

# 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

## Realtime Pointing Gesture Recognition Using Volume Intersection

保呂 毅 (東京大学)  
正 稲邑 哲也 (東京大学)

岡田 慧 (東京大学)  
正 稲葉 雅幸 (東京大学)

Tsuyoshi HORO, The University of Tokyo. horo@jsk.t.u-tokyo.ac.jp  
Kei OKADA, Tetsunari INAMURA, Masayuki INABA

We constructed a pointing gesture recognition system using multi cameras. This system reconstructs the human body from the multi camera images using volume intersection. It finds the human head and hand from the human body voxel data, and estimates the pointing gesture direction. This system doesn't depend on the user's position and the pointing direction. We also constructed a poster explanation system by applying the pointing gesture recognition system. In this system when the user points across at a poster, the explanation of the poster is played.

*Key Words:* Pointing Gesture, Gesture Recognition, Volume Intersection

### 1 はじめに

人と機械を繋ぐインターフェースとして、キーボードやマウスといった人にとって特殊な操作が必要なデバイスを用いるのではなく、音声やジェスチャといった、より人間の自然なコミュニケーションに近く、より直感的な操作を可能にするための研究が数多く報告されている [1] .

人のジェスチャを認識する方法として、人にモーションキャプチャのための特殊な装置を装着する方法と、カメラの画像から人の姿勢を推定する方法がある。前者は、関節の位置や角度を計測することで、比較的正確に姿勢を獲得することが出来るが、特殊な装置を装着する必要があり、人にかかる負担は少なくない。それに対し、後者は人に特殊な装置を装着する必要はなく、より負担の少ないインターフェースだと言える。

カメラを用いてポインティングジェスチャ (指さし) の認識を行うシステムとしては、スクリーンに対して指さしでインタラクションをする試みとして、操作者の横と上部にカメラを設置して認識する方法 [2] や、操作者の上部にステレオカメラを設置して認識する方法 [3] が提案されている。しかし、これらの方法で認識できるのは、スクリーンのような特定の領域・方向に限られており、広い空間での利用は考慮されていない。

そこで、本論文では、ある程度の広い空間内でポインティングジェスチャの認識が出来るように、まず、複数のカメラからの画像を用いて、視体積交差法を用いて人の立体形状を復元し、その立体形状をもとに、ポインティングジェスチャの認識を行う方法について述べる (図 1) .

### 2 視体積交差法を用いた立体復元

#### 2.1 視体積交差法

物体の立体形状を復元する方法として、ステレオ法と視体積交差法が挙げられる。ステレオ法では、複数のカメラ画像内の対応点を元に三角測量の原理で立体を復元する。

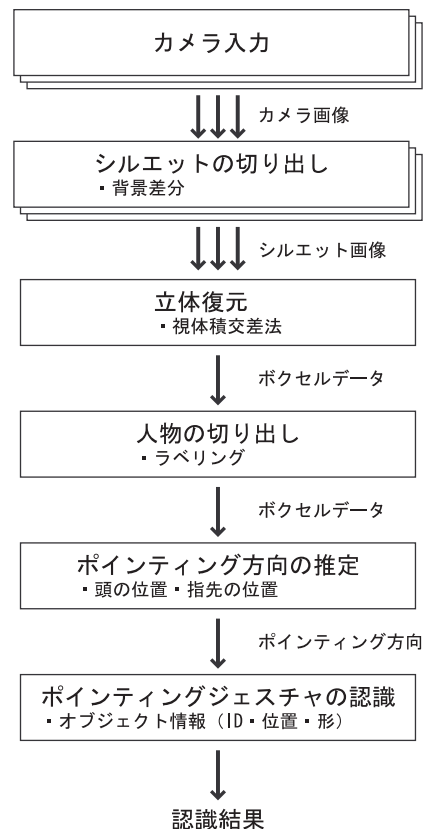


Fig. 1: Process of Pointing Gesture Recognition

山本ら [4] は室内に複数台設置されたステレオカメラを用いて室内の3次元復元を行い、それをもとに腕さしジェスチャの認識を行っている。しかし、ステレオカメラはそれほど一般的ではなく、非常に高価である。それに対し、視体積交差法は、物体のシルエットとカメラの視点から得られる錐体 (視体積) の共通部分である Visual Hall [5] を求

める方法で、比較的安価な普通のカメラを用いて、比較的容易に立体形状を復元できる。しかし、この Visual Hall は対象となる物体に外接する空間であり、物体の立体形状を正確に表現しているのではないという問題がある。とはいえ、凹部分以外の部分は、カメラの台数を増やすことで十分に正確な形状を復元できるので、本研究では、図2の様に人の周りにカメラを複数台設置し、視体積交差法を用いて立体復元を行うことにする。

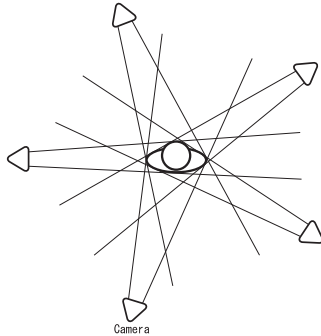


Fig. 2: Silhouette Volume Intersection

## 2.2 カメラキャリブレーション

シルエットから視体積を得るには、カメラキャリブレーションを行い、カメラ座標と世界座標の変換を測定する必要がある。測定する値は、カメラの内部パラメータと呼ばれるカメラの焦点距離と画像中心と画素サイズとレンズの歪み係数と、カメラの外部パラメータと呼ばれるカメラの中心座標と回転行列である。本研究では、定位置に固定されたカメラを用い、カメラの各パラメータは Zhang の手法 [6] を用いたカメラキャリブレーションを事前に行って測定しておいたものを使う。

## 2.3 人物のシルエットの切り出し

カメラ画像から、人物のシルエットを切り出すために、本研究では背景差分の手法を用いた (図3)。事前に背景画像を獲得しておき、背景画像との差が一定値以上の画素を人物のシルエットとして切り出す。



Fig. 3: Background Subtraction(Background · Input · Result)

## 2.4 立体復元

視体積交差法の処理は、認識する対象の3次元空間をボクセルで表現し、シルエット以外の部分をカットしていく、という方法で実装した。具体的には、ボクセルの各点をカメラの投影平面に投影したときに、シルエットの内部がカメラ画像の外部に投影されるときは残し、それ以外の場合、つまり、カメラ画像内のシルエット以外の部分に投影されるときは除去する、という処理をすべてのカメラに対して行う。

人物の周囲にカメラを8台設置し、一辺2cmのボクセルで表現された人物の復元結果を図4に示す。

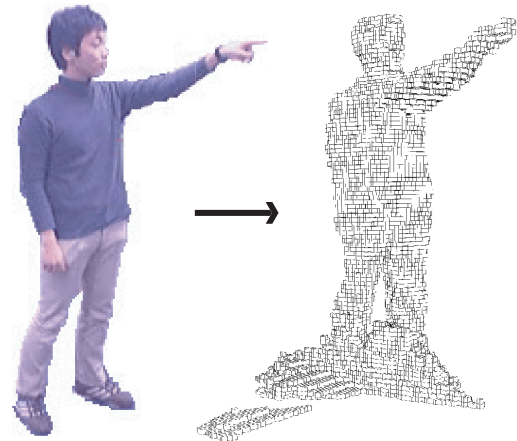


Fig. 4: Reconstructed Human Body

## 3 ポインティングジェスチャ認識

人がポインティングジェスチャを行うとき、肩から指先への延長線を指している場合と、頭(目)から指先への延長線を指している場合の二通りが考えられる。一般的に、対象物が遠方にあるときは頭からの延長線上で、対象物が近くにあるときは肩からの延長線上だと考えられるが、人によってまちまちである。本研究では、頭からの延長線の方が、人が正確に指示できると考え、ポインティングジェスチャの方向を頭から指先への延長線方向とする。

### 3.1 人物の切り出し

背景差分で得られた情報の中に、対象となる人物以外のシルエットが複数写り込んだとき、視体積交差法で得られたボクセルの中に、人物以外のノイズとなる物体が現れることがある。このようなノイズは、カメラにあまり写らない領域に多く現れる場合が多い。これを除去するために、ボクセル全体をラベリングし、一番大きな集合以外の部分を除去する。こうすることで、人物に接触していないノイズを除去することができる。

また、床に写った人物の影を、背景差分で人物のシルエットとして検出してしまっているため、図4にもあるように足元付近に大きなノイズが乗る。ポインティングジェスチャ認識において、足元付近の情報は必要がなく、後の

処理で体の重心を求めるときに邪魔になるので、一定の高さ未満のボクセルは除去しておく。

### 3.2 ポインティング方向の推定

ポインティング方向の推定は、頭の位置と指先の位置を求めることで行う(図5)。

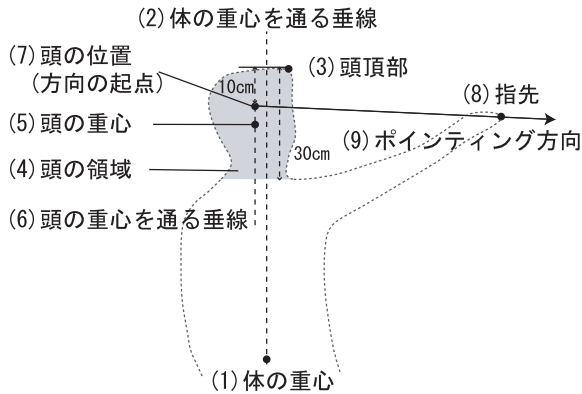


Fig. 5: Estimation of Pointing Direction

頭の位置を求めるために、まず、体の重心(1)を求める。次に、体の重心を通る垂線(2)から一定の距離以内で一番高い部分を頭頂部(3)とする。重心を通る垂線から一定の距離以内で頭頂部から30cm以内の高さ部分を頭の領域(4)として、その領域の重心を頭の重心(5)とする。頭の重心を通る垂線(6)上の、頭頂部の高さから10cm低い場所を、求める頭の位置(7)とする。

指先の位置(8)は、先ほど求めた体の重心を通る垂線からの距離が一番遠いボクセルの位置とする。そして、先ほど求めた頭の位置からこの指先の位置を結んだ延長線上を、ポインティングの方向(9)とする。求められたポインティング方向を図6に示す。

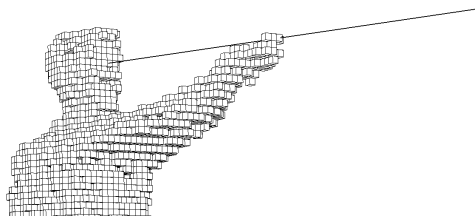


Fig. 6: Estimated Pointing Direction

### 3.3 ポインティングジェスチャの認識

以上の処理で求められたポインティングジェスチャの方向情報を元に、事前に登録されたオブジェクトのうち、何が指されたかを認識する。事前に登録しておくオブジェクトの情報は以下の三つである。

- オブジェクトのID
- オブジェクトの位置
- オブジェクトの形

現在対応しているオブジェクトの形は、長方形のパネル状の物体のみで、ポインティングジェスチャの方向にその長

方形が存在しているときに、そのオブジェクトが指されたと認識する。誤認識をさけるために、0.5秒以上指され続けた時のみ認識するようにしている。

## 4 システム構成

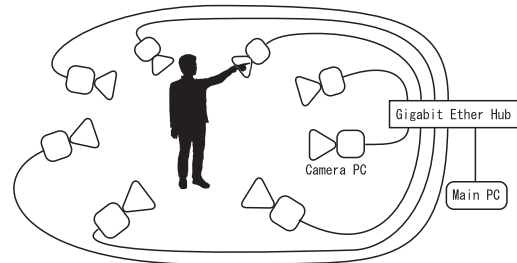


Fig. 7: Pointing Gesture Recognition System

システム構成図を図7に示す。認識する対象の空間の周囲にUSBカメラを接続したカメラPCを複数台設置する。各カメラPCにおいて、背景差分処理を行い、その結果をGigabit Ethernetを通してメインPCに送信する。そして、メインPCにおいて、立体復元を行い、ポインティングジェスチャの認識を行う。

## 5 ポインティングジェスチャ認識実験

人の周囲に設置されたカメラからの情報を元に人の立体形状を復元し、その立体形状をもとに認識を行うことで、ある程度の広さの空間における、あらゆる方向に対するポインティングジェスチャの認識が可能になった。そこで、その精度と処理速度を測るために、以下のような実験を行った。

### 5.1 実験方法

人物の周囲に8台のカメラを設置し、図8のようにパネルを設置し、各パネルに対して10回ポインティングを行い、認識率を測定した。パネルの大きさは幅90cm高さ120cmで、20m間隔で縦2列横6列で12枚設置した。認識の対象となる空間は、150cm×200cm×200cmで、ボクセルの一边は2cmとした。USBカメラはLogitech Quickcam Pro 4000を用い、解像度はQVGA(320×340)で、FPSは15である。カメラPCのCPUはIntel Pentium M 1.3GHz、メインPCのCPUはAMD Athlon 64 FX-57である。

### 5.2 実験の結果と考察

認識の処理はリアルタイムで行い、一秒間に約12回実行できた。認識結果を表1に示す。人の正面付近のパネルはほぼ確実に認識しているが、端のパネルの認識率は低くなっている。これは、正面付近のパネルは人から見たときに30度以上あるが、端のパネルは20度未満しかなく、認識のずれによる誤認識が発生していると考えられる。多少の誤差はあるものの、ある程度の精度でポインティングジェスチャの認識が出来ることが確認できた。

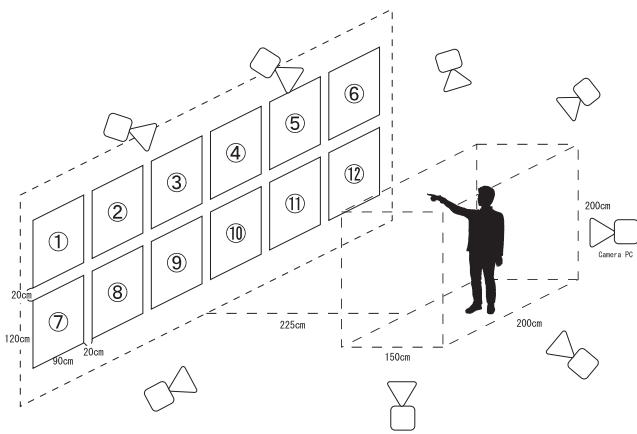


Fig. 8: Experimental System

パネル	1	2	3	4	5	6
認識率	70%	90%	100%	90%	90%	70%
パネル	7	8	9	10	11	12
認識率	70%	100%	100%	100%	100%	80%

Table 1: Result of Recognition

## 6 ポスター解説システム

今回作成したシステムを用いて、ポスターを指さすと、指されたポスターの解説が再生される、ポスター解説システムを作成した。図9は、部屋の周囲にポスターを9枚設置し、このシステムを使用している様子である。正面のプラズマディスプレイには、図10のように、事前に登録しておいたポスター9枚の画像と、立体復元された人の姿と、ポインティングジェスチャの認識結果が表示されている。この画面はOpenGLを用いて作画しており、人の色情報は、複数のカメラの中から、三次元空間を表示する際の仮想空間上のカメラの位置に一番近いカメラの色情報を使用している。

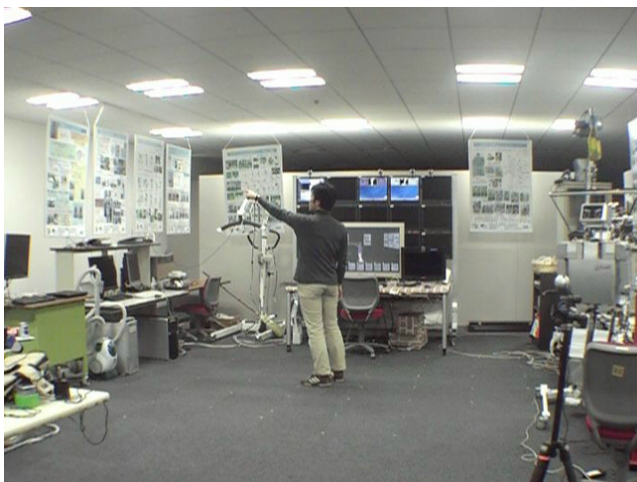


Fig. 9: Poster Explanation System



Fig. 10: Displayed Image

## 7 おわりに

人のポインティングジェスチャを認識する方法として、複数のカメラからの画像を用いて、視体積交差法を行って人の立体形状を復元し、その立体形状をもとに、ポインティングジェスチャの認識を行うシステムを構築した。また、このシステムの認識性能の評価を行い、ある程度の精度で認識が出来ることを確認した。さらに、このシステムを利用して、ポスターを指さすとポスターの解説が再生される、ポスター解説システムを作成した。

今後は、ポインティングジェスチャ以外のさまざまなジェスチャについても認識できるシステムを開発していきたい。

## 文献

- [1] 原島博, 井口征士, 工作舎. 感じる・楽しむ・創りだす 感性情報学. 工作舎, 2004.
- [2] M. Fukumoto, Y. Suenaga, and K. Mase. Fingerprinter: Pointing interface by image processing. In *Computer & Graphics*, Vol. 18(6), pp. 633–642, 1994.
- [3] 豊浦雅貴, 江端真行, 小池英樹, 岡兼司, 佐藤洋一. 3次元指先認識を用いたポインティングデバイスとその応用. 日本ソフトウェア科学会 WISS2004, pp. 133–134, 2004.
- [4] 山本雄, 依田育士, 坂上勝彦. ステレオ距離情報による腕さしジェスチャインタフェース. 信学会 技報 PRMU, 第103(452)巻, pp. 53–58, 2003.
- [5] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 2, pp. 150–162, 1994.
- [6] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, pp. 1330–1334, 2000.