

仮想三次元作業空間における 骨格認識を用いた操作手法に関する研究

Spatial operation using skeletal recognition for virtual 3D work space

齋藤 裕佑*¹
Yusuke Saito

西山 裕之
Hiroyuki Nishiyama

*¹東京理科大学理工学研究科経営工学専攻

Graduate School of Science and Technology, Tokyo University of Science

3D desktop application software for the general environment as well as professional work environment, such as Web services that design room layouts, has been increasing. However, the input method for existing applications requires large-scale facilities and special equipment. This study proposes an operation method using skeletal recognition technology with depth information only. In addition, the method compares placement and specifying rotation tasks with the wheel mouse to verify its operability and to evaluate its efficiency and usability. The efficiency of the method, excluding the adjustment and some usability aspects was found to be good, but the overall efficiency was poor compared with using a mouse because the operation of the space is unstable and the estimated position of the skeleton was inaccurate.

1. はじめに

近年、画面上に三次元空間を表現し、これを利用した作業環境やアプリケーションが充実しつつある。建築現場や機械の設計時におけるモデリングやパーツの合成などに利用される3dsMax*¹や、二次元的な地図と三次元的な周辺環境情報を複合させた GoogleEarth*²などが利用例として挙げられる。

しかしこのような仮想的な三次元作業環境においては、操作に必要な自由度が二次元作業空間に比べ増加してしまう。現在普及しているホイール付きのマウスのような、卓上での平面における操作に限定される場合、立体空間でのオブジェクト移動の際、奥行きか斜め方向への移動操作などの空間的な移動操作ができない。またオブジェクトの向きを操作する際、平面上では回転の中心となる軸をわざわざ選択するなどの措置が必要となり、向きを空間的に直接指定できない。

これらの問題に対し、“g-stalt”[7]などの、空間における操作はいくつか研究されてきた。しかしそれらの研究では必要な部位にマーカーをつけたり、部屋のいくつもの場所にカメラを設置するなどの大掛かりな設備や特殊な手持ちのデバイスが必要であった。そこで本研究では、単一の赤外線センサーデバイスによって物体の深度を取得し、人間の骨格を20部位に分け、それぞれの部位をセンサーからの三次元距離座標を取得することができる技術である骨格認識技術を利用した手法を提案する。コンピュータ上の仮想三次元作業空間に対し骨格認識技術を用い、机などの卓上面がない空間での操作において、利き手の三次元座標推移に対応した、三次元カーソルと二次元カーソルの複合カーソルによる操作を実装する。また有効性に関しては、仮想三次元作業空間のオブジェクトの配置操作及び向き合わせ操作を模したタスクをそれぞれ作成し、ホイール付きマウスとの比較実験を行うことで、その効率性や使用感について検討する。

2. 関連研究

2.1 仮想三次元作業環境に対する 既存のインターフェース

仮想三次元作業環境にて、オブジェクトを操作するためのインターフェースはいくつか提案されてきた。現在のホイール付きマウスの元となるローラーマウス [5]、デバイスレベルで平面移動と回転を分離した Globe Mouse[1] や、赤外線センサー及び傾きセンサーによって、空間にて高い自由度を利用できる Wii Remote*³などが存在する。ここで、仮想三次元空間では、視点やオブジェクト操作に限定しても拡大縮小や回転などの多数のモード切り替えが必要となる。しかしボタン数が限られるデバイスでは、複雑な作業を必要とする仮想三次元作業環境においてモード変更のための拡張性が乏しい。またこれらの研究では性能評価の際、オブジェクト操作のみであり、視点変更が必要な空間での性能評価は行われていなかった。そこで本研究では、ハンドジェスチャによる操作を想定した、空間でのいくつかの操作において、視点変更を含めたオブジェクト操作に関する実験を行い、有効性を検討する。

2.2 ジェスチャによる操作

身体を利用した操作は、近年では盛んに研究が行われている。木村らによる広視野ディスプレイとハンドジェスチャ操作を組み合わせたインタフェースについての考察及び試作 [2] や、大規模な空間でのジェスチャ操作である“g-stalt”[7] などがある。また、Nancel らは大画面における空中での地図閲覧操作において、デバイスの有無や動作方法の次元の違いにおける比較研究を行っており、上記の操作に関して、デバイスを用いない空中でのインタラクションの有効性が示されている [3]。本研究では、マウスの二つのボタンクリックを擬似的なハンドジェスチャと仮定し、骨格認識による空間での操作を実装する。

2.3 骨格認識を用いたインタラクション

近年、骨格認識技術を用いたインタラクションが注目を集め始めている。骨格認識とは、赤外線によるセンサーからの物体の深度情報を用いて、機械学習等によって人間の骨格を認識する技術である。Shotton らによる骨格の部位追跡の研究 [4]

連絡先: 齋藤裕佑, 東京理科大学, j7412611@ed.tus.ac.jp

*1 “3dsMax”, <http://www.autodesk.co.jp/>

*2 “GoogleEarth”, <http://www.google.co.jp/intl/ja/earth/index.html>*3 “WiiRemote”, <http://www.nintendo.co.jp/wii/>

により、30fpsのフレームレートで人間の部位を追跡することができるようになってきている。追跡している様子を図1に示す。骨格認識を用いた研究は現在盛んに行われており、山田らによるプレゼンテーション支援システム [6] などが挙げられる。本研究では手の位置追跡に骨格認識を用い、仮想三次元空間における直感的な操作を実現する。なお、現在深度情報を用いたハンドジェスチャ認識の研究は行われているものの、実用段階には至らないため、その代わりとしてマウスのボタンを利用し、本提案手法においてホイールは使用しない。

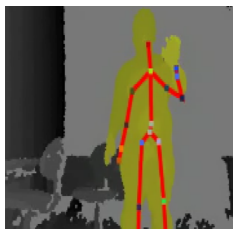


図 1: 骨格認識の様子

3. 設計方針

3.1 実装対象とする操作

仮想三次元空間を利用する製作環境において、使用するソフトウェアによってモードの変更の仕方や、操作方法が変わってしまう場合がある。本研究では、三次元モデリングソフトであるメタセコイア^{*4}を参考にし、以下の視点操作とオブジェクト操作を実装対象とする。

- 視点 (カメラ位置) の位置変更操作
 - － 視点の垂直および平行移動
 - － 視点の奥行きに対する平行移動
 - － 視点の曲線移動
- オブジェクトの位置変更操作
 - － オブジェクトの垂直および平行移動
 - － オブジェクトの向き変更のための回転

以上の動作においてマウスなどの卓上面での操作は平面上に制限されるため、オブジェクトの奥行き移動や調整の際に視点変更を強いられる。例えば、オブジェクトを右斜め上手前へと移動させる場合、開始地点と目標地点が同じ平面上に表示されるまで視点を回転させる必要がある。また向きを指定する場合においても、空間的に直接指定できない。

しかし空間における操作ならば、オブジェクトと目標地点が両方も画面内に表示されていれば、直接移動および向き指定をすることが可能になり、また画面外にあった場合を考慮しても、マウスと比べて視点を操作する時間は短くなるのではないかと仮説を立てた。本研究では、三次元カーソルを手の位置に追従させ、空間的なオブジェクト操作が可能な手法を実装する。

3.2 提案する手法

以上に示した通り、空間でのオブジェクト操作は直感的になり効率性が増すと予想できる。しかし空間での操作においては、視点を変更するための情報が乏しくなることや、安定面がなくブレが発生すること、また空間が把握しづらいなどの問題

*4 “メタセコイア”, <http://www.metaseq.net/metaseq/>

から、三次元カーソルのみでは逆に視点の操作がやりづらくなってしまふ。そこで、三次元カーソルだけではなく、視点移動の際、通常のカーソルと移動のための矢印表示を複合利用することによって、オブジェクト操作及び視点操作の両方に適した操作が可能なシステムの設計を目指す。

本システムの特徴は以下の通りである。また利用する作業空間と、三次元カーソルと二次元カーソルの複合の結果を図2に示す。二次元カーソルが矢印まで到達したとき、視点が動作するようになってきている。

- 三次元カーソルによる、空間でのオブジェクト移動と直接的な向きの指定
- 視点変更の際の補足表示と、二次元カーソルによる視点操作

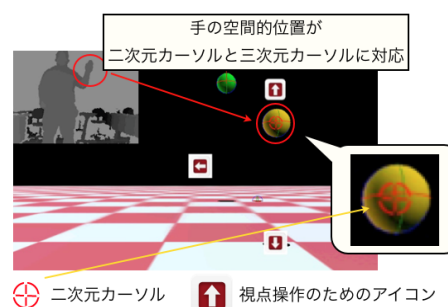


図 2: 複合カーソルの例

3.3 システム構成

先述した操作を骨格認識によって行う際の流れとしては、深度センサーによって取得した情報から骨格を追跡し、手の値に応じて二次元カーソル及び三次元カーソルを配置する。またオブジェクトや視点を操作する際に使用されるモード変更のためのトリガーとしての状態を取得し、その状態と複合カーソルの制御によって操作を実現する。以上の流れを踏まえ、本システムの構成として、深度センサーにより骨格情報を認識し取得する骨格情報取得部、得た骨格情報を元に複合カーソルへと変換する複合カーソル制御部、マウスにおける左右クリックの状態、つまりオブジェクトや視点などのモードを変更するためのトリガーとなる情報を取得する状態取得部、及びアプリケーション実行部に分かれる。システム構成図を以下の図3に示す。

4. 実装

4.1 骨格情報取得部

骨格情報取得部分では、深度センサーにより、センサーの視界内の物体との距離情報を取得する。人間がその視界に入った場合、ミドルウェアである kinect sdk beta2^{*5}によって骨格情報を追跡することができる。また、同位置に配置された通常のRGBカメラとの連動により、三次元位置情報を二次元画像へ正射影変換することによって、各部位の二次元画像に対する位置の割合を取得することができる。

4.2 複合カーソル制御部

普段マウスを持つ手を利き手とし、骨格情報取得部にてその部位のセンサーからの三次元位置を取得しておく。二次元カー

*5 “KinectSdkBeta2”, <http://www.microsoft.com/en-us/kinectforwindows/>

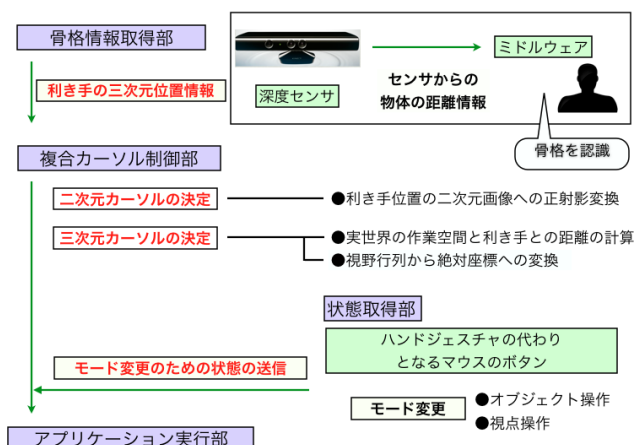


図 3: システム構成

ソルの位置は、利き手の三次元位置情報を先述した正射影変換を用いて二次元的な位置の割合を取得し、画面サイズと掛け合わせることで決定される。

また三次元カーソルは、二次元カーソルの位置と、利き手のセンサーからの距離情報によって決定される。仮想空間内のカーソルが利き手によって動かしているように見え、直感的な操作が可能となるよう、作業現場となる実空間の位置をセンサーから約 1.3m のところに設定し、その位置と利き手との距離によって三次元カーソルを配置する。

4.3 状態取得部

オブジェクトや視点を操作する際、それぞれのモードを切り替えるための情報が必要となる。本システムでは、三次元カーソルがオブジェクトに触れているときにクリック動作でオブジェクトの操作、触れていないときに同動作で視点の操作としている。ハンドジェスチャの代わりとなるボタンによるイベント判別は、オブジェクト操作の場合、左クリックを移動、右クリックを向き合わせとし、視点操作の場合、左クリックを垂直・平行移動及び奥行きに対する平行移動、右クリックを視点の回転移動としている。

4.4 アプリケーション実行部

アプリケーション実行部では、複合カーソル制御部によるカーソル制御と状態取得部によるモード変更処理を反映させる部分となる。

5. 評価実験

あるオブジェクトを指定の位置へ配置するタスクと、オブジェクトを指定の向きへ合わせるタスクの二つを行う。ホイール付きマウスと骨格認識によるインターフェースの両方に対して行い、配置に要する距離、および角度調整の難易度を 3 段階に分ける。角度調整の難易度の向き設定は易しいものから順に、目標を示す角度が、真下、真横、斜め手前 45 度 (左右はランダムである) の 3 段階である。効率性を表すものとして各タスク全体における時間だけでなく、オブジェクトの位置や向きが目標へ到達した時間や調整にかかった時間、エラーの回数等を測定する。

5.1 タスク内容について

配置のタスクでは、被験者は緑色で球体のオブジェクトを紫色で立方体の目標エリアへ運び、完全に入るようにする。開始

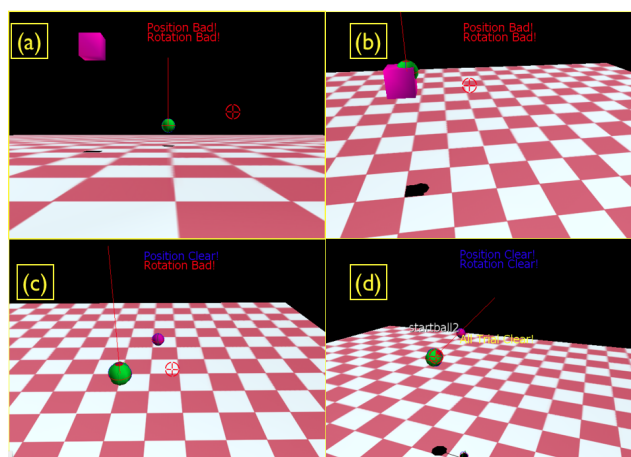


図 4: トライアル: (a) 目標エリアが見えるまでカメラを引く。(b) オブジェクトを目標のエリアへ運ぶ。(c) 位置合わせが終わり、目標の向きを示す紫色の球体が現れる。(d) 一連のタスクが終了する

直後はカメラがオブジェクトのすぐ手前にあり、いくらかカメラを引く動作をしないと目標エリアが見えないようになっており、左右どちらかにあるかはランダムとなっている (図 4(a))。オブジェクトの直径と紫色の一辺は、後者の方が前者より 0.2 ピクセルほど長く設定されている。ここで、オブジェクトの位置が 1.0 秒間立方体の範囲内にあり続けた時、オブジェクトの位置が安定していると見なす (図 4(b))。

向き合わせのタスクでは、配置のタスクが完了したと同時に紫色の立方体は消え、指定する方向を示す紫色の球体が出現する (図 4(c))。オブジェクトの方向を示す赤色の光線がその球体と交差し、オブジェクトと同様 1.0 秒間安定したとき、二つのタスクが終了したと見なされる (図 4(d))。

5.2 実験について

効率性を表す評価指標として、オブジェクトが最初に目標を通過するまでの時間、位置合わせ全体にかかった時間、向きが最初に合った時間、向き合わせ全体にかかった時間を測定する。さらに、位置や向きを安定させる際に目標を越えた回数をそれぞれ測定している。実験対象は 20 代前半の女性 1 人を含む 9 人であり、皆ほぼ毎日コンピュータを使用しており、かつ色盲ではない。実験データは、2 種類のインターフェース、3 つの距離、3 つの角度調整の難易度、9 人の被験者、3 回の反復、合計で 486 回分のデータを測定した。

また距離と角度調整の難易度の組み合わせが変わるごとに、1 から 5 ポイントのリッカート尺度による質問を行う。オブジェクトや視点それぞれの操作における操作感、各タスク全体を通した操作感、また疲労感をつけてもらい、また質問の度にコメントを残すことを奨励する。

実験の際マウスと骨格認識の両方において、被験者と画面との距離は約 3[m] に統一され、深度センサーを使う場合はセンサーから約 1.3[m] 付近を作業空間としている。ホイール付きマウスは Logicool 製で、深度センサーとして kinect を使用し、本システム及びトライアルの言語は C # で実装されている。また kinect に接続する PC の性能や開発環境は以下の表 1 のようになっている。

表 1: 実行環境

OS	Windows 7 Home Premium
CPU	Intel Core i7-2630 CPU @ 2.00GHz
メモリ	8.00GB
開発環境	Visual C # 2010 Express Edition

5.3 結果と考察

効率性に関する評価の結果を図 6 に示す。配置のタスクでは、距離が最小のタスクでは骨格認識を利用した方が効率的であり、特に目標までの移動時間平均は 23 %ほど軽減している(マウスが 12,492[ms] に対して骨格認識は 9,673[ms])。しかし距離が長くなるにつれ、目標への移動も調整にかかる時間も増えている。これは、距離の長さが増すほど多くの視点操作を必要とし、そしてその視点を調整する際、提案する操作手法では視点変更のためのアイコンまで二次元カーソルを移動する必要があり、視点変更ボタンを押してから実際に視点が動き始めるまでのタイムロスが発生してしまうからであると考えられる。また、骨格認識の操作では距離が大きいかほど超過の回数が増えている。これは距離が長くなりカメラからオブジェクトや目標の位置が遠くなると、カーソルのブレが激しくなることが原因として挙げられる。このブレは、手の位置が細かく安定しづらいというだけでなく、骨格を認識する際の誤差の影響もあるかも知れない。

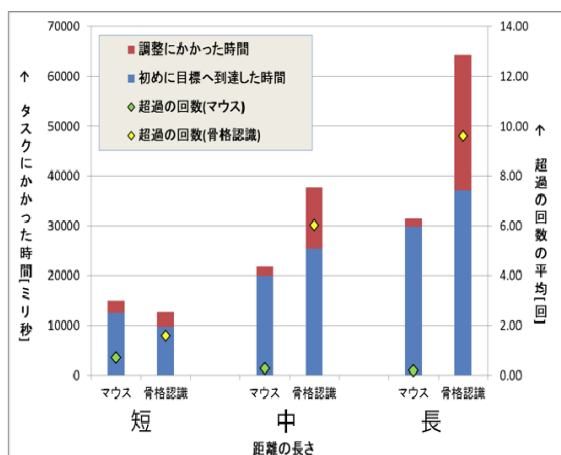


図 5: タスクにかかった時間：配置のタスク

また向き合わせのタスクにおいても、全体的な効率性はやや骨格認識の方が劣ってしまう。しかし角度が複雑になるにつれ、マウスではおおよその向きに向かせるまでの時間が大きくなるのに対し、骨格認識ではマウスほど複雑さの影響を受けていない。これは、向きを空間的に直接指定できることによるものだと考えられる。

また使用感に関しては、オブジェクトの向き操作に関しては一人あたり平均 0.59 ポイント程、骨格認識の方が容易であるという結果となった。しかし、オブジェクトの移動操作や視点操作全般に関してはマウスの方が使用しやすいと回答され、また疲労感に関しては、骨格認識の方が平均 0.80 ポイント程多く疲労を感じるという結果となった。さらにコメントより、腕を上げ続けることによる疲労を訴えるものもあった。しかし視点操作を除くオブジェクトの移動操作や向き合わせに関しては、空間における操作の方が楽だという意見や、直感的であるため慣れやすいといった意見もあった。

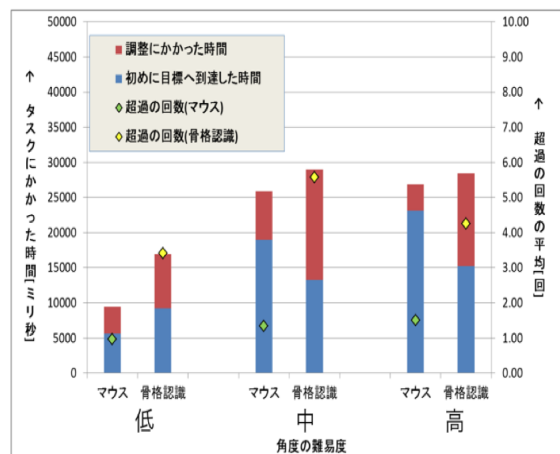


図 6: タスクにかかった時間：向き合わせのタスク

6. おわりに

本研究では、骨格認識を用いた仮想三次元空間でのオブジェクト及び視点における操作手法を設計・実装し、効率性や使用感に関してマウスとの比較実験を行った。複合カーソルを用いた空間での操作は、オブジェクト配置のタスクでは移動に関して有用性がみられたものの、視点操作に関してまだまだ課題点が多く、必ずしも効率的であるとは言えない。しかし向き合わせに関しては、空間では特に複雑な向きの指定において有効性を示唆することができた。今後は課題点である空間での視点操作の改善や、手の位置の認識精度の向上によって、骨格認識における操作性の向上を行う。

参考文献

- [1] Bernd Froehlich, et al. "The globefish and the globe-mouse: two new six degree of freedom input devices for graphics applications". In: ACM Conference on Human Factors in Computing Systems (CHI), pp. 191-199. ACM Press, 2006.
- [2] 木村 朝子, 鶴田 剛史, 酒井 理生, 鬼柳 牧子, 柴田 史久, 田村 秀行, "広視野電子作業空間に関する考察とシステム試作 マイノリティ・レポート型 I/F とその発展形", インタラクシオン 2005 論文集. pp. 143-150, 2005.
- [3] Mathieu Nancel, et al. "Mid-air Pan-and-Zoom on Wall-sized Displays". In CHI '11, 177-186. ACM, 2011.
- [4] Jamie Shotton, et al. "Real-Time Human Pose Recognition in Parts from Single Depth Images", Microsoft Research Cambridge And Xbox Incubation.
- [5] Dan Venolia. "Facile 3d direct manipulation", CHI. pp. 31-36. ACM Press, 1993.
- [6] 山田 裕之, 丹羽 佑輔, 白松 俊, 大園 忠親, 新谷 虎松. "ジェスチャーとキーイベントを利用したプレゼンテーション支援システムの実現". 日本ソフトウェア科学会第 28 回大会 講演論文集, 2011.
- [7] Jamie Zigelbaum, et al. "g-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface". In Proc. TEI '10, 261-264. ACM, 2010.