

## ジェスチャ入力に適した画像入力装置の提案と その3次元情報検出性能の検討

沼崎 俊一<sup>†</sup> 森下 明<sup>†</sup>  
梅木 直子<sup>†</sup> 土井 美和子<sup>†</sup>

人間とコンピュータとの円滑なインタフェースを実現するため、ジェスチャを認識するシステムが提案されている。しかし、低コストで、外部環境条件にロバストで実用的なシステムは少ない。我々は、ジェスチャを入力するための新しい入力デバイス、モーションプロセッサ<sup>TM</sup>を提案する。このデバイスは、近赤外光を発光させその反射光を取得する原理により、背景から対象物体の画像をリアルタイムに切り出すことができる。複雑な背景下でも誤認識することなく動作し、物体切り出しのための計算機パワーを必要としないので、低コストで構成できる。取得した反射光画像は距離的な情報を含んでおり、適切な処理により、3次元的なジェスチャの入力が可能である。モーションプロセッサが取得した反射光画像を、距離情報に精度良く変換するため変換モデルを明らかにする。実験により変換モデルのパラメータを取得し、そのパラメータにより、直線運動、平面内運動を検出した際の検出誤差を大幅に改善することを実証した。

### A Image Input Device for Gesture Interface and It's Performance of Detecting 3-Dimensional Information

SHUNICHI NUMAZAKI,<sup>†</sup> AKIRA MORISHITA,<sup>†</sup> NAOKO UMEKI<sup>†</sup>  
and MIWAKO DOI<sup>†</sup>

The intuitive interface like gestures are very important for human computer interaction. The conventional gesture recognition technologies have limits in practical uses because of their expensive cost and severe conditions. It is too difficult to use the conventional technologies for home PC applications. We developed a new input device "Motion Processor<sup>TM</sup>". The Motion Processor emits infrared light to the object and receives the reflected light. Therefore the Motion Processor can reject background and capture only the object's shape, motion and distance. It works well in complicated background and is low cost because it does not need a large amount of computer power. The detected image is not accurate range image, but includes distance information related to the object. This paper describes the conversion model to obtain the 3-dimensional information from the detected image. The results of the three experiments offer the parameters of the conversion model. The evaluation experiments show the proposed conversion model can obtain the accurate 3-dimensional image.

#### 1. はじめに

人間同士のコミュニケーションにおいては、テキスト、音声などのバーバル(言語的)チャネルだけでなく、表情や身振り/手振りなどのノンバーバル(非言語的)チャネルが重要な役割を果たしている<sup>1),2)</sup>。人間とコンピュータとのコミュニケーションにおいても、近年ノンバーバルチャネルの重要性が指摘されており、なかでも、身振りや手振りといったジェスチャ情報は重要であり、多くの研究が行われている。

従来のジェスチャ入力システムは、システム規模が大きかったり、処理速度が遅かったり、使用条件(環境)に制約があり、汎用にしかも安価に使用できるようなものが少ない。我々は、一般ユーザが使える程度のコストで構成可能であり、高速処理が可能なジェスチャ入力デバイスを開発した。これは、背景を除去した物体のみの形状を高速に取得できることが特徴である。本論文の前半ではこの新しい画像入力技術を用いたモーションプロセッサを提案する。モーションプロセッサは、その原理上、取得画像には3次元的な情報が含まれている。本論文の後半では、この取得画像から3次元情報を抽出するための変換モデルを提案し、そのモデルの有効性を実験的に検証した。以下、2章

<sup>†</sup> 東芝研究開発センター ヒューマンインタフェースラボラトリー  
Human Interface Laboratory, Corporate Research &  
Development Center, TOSHIBA Corp.

で従来技術を概説し, 3, 4 章でモーションプロセッサの原理・仕様について述べる. 5 章で取得画像と距離画像との差異を明確にし, 6 章で精度の良い 3 次元位置入力を行うための変換モデルのパラメータを求め, 効果を明らかにする.

## 2. 従来技術

ジェスチャ入力を目的としたシステムは数多く提案されているが, 個人レベルで使用できるような入力デバイスは少ない. グローブ型の装着型デバイスを用いたり<sup>3)</sup>, 色マーカーや発光体を手や体に取り付けて, これをカメラで追跡したり<sup>4),5)</sup> することで, ジェスチャを入力する方法は, 環境に依存せず安定的に検出できるメリットがある反面, ユーザにとっては煩わしい. 何も身につけないユーザをカメラで撮像し, そこから動きの情報を認識するシステムは, ユーザにとっては制約が少なく扱いやすい. カメラを 2 台用いて, 手の 3 次元的位置や動きを検出する手法<sup>6),7)</sup> も提案されているが, システムの規模が大きくなってしまふ. それに対し, 1 台のカメラで得られた画像に対し, 手の形状モデルをあてはめることにより, 位置や向きの推定を行う手法<sup>8),9)</sup> もあるが, 手形状が背景から容易に切り出せる環境が前提であり, 複雑な背景の下では安定して動作しない. 画像のフレーム間での, 局所領域の移動を手がかりに, 動きを検出するオプティカル・フローという技術を用いると, 複雑な背景下でも物体の動きを認識することができる<sup>10)</sup>. しかし, このような複雑な処理は, 最近の高性能なコンピュータを用いると実時間処理が可能とはいえ, かなりの処理コストがかかる. また, 大まかな動きは検出できるが, 物体の形そのものは捉えることができず, また 3 次元的位置を取得することもできない.

一方で, 距離画像を取得する装置についても多くの研究が行われている<sup>11)</sup>. スリット光やスポット光で物体上を走査し, 物体上の光をカメラで捉え, 三角測量の原理により距離を計算する方法<sup>12)~14)</sup> は, 非常に精度の良い距離画像を得ることができるが, リアルタイムに処理をするのが難しく, システムが高価になってしまう. カメラを 2 台用いて, それぞれの画像中の対応点を求め, 距離を計算する手法もあるが, 対応点を正確に求めるのが難しく, リアルタイム処理も難しい. また, カメラを用いて精度良く距離画像を得るための, カメラ・キャリブレーションについて詳細に報告されている<sup>15)</sup>. 本論文で要求される歪み補正は, 高速処理が可能な簡便なモデルであることが重要であり, また反射光量という異なる要素の補正も行うことが必

要になる.

## 3. モーションプロセッサの概要

我々は, ジェスチャを入力するシステムには以下の性質が重要と考えた. まず, 低コストであること. 一般家庭の中でユーザが用いるためには, 高性能なコンピュータを使わずに実現できなくてはならない. また, 複雑な背景の下でも安定して動作しなくてはならない. そして, ストレスなくジェスチャを入力できるためには高速に動作する必要があり, 自由度の高い人の動きを検出するためには 3 次元の動きを追跡できることが重要である.

我々は, 新しい原理を用いて, 手などの対象物をリアルタイムで背景より切り出して取得することができる動作入力装置を開発し, これを用いたジェスチャ入力デバイスを開発した<sup>16)</sup>. 我々はこれをモーションプロセッサと呼んでいる. モーションプロセッサは, 従来の画像処理ベースの物体抽出と異なり, CPU に演算負荷をかけずにリアルタイムで物体画像を切り出すことができる. 本章ではモーションプロセッサの原理と構成について述べる.

図 1 にモーションプロセッサの動作原理を示す. 図の LED は近赤外光を物体 ( 図中では手 ) に対し発光する. 発光された近赤外光は物体で反射され, 光学系 ( レンズ ) を経てイメージセンサ上に結像する. 反射光の強さは, 一般的に物体の距離が遠くなるに従い, 急激に減少する. したがって, このイメージセンサには, 背景は映らず, 物体のみが映る ( この画像を, 以下反射光画像と呼ぶ ). 近赤外光は不可視であるため, 操作者が使用中にまぶしさを感じることはない.

従来の CCD カメラ画像には, 背景が映っているため, これを除去し, 物体のみを抽出しなくては, その形状や動きを検出することはできなかった. そのため, 複雑な背景下で物体を認識するためには, 高度な画像処理技術を用いて, 背景を取り除かなくてはならなかった. モーションプロセッサは, 上述したように,

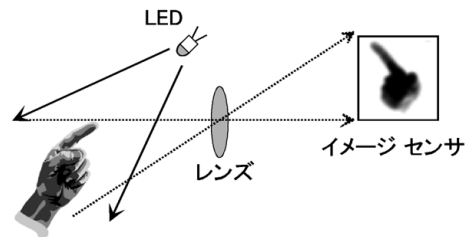


図 1 モーションプロセッサの原理

Fig. 1 The principle of Motion Processor.



図 2 モーションプロセッサの取得画像の例  
Fig. 2 Capture image of Motion Processor.

イメージセンサに結像する画像は、すでに背景が除去された物体画像であるため、計算量を大幅に減らすことができる。そのため、CPU に負荷をかけることなく高速にリアルタイムで取得できる。

モーションプロセッサによって検出された手の画像の例を図 2 に示す。背景は除去され、手形状のみが撮像されているのが分かる。

モーションプロセッサの機能は、物体形状を背景から切り出すだけではない。同じ物体からの反射光の強さは、物体までの距離の 2 乗にほぼ反比例する。したがって、モーションプロセッサの画像は、3 次元的な情報を含んでいる。

#### 4. モーションプロセッサの仕様・構成

モーションプロセッサは、カメラ部と PCI インタフェースカードから構成される。図 3 はカメラ部の外観写真である。中央のレンズの周囲に、近赤外光を発光する LED を配している。またレンズの奥に受光部のイメージセンサチップがある。インタフェースカードは、イメージセンサ制御用ロジック、イメージバッファ、PCI インタフェース機能などを搭載している。

表 1 は、モーションプロセッサの仕様を表している。画像サイズは  $64 \times 64$  画素である。モーションプロセッサは動きを認識するためのものであるので高い解像度は不要であり、また PC での処理の負荷を考慮すると、この程度が適切である。各画素 8 bit のデータで反射光量が出力される。背景からの物体の切り出しに CPU の負荷を必要としないため、高速処理が可能であり、ビデオレートを上回る、毎秒 50 枚の画像を得ることができる。このことは、動きの速いジェスチャも容易に検出できることを示している。データ量は毎秒 200 KByte であり、インタフェースとしては高速にデータ転送が可能な PCI インタフェースを採用した。反射光を用いているため、撮像距離範囲は 30 ~ 90 cm 程度である。ただし原理的には、強力な光源を使うことで距離を延ばすことは可能である。



図 3 モーションプロセッサの外観  
Fig. 3 Picture of Motion Processor.

表 1 モーションプロセッサの性能  
Table 1 Specification of Motion Processor.

分解能	$64 \times 64$ pixels
奥行き深度	256 階調
動作レート	毎秒 25 ~ 50 frames/sec
撮像距離	30 ~ 90 cm
画角	対角 約 $70^\circ$ (奥行 40 cm 地点で $40 \times 40 \text{ cm}^2$ )
本体サイズ	W75 × H75 × D78 [mm]
対応システム	PC/AT 互換機 (PCI バスを具備しているもの)
対応 OS	Microsoft Windows 95/98
接続方法	専用ボードで PCI バス接続

パソコン側には、対象物体が切り出された画像がリアルタイムで送られてくる。実際に、ジェスチャ入力システムを構築するためには、これらの画像に対し、様々な演算を行い、手の位置や動き、ジェスチャの種類などを認識するソフトウェアを開発する必要がある。そのための開発キットとして、モーションプロセッサ SDK (Software Development Kit) を開発した。SDK は 3 次元空間における対象物の重心位置、重心速度、体積の計算、FFT 演算などの基本的な画像演算から、パターンマッチング、オプティカル・フロー抽出などのより高度な処理を行う API (Application Programming Interface) 群から構成されている。

この SDK により、ソフトウェア開発者はアプリケーションを容易に開発できる。我々はこの SDK を用いていくつかのジェスチャ入力を行うアプリケーションを開発した<sup>17),18)</sup>。

#### 5. 反射光画像と距離画像の相違

上述のようにモーションプロセッサの取得画像は、反射光画像であり、物体までの距離と相関を持ち、距離情報を含んでいる。しかし、この画像は正確な「距離画像」ではなく、画素値が物体までの距離を示してはいない。その画像には様々な非線形要因を含み、距離画像に単純に線形変換できない。

本章以降で、モーションプロセッサ画像から距離情

報を取得するときの問題とその解決方法について述べる．本章では、モーションプロセッサの取得画像（反射光画像）と距離画像の違いについて述べ、次章で、変換モデルを用い高精度な3次元情報を再現できることを示す．

本論文では、既知の球状物体を撮像し、そこから球状物体の3次元位置を求める、というタスクを想定し、ここに関連する問題について論ずる．平面物体を用いた場合は、カメラ方向に対する面の傾きを考慮する必要があるのに比べ、球状物体はつねにカメラから見た向き・形状が同一であり、問題が単純化できるためこれを用いた．物体の3次元形状などさらに高度な3次元情報を取得する際の問題点については考察で述べる．

議論を始める前に、距離画像の定義を行う．一般に距離画像とは、画像の画素値が、そこに映っている物体までの距離を表すような画像を指す．ここでは画素値が表す距離は、カメラの光軸（中心軸）と垂直に交わる平面が、その画素の表す場所を通るときの、カメラと平面の距離、と定義する．カメラの光軸に垂直な平面があるとき、これを表す距離画像は、画素値がすべて一定であり、その値は、カメラと平面との距離である．

以下、反射光画像の性質について、距離画像との違いを基に説明する．

### 5.1 反射光量と距離の関係

まず、距離画像は、その画素値が物体までの距離を表すのに対し、反射光画像の画素は、物体上の対応する点からの反射光量を表す．センサの取得する反射光量は、反射光量を  $I$ 、距離を  $L$  とすると、式(1)の関係になる．ここで  $k$  は定数である．

$$I = \frac{k}{L^2} \quad (1)$$

式(1)により、反射光量から距離を計算できる．式(1)の  $k$  は、物体表面の反射率や反射特性、また、物体面のカメラ方向に対する傾きなどの条件によって異なる．しかし物体表面が完全拡散反射面に近い特性を持っていると仮定し、面の傾きをそれほど考慮しなくてよい条件下であれば、比較的良好に距離情報を再現できる．人間の手はそれに近い特性を持っており、手の位置・動きを捉えるという典型的な使い方の下では、比較的正確に距離情報を捉えることができる．

### 5.2 周辺部の反射光量低下

距離画像の定義では、カメラの光軸と垂直に交わる平面上の任意の位置にある物体の距離値はすべて等しい．しかし、実際にこのような物体を撮像した反射光

画像上では画素値は一定とはならず、中心に比べ周辺部の値が小さくなる．これにはいくつかの要因がある．第1に、中心部と周辺部で距離が異なることである．つまり、光軸に垂直な平面上では、カメラの光軸との交点が最もカメラに近く、そこから離れるに従って、距離が大きくなるため反射光量は小さくなる．第2にレンズの特性があげられる．レンズには通常、周辺光量低下という特性が見られる．これは、中心方向からの光に比べ、周辺方向からの光の量は少なくなり、撮像した像の周辺部が中心に比べ暗くなるという現象である．第3に発光源の指向特性がある．通常LEDは中心方向が最も明るく、周辺方向は暗い．モーションプロセッサの発光源として使用しているLEDは、比較的指向性が広く、中心に比べ周辺が緩やかに暗くなるタイプのものである．それでも周辺部にいくに従って、光量は低下する．

以上述べた複数の要因により、反射光の強さは画像中心に比べて、周辺部は小さくなる．しかし、この反射光量の低下は、画面中心からの距離に依存するので、比較的モデル化しやすい．

### 5.3 位置の歪み

レンズには歪曲という特性がある．これは、矩形物体をレンズに通すと、結像面で形が歪み、各辺が膨らむ樽型、各辺がへこむ糸巻型、あるいはそれらが複合されたような形に変形する．これは、実空間上での直線的な動きが、結像面上では曲がってしまうことを表しており、補正が必要である．

### 5.4 座標空間の違い

ところで、距離画像の表す空間は、四角錐であり、実空間とは座標系が一致しない．カメラの画角が小さく、ある程度距離のある物体を撮像する場合は、距離画像を直接実空間と考えられる．ところがカメラの画角が大きいと、この四角錐は、かなり末広がりの形になる．たとえば、斜め方向から、カメラに向かってくる物体は、距離画像上ではその位置は変わらないが、実空間では斜め方向に運動している．モーションプロセッサは画角が大きいため、この影響を無視できない．

実際の空間での動きを取り出すには、距離画像から実空間への変換が必要である．これを図4を用いて説明する．ここでは簡単のため、真上から見た図、つまり  $xz$  平面に射影した形で考える．直線  $m_1, m_2$  の間がモーションプロセッサの視野である．今、物体が  $P$  の位置にあるとする．カメラは  $O$  の位置にあり、カメラの光軸は  $z$  軸である．このとき、カメラの光軸と物体の方向の成す角度を  $\theta$  とし、物体の距離値を  $d$  とする．物体が直線  $t$  上のどこにあっても、距離画像

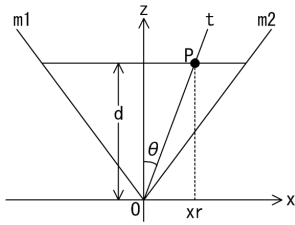


図4 距離画像空間と実空間の関係

Fig. 4 Relation between range image space and real space.

上での  $x$  座標  $x_k$  は等しいが、実空間上での  $x$  方向の位置  $x_r$  は  $d$  によって変化し、 $d \tan \theta$  となる。

実際の変換では、距離画像中の位置より、 $\tan \theta$  を求め、 $d$  と  $\tan \theta$  から実空間上での位置を求める。理想的には、この  $\tan \theta$  は、 $x_k$  に比例するが、前節で述べた位置歪みの影響により、厳密に比例しない。歪みを補正することは、 $\tan \theta$  を補正することと同じであるので、距離画像中の位置から  $\tan \theta$  への変換の際に、この補正を同時に行う。位置を補正した後、 $\tan \theta$  へ変換するより、手順が1つ少なく、実装上は都合が良い。

## 6. 変換モデル作成と評価実験

本章では、既知の位置にある対象物体を撮影した画像から、変換モデルを作成する。また、作成した変換モデルを用いることにより、物体の3次元位置を再現できることを評価実験によって確認する。

### 6.1 変換モデル

前章で、反射光画像と距離画像の違い、および実空間との座標系の違いについて述べた。これより、以下の3段階のステップで、変換を行うことにする。初めに、球状物体の位置および画素値(反射光量に対応する)が求まっているとする。

- (1) 周辺光量低下の補正を行う。
- (2) 反射光を距離に変換する。
- (3) 位置歪みを補正し、実空間への座標変換を行う。

まず、変換モデルのためのデータを取得するため、以下の3つの実験を行う。

データ取得実験1 実際の物体距離と、計測された反射光量との関係を得る。

データ取得実験2 カメラの光軸と垂直なある平面上を物体が移動するとき、反射光量がどのように変化するかを調べる。

データ取得実験3 反射光画像中の物体位置と、 $\theta$  の関係を得る。

次に、変換モデルを作成し、これを用いて実際に物体の位置を測定し、その有効性を確かめる。本実験で

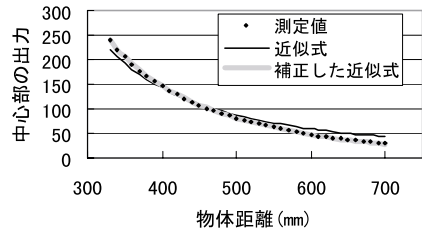


図5 物体距離と反射光量の関係

Fig. 5 Distance of object and amount of reflected light.

は対象物体として、直径8cmの球体を用いた。表面には黄色の塗装が施してあり、人の手と同程度の反射特性を持っている。鏡面性はない。

各位置における、計測された物体の反射光量は、物体の内部にある点の中で最も明るい5画素の平均値とした。最も明るい画素は、ほぼカメラから見て球面中央に位置するが、光軸と垂直な面での反射光量を計測するためにこの画素を利用することにした。最大値の画素のみを使用すると、ノイズによる誤差が大きくなるため、上位5画素の平均とした。反射光画像中の物体位置は、物体内部の最も明るい10画素の重心とした。位置についてはノイズの影響をさらに減らすために上位10画素を用いることとした。

### 6.2 データ取得実験1: 物体距離と反射光量

カメラの光軸上に球体を配置し、その距離を1cm刻みで変化させ、物体の反射光量を測定した。測定結果を図5に表す。

このグラフを見ると、かなり理論曲線に近い特性が得られているが、グラフの傾きがやや異なっていることが分かる。式(1)に定数項を加えて、式(2)を用いて近似した曲線を薄い曲線で表した。式(2)を用いるとかなり精度良く近似できることが分かる。

$$I = \frac{k}{L^2} + c \quad (2)$$

### 6.3 データ取得実験2: 上下左右の物体位置と反射光量

カメラから40cmにある平面上に対象物体を置き測定した。球体の中心位置は、カメラの正面を中心とし、上下左右に30mm刻みで動かし、計169点において計測した。

この結果を表したのが、図6である。2本の水平軸で、物体の位置を表し、そのときの出力を垂直軸で表している。

カメラに対向する平面上での物体の動きによる反射光量の変化は、画像中心を中心として同心円上に同じ大きさで現れる。すなわち反射光量は、画像中心から

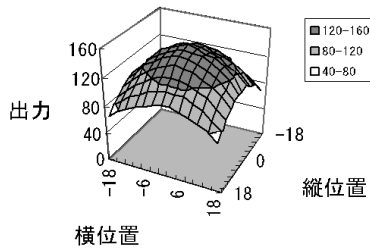


図 6 物体位置による反射光量の変化

Fig. 6 Variations of reflected light in position of object.

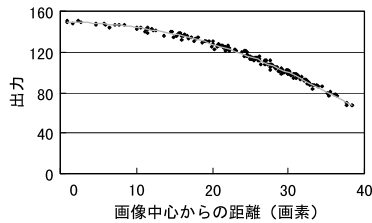


図 7 中心からの距離と反射光量低下の関係

Fig. 7 Relation between object position from image center and amount of reflected light.

の距離によって決まる。図 6 のデータを、画面中心からの距離とその距離における反射光量の関係にしたものが、図 7 である。

図中の曲線は、これらのデータにあてはめた式であり、2 次式で表すことができた。この曲線の落ち込んだ部分を、画面中心と同じ高さになるように持ち上げるような変換を行う。この 2 次曲線は垂直軸に関して対象であるので、画像中心からの距離が  $t$  であるような点に対しては、式 (3) で表せる係数を、画素値に乗ずることによって補正できる。

$$\frac{a}{t^2 + a} \quad (3)$$

#### 6.4 データ取得実験 3：測定された位置の誤差

データ取得実験 2 と同じ条件で測定する。ここでは、反射光量ではなく、測定された物体位置と実際の物体位置との関係に着目する。図 8 に、測定された物体位置をプロットする。理想的には、これらの点は格子状に並ぶが、レンズの歪みのため、端の方にある点が、中央に寄っているのが分かる。

反射光量の減少と同様に、位置に関する歪みも、画像中心を中心として同心円上に同じ大きさで現れる。反射光画像上における、物体位置の画像中心からの距離と、実物体とカメラを結ぶ線が光軸となす角 (図 4 の  $\theta$ ) の tangent の関係を図 9 に示す。また、近似曲線をあわせて表示した。この近似曲線は、位置歪みがなければ、直線になる。この曲線は、原点を通過する

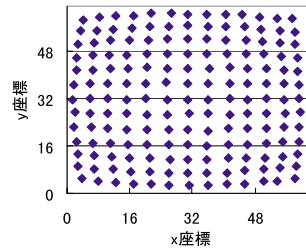


図 8 物体位置

Fig. 8 Positions of objects.

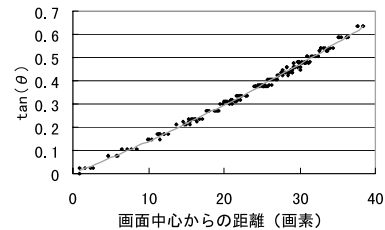


図 9 測定位置と実際の位置の中心からの距離

Fig. 9 Relation between detected positions and real positions.

2 次曲線で適切に近似され、画像中心からの距離を  $t$  とし、 $\tan \theta$  との関係は式 (4) で表すことができる。

$$\tan \theta = et^2 + f \quad (4)$$

#### 6.5 得られた変換モデルと精度

以上の 3 つのデータ取得実験により、任意の位置にある物体に対し、その 3 次元位置を特定するための変換モデルを得た。反射光画像より得られる、対象物体の画像中の重心位置  $x_0, y_0$  と、反射光量  $I_0$  から、実空間上の物体位置  $x, y, d$  を求めるには以下の処理を行う。画像中心を  $(x_c, y_c)$  とし、画像中心からの距離  $t$  を、式 (5) によって求める。

$$t = \sqrt{(x_0 - x_c)^2 + (y_0 - y_c)^2} \quad (5)$$

ここで得られた  $t$  を用いて、式 (6) で、反射光量の補正を行う。

$$I = \frac{a}{t^2 + a} I_0 \quad (6)$$

また、実空間上での物体方向のカメラ光軸との角度を、式 (4) で求め、物体までの距離  $d$  を、式 (7) で求める。この式は、式 (2) の変形である。

$$d = L = \sqrt{\frac{k}{I - c}} \quad (7)$$

物体のカメラ光軸からの距離は、 $d \tan \theta$  であるが、カメラ光軸を軸にしての回転移動はないので、式 (8)、(9) によって、物体位置が求まる。

$$x = d \tan \theta \frac{x_0 - x_c}{t} \tag{8}$$

$$y = d \tan \theta \frac{y_0 - y_c}{t} \tag{9}$$

ここで、分解能および精度について述べる。反射光画像のサイズは  $64 \times 64$  であり、各画素は 256 階調であるが、物体の位置、反射光量の初期値としては、領域の重心、平均を用いるため、分解能は十分にあると考えられる。測定結果をいくつかの近似曲線でモデル化した。その誤差は式 (7) で、反射光量を距離に変換するときで、 $\pm 5 \text{ mm}$  程度である。65 cm 以上では近似曲線の適合がやや悪く  $\pm 10 \text{ mm}$  であった。式 (6) で反射光量を補正するときの誤差は、40 cm の距離の場合で  $\pm 5$  であった。式 (4) で  $\tan \theta$  の補正を行うときの誤差は、 $\pm 0.02$  であったが、これは角度にして  $\pm 1.15$  度であり、40 cm の距離では  $\pm 8 \text{ mm}$  に相当する。

### 6.6 評価実験 1

この変換モデルを用いて、3 次元空間内での実際の動きを検出する実験を行った。変換モデル作成に用いたのと同じ球体物体を使い、撮像空間を斜めに直線的に横切るように動かし、一定間隔で撮像し、変換モデルを用いて変換した。結果を、図 10 に表す。求められた 3 次元位置の系列を、距離と  $x$  方向の位置、距離と  $y$  方向の位置に分けてグラフ化した。直線的に動いているので、グラフは直線的であるほど良い。また、各点の間の距離は等しいほど良い。

比較のために、反射光量の距離への変換のみを行い、反射光量の補正、位置の補正は行わない場合の結果を図 11 に示す。変換モデルを用いることにより、直線運動を正しく検出できることが分かる。

### 6.7 評価実験 2

さらに、球体を平面上に並べ、これをモーションプロセッサで撮像し、平面を再現できるか実験を行った。直径 2 cm の白色球体を 5 cm 間隔で  $5 \times 5$  の格子状に並べた。全体の大きさは、20 cm 角である。これをモーションプロセッサに対し、斜めに向け、撮像した。変換モデルを用いると、図 12 のように平面を再現できる。一方、反射光を距離値に変換しただけの場合は図 13 のようになる。格子の各交点が、求められた球体物体の位置であり、これらの図では、3 次元的に表示している。格子の下部に見える、薄い点は水平面に射影した影である。評価実験 1 と同様に、変換モデルの効果が現れていることが分かる。なお、評価実験 2 では使用する物体が評価実験 1 とは異なるため、新たに変換パラメータを取得し直した。

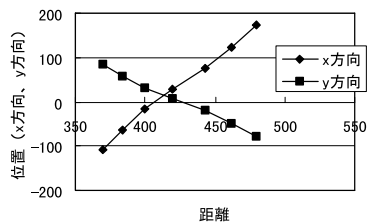


図 10 変換モデルを用いて直線運動を検出した結果  
Fig. 10 Detection of object moving in straight line with conversion model.

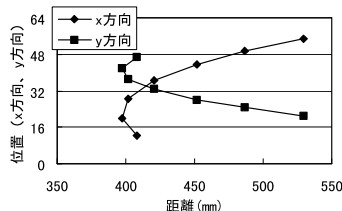


図 11 反射光量、位置の補正を行わない場合の、直線運動検出結果  
Fig. 11 Detection of object moving in straight line without conversion model.

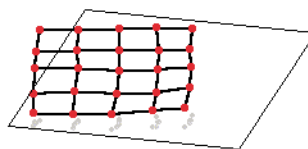


図 12 変換モデルを用いて平面状に並んだ球体を検出した結果  
Fig. 12 Detection of objects placed in grid with conversion model.

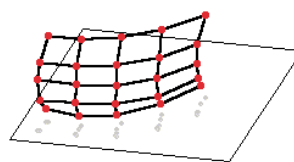


図 13 反射光量、位置の補正を行わない場合の、平面状に並んだ球体を検出した結果  
Fig. 13 Detection of objects placed in grid without conversion model.

## 7. 考 察

6 章の実験により、鏡面反射しない表面特性を持つ球形物体が対象物体である場合、その位置を変換モデルを用いて精度良く求めることができることが示された。変換モデルを用いない場合の結果を見ると、画像

周辺部の距離が、実際より遠くなっているのが分かる。反射光画像には、周辺ほど反射光量が小さくなる特性があると述べたが、これが大きく影響しているものと考えられる。

変換モデルの誤差は ± 数 mm 程度あり、高精度のレンジファインダにかなう性能ではない。しかし、非常に低コストなシステムであり、しかもリアルタイムに処理することが容易である。ジェスチャ入力装置の 3 次元情報取得性能としては十分である。

本モデルによって、3 次元位置を求められることを示したが、これはたとえば、既知の球体物体を黒い棒（これは近赤外光を反射しない）の先に取り付け、その棒を振ることで、3 次元的な位置・運動を精度良く入力できることを示している。人間の手や身体は球体物体によりも複雑な形状をしているので、細かい部分で見ると、本論文で述べたモデルで 3 次元位置が正しく求まるとは限らない。実際、物体表面のカメラ方向から見た傾きが影響するので、同じ距離にあっても、反射光量が異なる場合がある。しかし、モーションプロセッサの典型的な使い方は手の動きをとるものであり、手全体を本モデルにおける球体と位置づけると、本モデルは実際的な手法である。たとえば、握りこぶしの 3 次元位置を追跡するボクシングのようなジェスチャを仮定すると、本手法は有効に機能する。

## 8. ま と め

我々は、独自の原理により、背景から切り出された物体画像を高速に得ることができる、新しい画像入力技術を開発し、これを用いたジェスチャ入力デバイス、モーションプロセッサを開発した。

さらに、モーションプロセッサが取得する反射光画像は 3 次元的な情報を持つことに着目し、その画像の性質を分析し、距離画像との違いを明らかにした。これにより得られた変換モデルを用いることで、高い 3 次元位置検出性能を持つことを示した。

本論文では、物体表面がカメラ方向に対し斜めを向いている場合や、モデルを作成した物体と異なる物体を撮像した場合の影響などについては議論しなかった。これらの点に関しては今後、明らかにしていく。また、回転運動などのより高度な 3 次元情報を獲得するためのモデルについても、今後研究を行っていく予定である。

## 参 考 文 献

1) Mehrabian, A.: Communication without words, *Psychology Today*, Vol.2, pp.52-55

- (1968).
- 2) Birdwhistell, R.: *Kinetics and Context*, University of Pennsylvania Press (1970).
- 3) 西野浩明, 凍田和美, 宇津宮孝一: 両手ジェスチャで変形可能な 3 次元形状表現法, 情報処理学会論文誌, Vol.40, No.2, pp.698-701 (1999).
- 4) Cipolla, R., et al.: Qualitative visual interpretation of 3D hand gestures using motion parallax, *Proc. MVA '92*, pp.477-482 (1992).
- 5) 金森 務ほか: モーションキャプチャ「DigitEye 3D」の実装, 電子情報通信学会論文誌, Vol.J81-D-II, No.5, pp.804-809 (1998).
- 6) 内海章ほか: 多数カメラを用いた手形状認識とその仮想空間インタフェースへの応用, 情報処理学会論文誌, Vol.40, No.2, pp.585-593 (1999).
- 7) Segen, J., et al.: Fast and accurate 3D gesture recognition interface, *Proc. ICPR '98*, pp.86-91 (1998).
- 8) 岩井儀雄ほか: 単眼動画からの手の 3 次元運動と位置の推定, 電子情報通信学会論文誌, Vol.J80-D-II, No.1, pp.44-55 (1997).
- 9) Kuch, J., et al.: Virtual Gun: A Vision Based Human Computer Interface Using the Human Hand, *MVA '94 IAPR Workshop on Machine Vision Applications*, pp.196-199 (1994).
- 10) Cutler, R. and Turk, M.: View-based Interpretation of Real-time Optical Flow for Gesture Recognition, *Proc. IEEE Conference on Automatic Face and Gesture Recognition*, pp.416-421 (1998).
- 11) 井口征士ほか: 3 次元画像計測, 昭晃堂 (1990).
- 12) Lewis, R.A. et al.: A scanning laser range finder for a robotic vision, *Proc. 5th IJCAI*, pp.762-768 (1977).
- 13) Rioux, M.: Laser range finder based on synchronized scanners, *Applied Optics*, Vol.23, No.21, pp.3837-3844 (1984).
- 14) 荒木和男ほか: 高速 3 次元形状計測装置の試作, 信学論, Vol.J71-D, No.10, pp.2059-2068 (1988).
- 15) Weng, J., et al.: Camera calibration with distortion models and accuracy evaluation, *Trans. PAMI*, Vol.14, pp.965-980 (1992).
- 16) Numazaki, S., et al.: A Kinetic and 3D Image Input Device, *Proc. CHI '98 (SUMMARY)*, pp.237-238 (1998).
- 17) Umeki, N., et al.: A Motion Interface Approach Based on User's Tempo, *Proc. CHI '99 (Extended Abstract)*, pp.23-24 (1999).
- 18) 三原功雄ほか: ハンドアクションを用いた直感的な情報入力装置, 第 14 回ヒューマン・インタフェース・シンポジウム, pp.301-304 (1998).

(平成 11 年 10 月 29 日受付)

(平成 12 年 4 月 6 日採録)





沼崎 俊一 (正会員)

1992年東京工業大学総合理工学研究科修士課程修了。同年(株)東芝に入社。画像センサ, 画像解析, HI 入力デバイスに関する研究・開発に従事。電子情報通信学会会員。



梅木 直子

1991年千葉大学文学部行動科学科卒業。同年(株)東芝に入社。VR, CG, 動作入力 IF 技術に関する研究・開発に従事。



森下 明

1991年慶応義塾大学大学院計算機科学研究科修士課程修了。同年(株)東芝に入社。並列プログラミング, HI 入力デバイスに関する研究・開発に従事。



土井美和子 (正会員)

1979年東京大学大学院工学系研究科修士課程修了。同年(株)東芝に入社。文書処理, CG/VR, 入力デバイス, ウェアラブルコンピュータを中心とした HI 研究・開発に従事。ACM, 電子情報通信学会, HI 学会, 計測自動制御学会各会員。慶應義塾大学非常勤講師。

