を対象物体が折りたたんだ最小状態でも入りきれない大きさとする。このことにより、対象物体と部分空間の重なり判定は、対象物体と部分空間の境界面との交差判定に置換えることができる。さらに、対象物体の最小部品サイズから決定した間隔 d で境界面上のボクセルをサンプリング調査することで、正確かつ高速な状態判定処理を実現している。

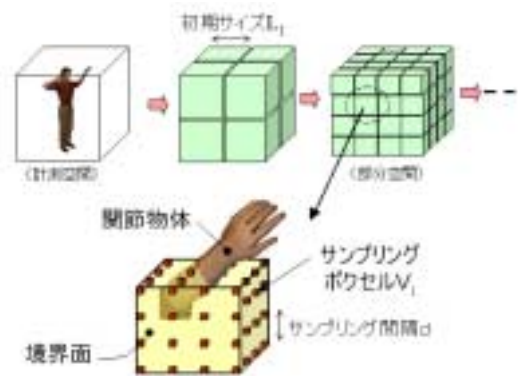


図4 境界ボクセル法の原理

境界ボクセル法の具体的な処理手順を図5に示す。

(1)基本パラメータの決定と初期化

基本パラメータである初期サイズ L_1 とサンプリング間隔 d を決定し、全ての部分空間の状態を空状態で初期化する。

(2)部分空間分割

m 回目に分割して発生する部分空間を P_m とする。最初の分割($m=1$)では、計測空間を初期サイズ L_1 の部分空間 P_1 に分割する。それ以降 m 回目の分割では、次の処理(3)で占有状態と判断した部分空間 P_{m-1} のみをより小さい部分空間 P_m に分割する。この処理を合計 N 回 (N は分割レベル) 繰り返して、最終的な部分空間 P_N がボクセルに一致する。

(3)部分空間の状態判定

ある部分空間 P_m の境界面にあるボクセルのうち間隔 d で間引いたボクセル V_i について、視体積交差法で状態を判定する。 V_i のうち少なくとも1つが占有状態ならば部分空間 P_m は占有状態と判断し、全ての V_i が空状態ならば部分空間 P_m は空状態と判断する。

```

基本パラメータ  $L_1, d$  の決定 ..(1)
全ての部分空間  $P$  の状態を「空」状態とする
計測空間  $P_0$  をサイズ  $L_1$  の部分空間  $P_1$  に分割 ..(2a)
for 分割した全ての部分空間  $P_1$  について do ..(3a)
    if サンプリングボクセル  $V_i$  のうち少なくとも1つは「占有」状態?
        then  $P_1$  を「占有」状態とする
    fi
end
for 分割回数  $m=2 \sim N$  について do ..(2b)
    for 全ての部分空間  $P_{m-1}$  について do
        if  $P_{m-1}$  は「占有」状態?
            then  $P_{m-1}$  を  $P_m$  に分割
        fi
    end
    for 分割した全ての部分空間  $P_m$  について do ..(3b)
        if サンプリングボクセル  $V_i$  のうち少なくとも1つは「占有」状態?
            then  $P_m$  を「占有」状態とする
        fi
    end
end
end
    
```

図5 境界ボクセル法の処理フロー

全ボクセルを処理対象とする従来の視体積交差法と境界ボクセル法との処理量比較[4]によると、従来手法と境界ボクセル法で処理対象となるボクセルの数 T_N と S_N は、

$$T_N = 2^{3N} \quad (2)$$

$$S_N = 3N \cdot 2^{2(N-k)} \quad (3)$$

である。(但し、計測空間は立方体で、再帰的に8つの部分空間に等分し、その分割回数は合計 N 回とした場合。)

処理対象となるボクセル数のオーダーは、従来手法では $O(2^{3N})$ なのに対して、境界ボクセル法では $O(N(2^N)^2)$ まで削減できている。

分割レベル N とサンプリング間隔 $d(=2^k)$ を変化させた場合の式 2,3 の値を図6に示す。分割レベル N が大きいほど境界ボクセル法は従来手法とくらべて処理量が大幅に減少している。また、サンプリング間隔が大きいほど境界ボクセル法の処理量はさらに減少していることがわかる。

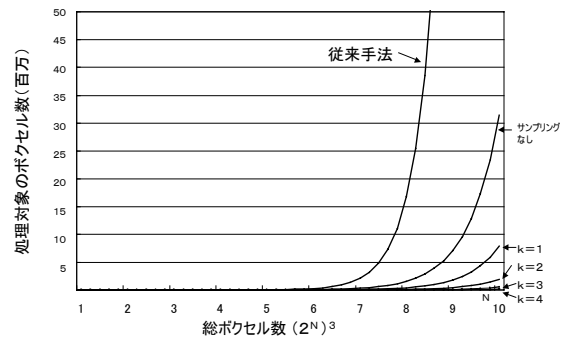


図6 処理対象のボクセル数の比較

3.4 ポリゴン生成

ボクセルデータからMC法を使ってポリゴンデータを生成する。

MC法は、隣接する8つのボクセルの重心位置を頂点とする立方体(キューブ)を想定し、対象物体の表面と交わるキューブ(以下、交差キューブ)を探索する。交差キューブにポリゴンを配置することで、ポリゴン集合体として形状復元する手法である。

MC法は計算量が多く、実時間処理は難しい。1人の人物を対象とした場合、MC法の処理時間全体の約75%を、全体の僅か数%である交差キューブの探索に消費しているという調査結果がある[6]。このことから実時間処理の最大障害要因は、全てのキューブに対する交差キューブの探索処理であると考えられる。

そこで我々は、境界ボクセル法の内部データであるOctree構造にMC法のアルゴリズムを適用することによって、ポリゴン生成の高速化を行った[6]。

つまり、全ボクセルからOctree構造により占有ボクセルを高速に探索し、その占有ボクセルが関与するキューブについてのみ交差判定を行う。こうすることで、大多数を占める空キューブを処理対象から除外することができる。従来手法と比較において、分割レベル7で約4倍の性能を実験的に確認している。

3.5 仮想世界コンテンツ

仮想世界は3種類のオブジェクトで構成されている。

アバタ：ユーザの化身であり、ポリゴン生成モジュールによってフレーム毎にモデリングされる。

目標オブジェクト：小動物(蝶)のオブジェクトで、羽ばたく動きを音と映像で表現する。

背景オブジェクト：公園にある道、木々、池などを構成し臨場感を演出する。

仮想世界で表現された公園にユーザはアバタとして入る。アバタの周囲には、ランダムに蝶が飛来する。蝶はパタパタという効果音とともに一定時間出現する。その間にアバタは身体の一部を蝶に接触させて捕獲し得点を得る。アバタと蝶との接触判定はポリゴン同士の重なりとして判定する。以上のインタラクションをあらかじめ設定した時間繰り返す。オペレータの操作により、任意の視点から映像をユーザに提示することが可能である。(図7)



図7 仮想世界の映像

4 システムアーキテクチャ

4.1 パイプライン処理

システム全体の処理を次の4つのステージに分割して2台のPCでパイプライン処理した。

画像キャプチャ(CA)：全カメラ画像をPCのユーザメモリに取込み

シルエット抽出(SE)：カメラ画像から対象物のシルエット部分のみを抽出

ボクセル再構成(VR)：シルエット画像からのボクセルデータを生成

仮想世界制御(PG)：ポリゴンデータの生成と衝突判定、および仮想世界映像の生成

各パイプラインでは、図8に示すように、異なるフレーム番号のデータを同時並列に処理できる。つまり、フレーム1のSEステージ処理中に次のフレームのトリガが発生しても、フレーム1のSE以降の処理終了を待つことなく、フレーム2の処理を開始できる。

各PCには、2つのCPUを搭載しており、マルチスレッドプログラミングで各ステージの処理を記述することで、OSの持つスケジューリング機能によりCPUの利用率を向上させ、全体のスループット向上が期待できる。

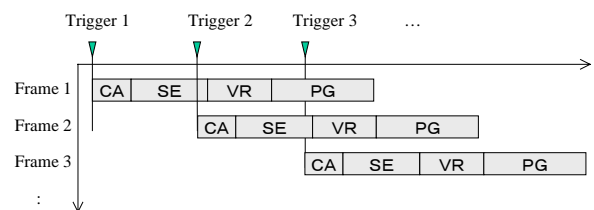


図8 パイプライン処理

4.2 基本パラメータの決定

境界ボクセル法のパラメータを次のように決定した。

計測空間：大人が立位で運動ができる広さを考慮して、一辺が2400[mm]の立方体の空間とした。

部分空間の初期サイズ L_1 ：大人が背を丸めた姿勢から $L_1=800[mm]$ とした。

最初の分割数：各辺を(計測空間)/(部分空間の初期サイズ) = 3分割した。

ボクセルサイズ a ：最初に計測空間を $27(=3^3)$

分割し、以降再帰的に8分割するとして、分割レベル7の部分空間をボクセルと一致させた。この時の部分空間の一辺の長さ、すなわちボクセルサイズは $a=12.5[\text{mm}]$ となる。なお、撮影画像サイズの縦方向は $240[\text{pixel}]$ であり、1ピクセルが対応する計測空間は $2400[\text{mm}] / 240[\text{pixel}] = 10[\text{mm}/\text{pixel}]$ でボクセルサイズとほぼ一致する。よって撮影画像1ピクセルが1ボクセルに対応すると見なしてよい。

サンプリング間隔 d : 手首を最小部品とし、境界面で垂直に切断した最小断面からサンプリング間隔を $d=50[\text{mm}]$ とした。ボクセルサイズ a を考慮して $50[\text{mm}] \times 2^k$ から $k=2$ とした。

5 評価実験

本システムの性能を評価するため、実際の人物をカメラで撮影して仮想世界でのインタラクション実験を行った。

5.1 実験条件

次の条件で実験をおこなった。

背景環境 : 実験室での雑多な静止背景とした。

空間分解能 : ボクセルサイズを 50, 25, 12.5 [mm] とした。(分割レベル $N=5,6,7$)

カメラ : IEEE1394デジタルカメラ 4台

コンピュータ : AT互換機2台

- ・PC1:PentiumIII 850MHz dual 主記憶384MB
- ・PC2:PentiumIII 1GHz dual 主記憶512MB
- ・OS: MS-Windows NT
- ・描画ライブラリ:Cosmo3D, OpenGL

数分間のインタラクション動作を行い、そのうち100フレームについて、各ステージの処理に要した時間と、PGで生成したポリゴン数を計測した。

表1 アバタのポリゴン数

分割レベル N	ポリゴン数 [個 / frame]
5	約 4,000
6	約 12,000
7	約 70,000

5.2 結果

計測結果を図9および表1に示す。CA およ

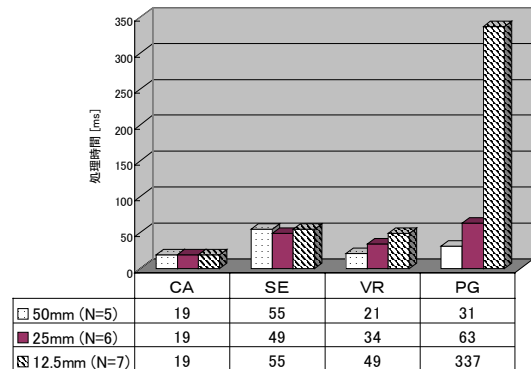


図9 各ステージの処理時間 (単位ms)

び SE の処理時間は分割レベル N には依存していないが、VR および PG は分割レベル N の増大とともに処理時間も増加している。特に PG の処理時間は発生するポリゴン数とほぼ比例している。

また、1フレーム当りのシステムのスループットとレイテンシを表2に示す。4つのステージに分割してパイプライン処理することで、逐次処理した場合と比較して短時間で処理を完了し、スループットが向上できた。

分割レベル $N=6$ ($96 \times 96 \times 96$ 分割; ボクセルサイズ $25[\text{mm}]$) で毎秒 16 フレームの処理を実現できている。この時のレイテンシは $165[\text{ms}]$ であった。同様に $N=7$ ($192 \times 192 \times 192$, $12.5[\text{mm}]$) では毎秒 3 フレーム、レイテンシ $460[\text{ms}]$ の性能となっている。現在のシステム構成では $N=6$ 以上でのボトルネックは PG の処理時間である。PG をさらに複数ステージに分割することでポリゴン数が増えた場合にもスループットを向上が期待できる。

表2 システム性能

N	ボクセルサイズ [mm]	スループット [frame/sec]	レイテンシ [ms]
5	50	18	126
6	25	16	165
7	12.5	3	460

6 まとめ

本報では、多視点で撮影したカメラ画像から実時間でアバタをポリゴンモデリングし、仮想世界において仮想物体とのインタラクションを行うシステムについて述べた。境界ポ

クセル法によるポリゴン形状復元処理は、高速性と高精度な復元精度の両立が求められる仮想世界インタラクティブシステムの実現において有効であることを明らかにした。

システム全体の処理を4つのステージに分割してパイプライン処理することで、4台のカメラで撮影した人物を25[mm]の空間精度(96×96×96分割)でボクセル再構成し、約

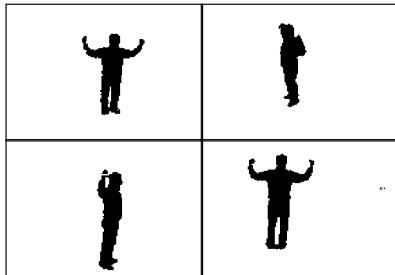
12,000ポリゴンで形状復元して仮想世界映像を生成した場合、毎秒16フレームのスループットと165[ms]のレイテンシの性能が2台のPCで実現できた。

参考文献

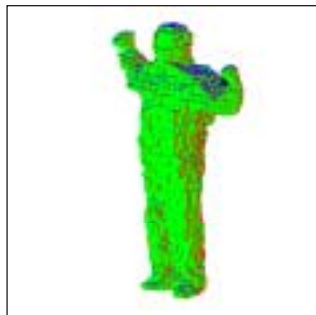
- [1] Saied Moezzi, Li-Cheng Tai, and Philippe Gerard: "Virtual View Generation for 3D Digital Video", IEEE MultiMedia, vol.4, no.1, pp.18-26, 1997.
- [2] 王小軍, 圓藤康平, 和田俊和, 松山隆司: "3次元ビデオ映像の能動的実時間撮影と対話的編集・表示", 信学技報, PRMU2000-187, pp.9-16, 2001.
- [3] German K.M. Cheung, Takeo Kanade: "A Real time System for Robust 3D Voxel Reconstruction of Human Motion", Proceedings of CVPR'2000, pp.714-720, June. 2000.
- [4] 濱崎省吾, 吉田裕之: "部分空間の境界面と関節物体の交差判定による高速3次元再構成", 第6回画像センシングシンポジウム講演論文集, pp.231-236, June. 2000.
- [5] W.E. Lorensen and H.E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm", Computer Graphics, Vol21, No.4, pp.163-169, July 1987.
- [6] 濱崎省吾, 吉田裕之, 重永信一: "多視点シルエット画像からの高速な3次元形状復元手法", 第7回画像センシングシンポジウム講演論文集, pp.59-64, June. 2001.
- [7] 1394 TRADE ASSOCIATION: "1394-based Digital Camera Specification version 1.2", <http://www.1394ta.com/>.
- [8] 吉本廣雅, 有田大作, 谷口倫一郎: "1394カメラを利用した多視点動画像獲得環境", 第6回画像センシングシンポジウム講演論文集, pp.285-290, June. 2000.
- [9] 岡松, 大橋, 乃万, 松尾, 江島: "奥行き情報を利用した人物追跡システム", 電子情報通信学会 PRMU 研究会, PRMU99-210, pp.45-52, Jan.2000.



(a) 撮影画像



(b) シルエット画像



(c) ボクセルデータ



(d) 仮想世界映像(ポリゴンデータ)

図10 ボクセル再構成、ポリゴン形状復元