

# 人と環境への適応技術研究のための人の実時間立体復元と認識システム環境

保呂毅 矢口裕明 岡田慧 稲葉雅幸(東京大学)

## A System Construction for Research of Adaptive Technology for Human and Environment using Real-Time 3D Reconstruction and Recognition of Human Behavior

Tsuyoshi HORO, \*Hiroaki YAGUCHI, Kei OKADA, Masayuki INABA (University of Tokyo)

**Abstract**— In this paper, we construct a Real-Time 3D reconstruction system of human body using visual cone intersection with multicamera, and propose a method of gesture recognition from finger motion detected from 3D human model. We show evaluation about our system using some applications.

**Key Words:** Multi Cameras, User Interface, Visual Cone Intersection

### 1. はじめに

人間の意図をコンピュータに伝えるためのもっとも直感的な方法のひとつとしてジェスチャ認識があげられる。ジェスチャ認識の方法としてモーションキャプチャのような人体に特殊なデバイスを装着する方法が多くとられているが、装着者への負担やデバイス全体の大規模化が問題となる。

また特殊なデバイスを装着する代わりにカメラを用いて人の姿勢を推定する手法もある。この際1台のカメラのみを使う方法と複数台のカメラを用いる方法がある。1台のカメラのみを使う場合、二次元情報から三次元情報の復元を行う必要があり、計算量の増加による実時間性の低下やオクルージョンや狭い視野の影響による姿勢推定の困難さを考慮するとユーザインタフェースとしては好ましくない。

本研究では複数カメラを人間の周囲に配置し、視体積交差法による人間の立体形状を復元したのち、三次元モデルの解析、特にユーザインタフェースとして用いることを考え腕の動きの検出とそれによるジェスチャ認識手法を提案し、実用可能なシステムを構築する。

### 2. マルチカメラシステム構成

本研究で構築したシステムを Fig. 1 に、処理の流れを Fig. 2 に示す。構築したシステムは8台のカメラサーバPCと1台の立体処理用PCからなる。それぞれのPCはLANで接続されており、カメラサーバからは背景差分によるシルエット画像が処理用PCに送信され、処理用PCでは立体復元と認識が行われる。

### 3. 視体積交差法による立体復元

8台のカメラからのシルエット画像を用いた視体積交差法による立体復元を行う。視体積交差法では、まず各々のカメラに対して背景差分によって得られたシルエット画像を三次元空間に投影し、視体積と呼ばれる柱状の立体を復元する。次に各視体積の重複する部分である Visual Hall をもとめ、これに基づいてボクセル空間に投票を行い、立体復元をおこなう (Fig. 3)。

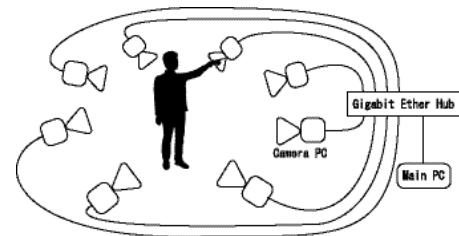


Fig.1 System Overview

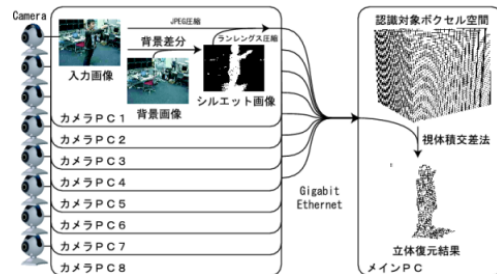


Fig.2 Flow of Process

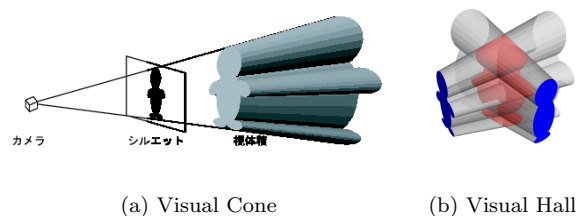


Fig.3 Image of Visual Cone Intersection

### 4. 人物領域の切り出しと複数人追跡

視体積交差法によって得られたボクセルから人物領域を取り出す。まず、人間の体はある一定以上の体積を持った連続体であると考え、空間全体をラベリング

し、しきい値以下の体積のボクセル集合をノイズとして削除する．これにより、二人の人物を個別に検出することができる．

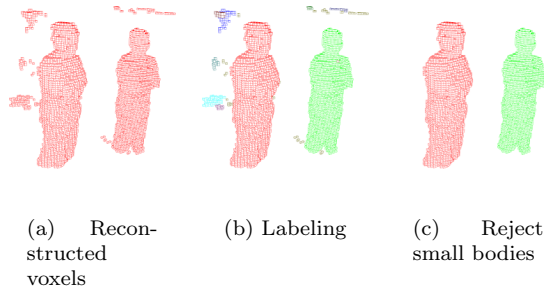


Fig.4 Detect 2 persons from reconstructed voxels

この際問題となるのは、複数人が空間内に存在し、且つ体が触れ合っている状況ではひとつの物体として認識されることである．そこで、本研究では Fig. 5 に示すような方法で領域の分割を行う．

まず、2人が接触する瞬間の1フレーム前が(1)の様に赤と緑の2つの領域に分かれていたとする．そして、接触した直後のフレームが(2)の灰色の領域であるとする．なお、赤や緑で縁取られている領域は、前フレームでのラベルを表している．灰色の領域はすべて接続しているので、ラベリング処理をしても一つの領域とラベル付けされてしまう．そこで、灰色の領域の各ボクセルに対して、前フレームでラベル付けされた一番近いボクセルのラベルを割り当てるという処理を行う．そのようにすると、各ボクセルが(3)の様にラベル付けされ、2つの領域に分けることができる．そしてさらに次のフレームが(4)の灰色の領域であるとする、同様の処理をして(5)の様にラベル付けできる．さらに領域全体の移動量を差し引いて全フレームのラベルを参照することで接触しながら平行移動しても正しくラベル付けを行うことができる．

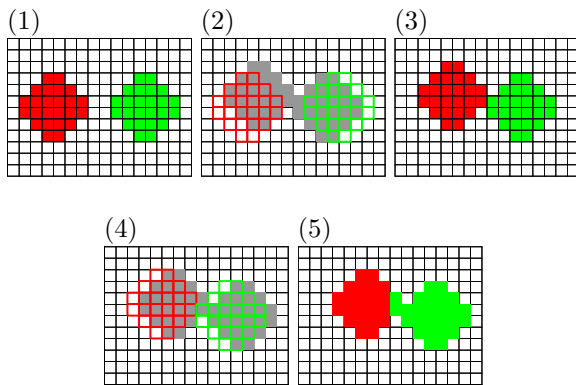


Fig.5 Voxel Tracking

## 5. 指先の認識とユーザーインターフェース

復元された三次元モデルに対し、人体モデルを仮定することで体の各部位の三次元位置を計測する．本研究ではユーザーインターフェースとしての性能を重視し、腕

によるジェスチャの認識に注目する．仮定する人体モデルと指差し認識の過程を Fig. 6 に示す．

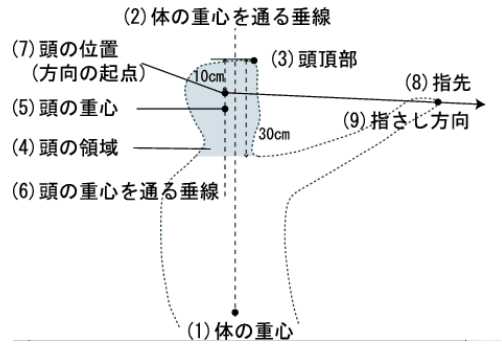


Fig.6 Recognition of Finger Pointing

まず、人体モデルから頭部と指先の位置を計測する．切り出されたボクセルから体の重心(1)を求め、その重心を通る垂線(2)から一定の距離以内における最大高さを頭頂部高さ(3)とし、そこから30cm下方までの集合を頭部とする(4)．頭部重心(5)を通る垂線(6)上の頭頂部高さから10cm下方の点を頭部の位置(7)とする．指先(8)はボクセルの中で体の重心を通る垂線からの最遠点とする．

次に計測された頭部と指先の位置から、指差し方向のベクトルをもとめる(9)．先に計算した頭部の位置は目線の高さを想定しているため、指差し方向は目と指先を結ぶ線分の方向となる．

さらに、指によるジェスチャを認識するために指先軌跡の検出を行う．一般的な成人男性の体格を基に体幹を中心とする半径50cmの円筒を設定し、その円筒から立体がはみ出たときにジェスチャが行われていると仮定し、軌跡の検出を行う(Fig. 7)．本研究ではあらかじめ八方向へのストロークと回転の軌跡(Fig. 8)、および36文字の英数字の軌跡(Fig. 9)をあたえ、検出した軌跡とのDPマッチングによりジェスチャ認識を行う．

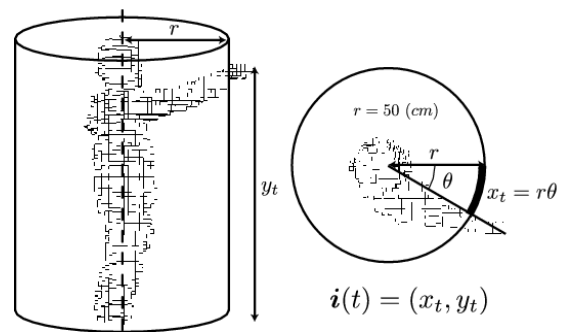


Fig.7 Detection of finger locus

## 6. 実験

### 6.1 複数人物の追跡

複数人物の追跡機能の検証のため、三人の被験者が順番にボクセル空間内に登場し、順番に退場するシチュエーションで追跡を行った．追跡の様子を Fig. 10 に

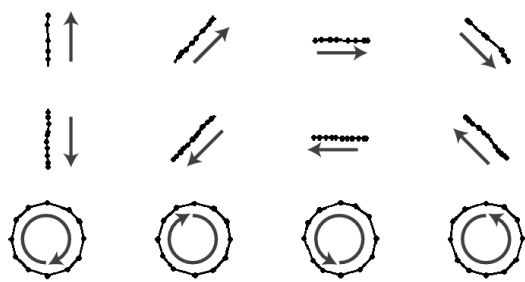


Fig.8 Registered Loci (Direction and Rotation)

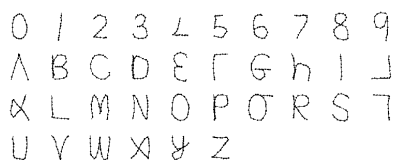


Fig.9 Registered Loci (Alphabets and Numbers)

示す．図中ボクセルのラベリングは色によって表現されている．三人のラベリングと追跡には成功しているが，影の部分に発生する視体積の膨張などによって切り出された各人の立体形状はゆがんでいる．

## 6.2 ジェスチャ認識

ジェスチャ認識機能の検証のために，Fig. 11のようにパネルを設置し，二人の被験者に各パネルに対する指差しを10回ずつ行ってもらい，認識率を測定した．測定結果をTable 1に示す．結果より，人物正面付近のパネルはほぼ確実に認識できたが，端のパネルの認識率が低いことがわかる．特にパネル1における認識ミスではパネル1とパネル2の間に，パネル6および12ではパネル6とパネル12の間に指差しが行われていると認識されているケースが多かった．これは正面付近のパネルは人間から見たときに $20^\circ$ 以上の範囲にあるが，端のパネルは $10^\circ$ 程度しかなく，認識した指差し方向のずれから誤認識が発生していると考えられる．

Panel No.	1	2	3	4	5	6
Recog. Rate	85%	100%	100%	100%	100%	80%
Panel No.	7	8	9	10	11	12
Recog. Rate	100%	100%	100%	100%	100%	95%

Table 1 Result of finger pointing experiments

次に，指先軌跡の検出による文字書き認識実験を行う．三人の被験者に英字26字および数字10字を三回ずつ入力してもらい，認識率を測定する．

軌跡の検出処理は1秒間に約22回実行できた，文字認識の認識率は約96.6%で，数字は95.6%(86/90)，英字は97.0%(227/234)であった．誤認識の結果をTable 2に示す．例えば「L U」というのは被験者はLを入力しようとしたがシステムはUと認識した，という意味である．

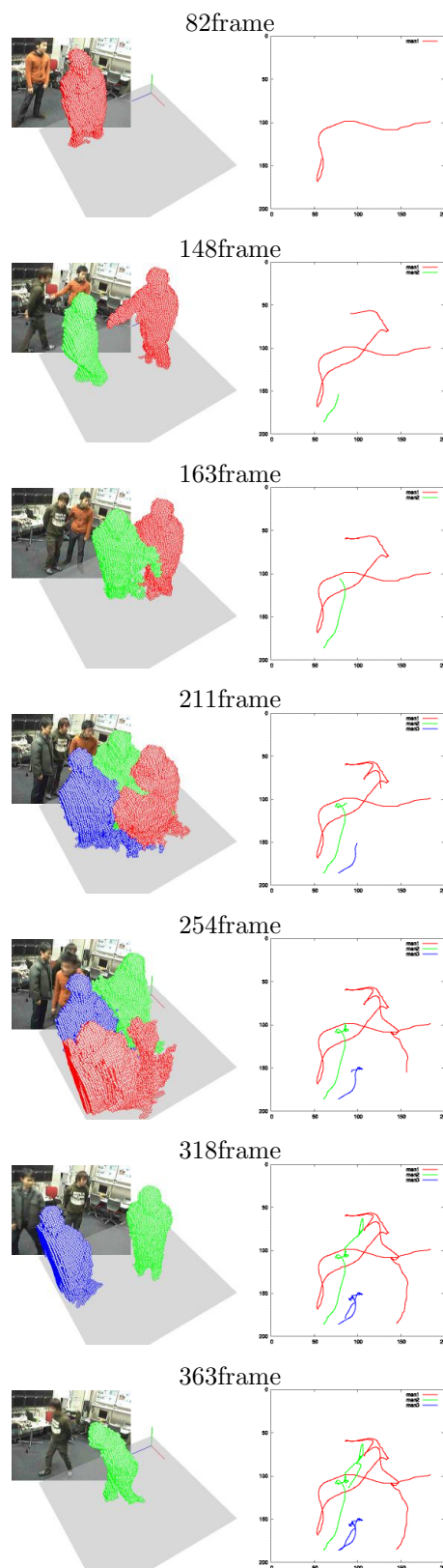


Fig.10 Result of Tracking 3 persons

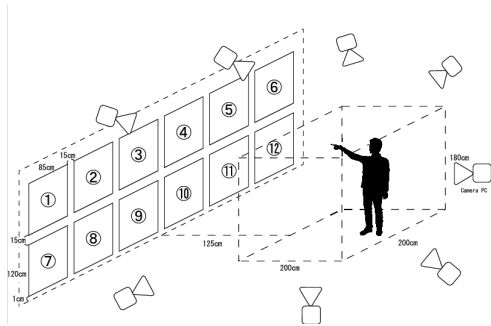


Fig.11 Environment of finger pointing experiments

	Testee 1	Testee 2	Testee 3
Alphabet (Trial 1)	L U, M H	R B	D P
Alphabet (Trial 2)	U V	L U	
Alphabet (Trial 3)			B P
Number (Trial 1)	2 4, 9 5		
Number (Trial 2)		6 0	
Number (Trial 3)		5 9	

Table 2 False cases of Stroke Recognition



Fig.13 Robot System



(a) Remote Control

(b) Ride on Control

Fig.14 Robot Controller

## 7. システムのアプリケーション応用

### 7.1 腕振りによる地図アプリケーション操作

腕振りジェスチャ認識機能を用いたアプリケーションとして、地図アプリケーションの操作を実現した。利用するストロークは Fig. 8 に示したものである。八方向への腕振りを地図のスクロールに、回転を拡大縮小に割り当てた。



Fig.12 Arm Swing Gesture System for Map Application

### 7.2 方向指示によるロボット操作インタフェース

方向指示ジェスチャ認識機能を応用し、ロボット操作インタフェースを作成した。操作対象となるロボットを Fig. 13 に示す。このロボットはオムニホイールを搭載しており、二次元空間の全方位にホロノミックな動作が可能である。ロボットに対する操作として、遠隔操作と搭乗操作の二つの操作方法を実装した。

遠隔操作時は Fig. 8 のストロークを利用し、八方向への腕振りを並進移動に、回転を回転移動に割り当てた。搭乗操作時はジェスチャ認識用に設定された円筒範囲から腕が出た方向としまわれた方向により意図

解釈を行う。腕を進みたい方向に出し、そのまま戻すと腕を出した方向に直進し、腕を出してから回転したい方向に動かし、腕を戻すと動かしした方向に回転する。

ロボット搭乗時のジェスチャ認識における問題として、人体モデルが直立時に合わせて与えられているため、着座姿勢およびロボット自体が含まれることによるモデルの逸脱が発生することがあげられる。このような状況を回避するため、ロボットの高さ基準とした腰の高さより上にボクセル空間を設定することにする。ここでは 70cm と設定した。

## 8. 結論

複数カメラによる視体積交差法を用いた立体復元システムと、それを用いた人体モデルに基づくジェスチャ認識システムの構築を行った。構築したシステムを用いたユーザインタフェースを複数の事例に対して応用し、またラベリングによる複数人数の同時認識を可能とした。今後は人間以外の物体、特にロボットの認識や認識可能なジェスチャの複雑化などインタフェースとしての応用を進めていく。

## 参考文献

- [1] 保呂 毅, 岡田 慧, 稲田 哲也, 稲葉 雅幸, “視体積交差法を用いた実時間ポインティングジェスチャ認識”, 日本機械学会ロボティクス・メカトロニクス講演会 2006, 2006.
- [2] 保呂 毅, 稲葉 雅幸, “複数カメラを用いた手書き文字認識システム”, 日本ソフトウェア科学会 WISS2006, 2006.