

## 反射光画像を用いた3次元位置検出手法に関する検討

沼崎俊一、森下明、梅木直子、土井美和子

(株)東芝 研究開発センター

〒210-8582 神奈川県川崎市幸区小向東芝町1番地

TEL. 044-549-2242 FAX. 044-520-1308

numazaki@eel.rdc.toshiba.co.jp

我々は新しい画像入力デバイス、モーションプロセッサを開発した。これはコンピュータへの簡単で、直感的な入力方法をユーザに提供するものである。モーションプロセッサは近赤外光を発光し、その反射光を撮像することによって、背景から切り出された対象物の3次元形状をリアルタイムに取得できるため、ジェスチャ入力などを行うシステムを容易に構築できる。さらに、より精度の良い距離情報を得るためには、本デバイスの撮像する画像の持つ非線形性を解析し、これらを補正するのが望ましい。本論文では、この要因を明らかにするとともに、補正モデルを作成し、対象物体の3次元位置検出を精度良く行えることを示す。

## A Method of Detecting Object Position in 3-D Space by Using Reflected Light Image

Shunichi Numazaki, Akira Morishita, Naoko Umeki, Miwako Doi

Research and Development Center, Toshiba Corp.

1 KomukaiToshiba-Cho, Saiwai-Ku, Kawasaki-Shi, Kanagawa 210-8582, Japan

TEL. 044-549-2242 FAX. 044-520-1308

numazaki@eel.rdc.toshiba.co.jp

We developed a new input device "Motion Processor". The Motion Processor presents an easy and intuitive input method to users. It throws 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. But if we require more accurate range information, it is necessary to correct the image based on the analysis of non-linear characteristic of the image Motion Processor generates. In this article, we explain the factors of the difference between the image of Motion Processor and range image, then make a model for correcting range information.

## 1. はじめに

今日、コンピュータは広く普及し、専門家やオフィスでの使用にとどまらず、一般家庭で広く使用されるようになった。しかし、多くの人々にとって、コンピュータを使いこなせていないのが現状である。

日常生活の中で人間同士が身振り、手振りを使ってコミュニケーションを行うのと同じように、身振り、手振りなどを用いて、直感的に、自然に、コンピュータを操作できる機構が切望されている。

このような動機から、従来からジェスチャーなどの認識に関して様々な研究が行われてきた[1,2,4]。しかし、撮像した画像から手などの対象物のみを抽出することは大変困難で、多大な時間と計算量を必要とするという問題があった。リアルタイムに人間のジェスチャーなどを認識するには、ブルーバックを用いるなどの制約が多かった。

そこでこのような制約なく対象画像の切り出しを行えるデバイス「モーションプロセッサ」を開発した[3]。

モーションプロセッサは、光の反射の原理を用いて、背景画像を取り除くことで、対象物の画像のみを取り出す新規画像処理技術を採用した情報入力装置である。反射光量は、物体とカメラの距離によって変化するので、物体の3次元的な形状情報も取得できる。

モーションプロセッサにより、手の形状の切り出し、3次元空間内での手の動きの認識を行える。これをCG技術などを組み合わせ、直感的な操作を実現した[5]。

モーションプロセッサの取得画像は距離の2乗に反比例する反射光量の画像であるので、厳密な距離画像ではない。モーション

プロセッサが持つ非線形性を除く補正を行うため、本論文では、非線形要因と補正モデルを示す。さらに、実測したデータにより、補正モデルを作成し、その有効性を確認する。

以下、2節ではモーションプロセッサの概要を説明し、3節ではその画像に含まれる非線形要因について説明する。4節では実験により、補正モデルの作成と有効性の検証を行う。

## 2. モーションプロセッサの概要

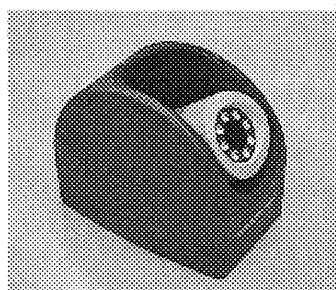


Fig. 1 モーションプロセッサ試作機

Fig.1 は我々が試作したモーションプロセッサである。中央のレンズの周囲に、近赤外光を発するLEDを配している。またレンズの奥に受光部のカスタムセンサチップがある。

Fig.2 にモーションプロセッサの動作原理を示す。LEDから発光された赤外光が物体

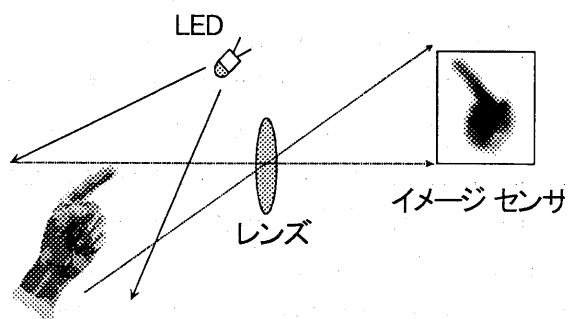


Fig. 2 動作原理



Fig. 3 検出された手の画像

で反射され、センサ部分がその反射光を受光する。

このようにして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 Fig.3 に示す。

図の様に、手の部分のみが検出されており、背景からの反射光は距離の2乗に反比例するので、微弱であり検出されていない。この原理により背景からの対象物の切り出しを低コストでリアルタイムに実現した。

試作したモーションプロセッサの性能を Tab.1 に示す。パソコンには、64x64 のデータアレイに 0 から 255 までの数値が格納されて送信される（ただし、下位 1bit は無効である）。

Tab.1 試作機の性能

分解能	64×64 pixels
距離深度	128 階調
撮像レンジ	30 ~ 90 cm
動作レート	30 ~ 50 frames / sec
本体サイズ	幅 90×高 120×奥行 140mm
対応 I/F	PCMCIA
対応 OS	Windows NT

パソコン側では、モーションプロセッサから送信された画像に対し、様々な演算を行い、手のジェスチャーなど対象物の 3 次

元的な形状や動きを認識することができる。

モーションプロセッサ SDK (Software Development Kit) と呼ぶ形状・動き認識のための様々な演算を行うためのライブラリを開発した。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 演算といった低レベルの演算を行う部分から、パターンマッチング、オプティカル・フロー抽出などといった高次の認識を行う API (Application Programming Interface) 群から構成されている。

Fig. 4 は、一例として SDK を用いて手の重心座標を求めているところである。

この SDK により、ソフトウェア開発者は、モーションプロセッサのアプリケーションを容易に開発できる。実際に、我々は SDK を用いてマルチメディア・アプリケーションを試作した。

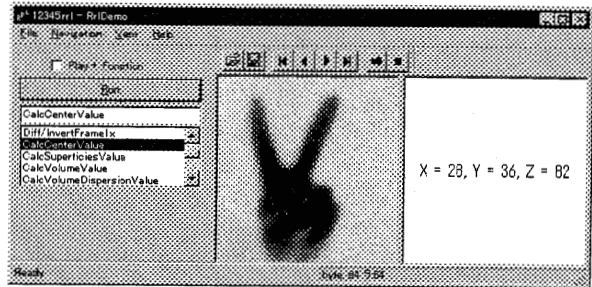


Fig. 4 手の重心位置の計算例

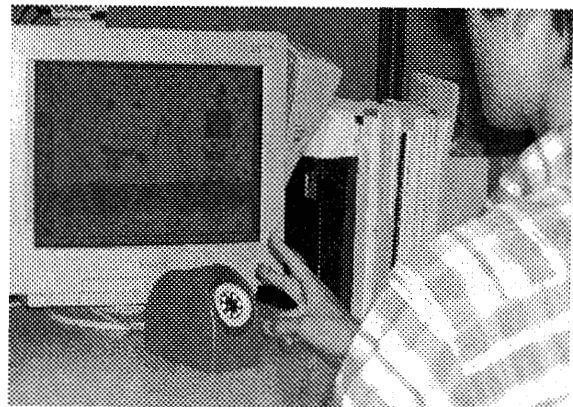


Fig. 5 じゃんけんゲーム

Fig.5は、作成したじゃんけんゲームをしている図である。切り出された手の形を認識し、勝敗を判断するとともに、手の奥行き方向の動きを検出して、人間のじゃんけんを出すタイミングを認識している。

### 3. 取得画像の非線形要因と補正モデル

#### 3.1 非線形要因

本節では、モーションプロセッサの取得画像の非線形要因を明らかにする。

まず、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わすような画像を指す。ここでは画素値が表わす距離は、カメラの光軸（中心軸）と垂直に交わる平面が、その画素の表わす場所を通るときの、カメラと平面の距離、と定義する。つまり、Fig.6において、平面  $m$  はカメラの光軸と直行する。カメラと平面  $m$  との距離を  $d$  とする。平面  $m$  上の点  $P$  に物体がある時、点  $P$  の位置に対応する画素値は  $d$  となる。

次に、モーションプロセッサ取得画像の非線形要因を説明する。第 1 の要因は、同

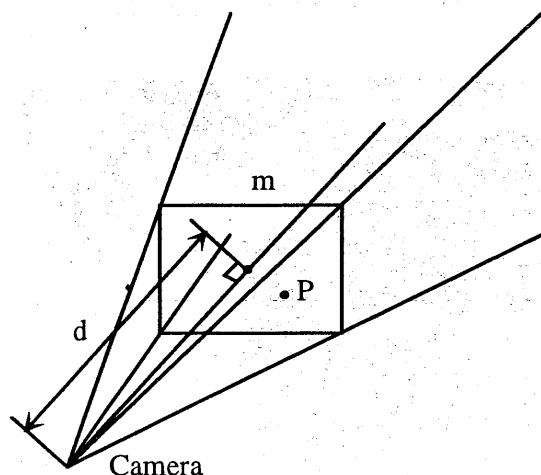


Fig. 6 距離画像の定義

一平面での反射光量の変化である。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、(1)式の関係になる。ここで  $k$  は定数である。

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

(1)式により、反射光量から距離を計算できるはずである。が、厳密には距離以外に、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反射光量は左右される。しかし物体が完全拡散反射面に近い性質の面に覆われていると仮定し、面の傾きを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取得できる。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下では、比較的正确に距離情報を捕えることができる。

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。が、実際にモーションプロセッサの取得画像上では同じにならない。つまり(1)式であらわされるように、同一の反射光量になるのは、カメラを中心とした球面になるためである。つまり、垂直面では、カメラの光軸の交点があってもカメラに近く、そこから離れるにしたがって、距離が大きくなり、反射光量は小さくなる。

第 2 の非線形要因は、レンズである。レンズには通常、周辺光量低下という現象が見られる。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 の要因は発光源の指向特性である。使用している LED は、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。

### 3. 2 補正モデル

以上述べた要因により、光軸の垂直面にある物体からの反射光量が等しくならない。次にこの非線形の補正モデルを考える。この反射光量の低下は、画面中心に対し、対称な形で起こるので、比較的モデル化しやすい。実際に物体を計測して、補正モデルを作成できる。この補正については、次章で述べる。

ただし、厳密には、対称物体の反射特性が変われば、補正モデルも変わる。しかし、主に撮影対象となる物体（たとえば手）に対し、補正モデルを作っておくことは現実的である。

次に、位置の補正モデルについて述べる。

ところで、距離画像の表わす空間は、四角錐である。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、この四角錐のうち、対象物体の存在する領域付近を切り取った形は直方体に近いので、そのまま実空間の距離情報と考えても違和感が無い。ところがカメラの画角が大きいと、この四角錐は、かなり末広がりな形になる。カメラの光軸上をまっすぐ近づいてくる物体は、距離画像上での位置が変わらないが、周辺にあり光軸と平行な動きで近づいてくる物体は、距離画像上ではどんどん周辺に動いていく。

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。Fig. 7 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。今、物体が  $P$  の位置にあるとき、距離画像上での  $x$  座標と距離値

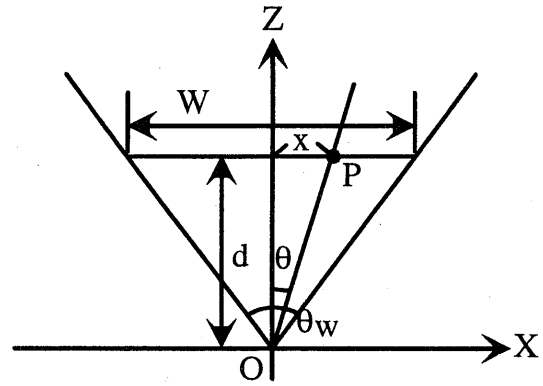


Fig. 7 距離画像空間と実空間との関係

$d$  は図中に書かれるように決まる。ここでカメラ  $O$  と物体を結ぶ方向が光軸方向と成す角を  $\theta$  とすると、 $\theta$  は反射画像中の物体位置の  $x$  座標で決まり、

$$\frac{x}{W} = \frac{\sin \theta}{\sin \theta_w} \quad (2)$$

という関係を満たす。ここで、 $W$  とは画像の  $x$  方向の画素数、 $\theta_w$  視野角である。したがって、実空間座標での位置は、 $x$  座標は  $d \sin \theta$ 、 $z$  座標は  $d$  となる。

ただし、式(2)には、レンズ歪曲の影響が考慮されていず、厳密にはこの式は成り立たない。が、この補正を行うは、 $x$  と  $\theta$  の関係を補正情報を含めモデル化すればよい。

## 4. 補正モデル作成と評価実験

本実験の目的は、既知の位置にある対象物体を撮影した画像から、補正モデルを作成し、その妥当性を検証することである。

まず 3 つの実験を行い、補正モデルのためのデータを取得する。ここで行う 3 つの実験は以下である。

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。
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。

### 3. 反射画像中の物体位置と、 $\theta$ の関係を得る

次に、補正モデルを作成し、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されており、人の手と同程度の反射特性を持っている。鏡面性はない。

球体を用いた理由は、面の傾きの影響を排除するためである。平面を撮像したときには同じ距離にあってもその面の傾きによって明るさが異なる問題がある。球体では、カメラ側から見た時の球体の中心近傍は常にカメラ方向に対し垂直となり、面の傾きの影響を排除できる。

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。計測された物体位置は、物体内部の最も明るい 10 画素の平均とした。

#### 実験 1：物体距離と反射光量

カメラの正面方向に球体を配置し、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 Fig.8 に表わす。

このグラフを見ると、かなり理論曲線に

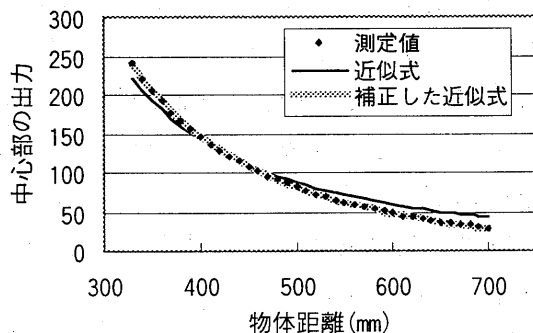


Fig. 8 距離と反射光量の関係

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。(1)式に定数項を加えて、(3)式を用いて近似した曲線を薄い曲線で表わした。(3)式を用いるとかなり精度良く近似できた。したがって、補正モデルとして(3)式を用いることにする。

$$I = \frac{k}{L^2} + c \quad (3)$$

#### 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。

この結果を表わしたのが、Fig.9 である。2本の水平軸で、物体の位置を表わし、その時の出力を垂直軸で表わしている。

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、基本的に対称的であり、画面中心からの距離のみに依存する。Fig.9 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、Fig.10 である。

図中の曲線は、これらのデータに当てはめた関係式である。これは 2 次式で表わすことができた。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

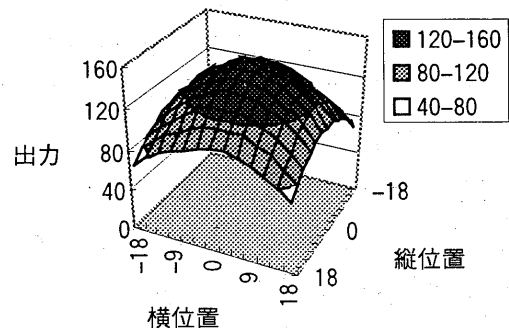


Fig. 9 物体位置による反射光量の変化

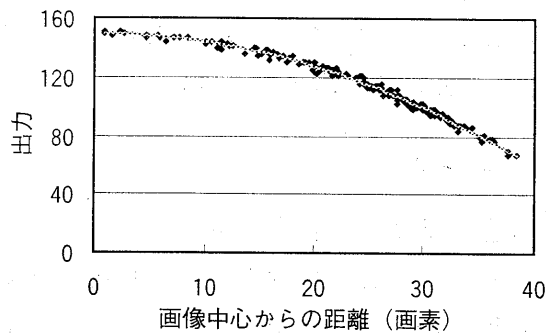


Fig. 10 中心からの距離と反射光量低下の関係

上げるような補正を行う。

### 実験 3：測定された位置の誤差

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。Fig. 11 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対称的である。画像上での物体位置と、実物体とカメラを結ぶ線が光軸と成す角の関係をモデル化する。実際にはこの角

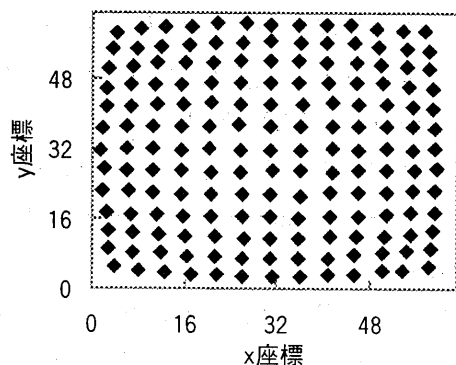


Fig. 11 測定された物体位置

度の  $\tan(\theta)$  が分かっているならば良い。この関係を Fig. 12 に示す。また、近似曲線を合わせて表示した。レンズ歪曲が無ければ、直線になるはずであるが、2 次の項を付け加え、2 次式でモデル化した。

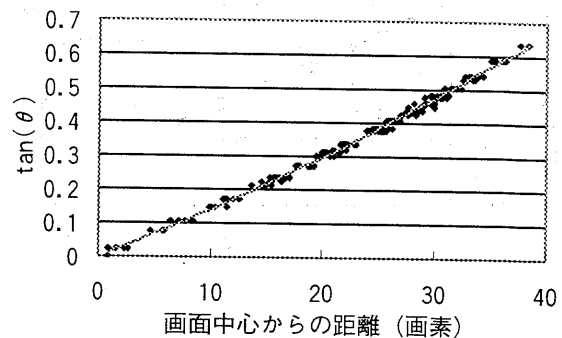


Fig. 12 測定位置と実際の位置の中心からの距離

以上の実験により、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計算式が得られた。計算式は大きく 3 つに分けられる。まず、反射光量を距離情報に変換する計算式、縦横方向の物体位置による反射光量低下を補正する計算式、そして、検出された縦横方向の物体位置を補正する計算式である。

この式を用いて、3 次元空間内での実際の動きを補正する実験を行った。補正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、補正モデルを用いて補正した。結果を、Fig. 13 に表わす。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ものである。直線的に動くので、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 Fig.14 に示す。補正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。

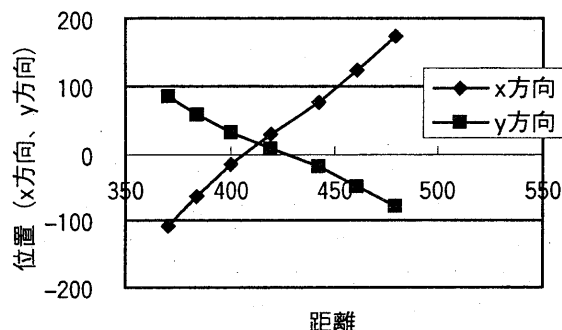


Fig. 13 補正モデルを用いて直線運動を検出した結果

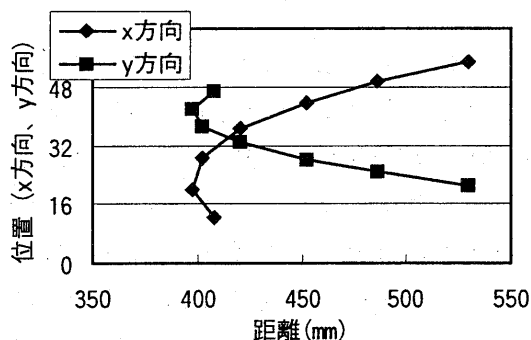


Fig. 14 反射光量、位置の補正を行わない場合の、直線運動検出結果

## 5. 考察

以上の実験により、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、その位置を補正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。これは限定条件のようにも思える。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、この点で実際的な手法である。例えば、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

チャを考えると、握りこぶしを球体の物体とみなすことで、4章で示した補正モデルで、3次元位置を求められる。

## 6. おわりに

本論文では、反射光を用いた画像入力装置において、より精度の良い距離画像を取得するための方法について述べた。本装置の持つ非線形性を明らかにし、これを補正するモデルを作成することにより、良好な3次元位置検出が行えることを示した。これにより、モーションプロセッサが手の3次元位置の精度を必要とするアプリケーションに対しても有用であることを示した。

今後は、物体の表面反射特性がモデル作成時と異なる場合や、物体表面の傾きがカメラに対し変化する場合について検討予定である。

## 参考文献

- [1] Cipolla, R., Okamoto, Y. and Kuno, Y. Qualitative visual interpretation of 3D hand gestures using motion parallax. *Proceedings of MVA '92*, pp.477-482, 1992.
- [2] 鹿毛裕史、三宅康也、久間和生、"人工網膜チップを用いたジェスチャ認識によるゲーム操作", 1997年信学会総合大会, pp.364, 1997
- [3] Numazaki, S., Morishita, A., Umeki, N. et al. A Kinetic and 3D Image Input Device. *CHI 98 SUMMARY*, pp.237-238, 1998
- [4] Nesi, P., Bimbo, A.D. A Vision-based 3-D Mouse. *Human Computer Studies*, Vol.44, pp.73-91, 1996
- [5] 梅木直子、森下明、沼崎俊一、土井美和子、"3次元画像入力装置とその利用について", 情処研報, ヒューマンインタフェース 79-8, pp.43-48, 1998