

## 装着型カメラを用いた頭部位置・姿勢の計測法

中崎 裕介<sup>†</sup> 和田 俊和<sup>†</sup>

ディスプレイを観測する人物の目の3次元位置が計測できれば、視点位置に依存した映像を提示することで、運動視差を作り出すことができ、3次元物体の表示を行うことができる。我々は、能動ステレオカメラを用いて眼の位置を計測する方法を提案したが、この方法では装置が高価になってしまうため、一般に普及することは困難である。そこで、本研究では頭部に装着したカメラでディスプレイの矩形を観測し、この矩形の変形から実時間でカメラの位置と向きを計測する方法を提案する。この方法では、首振り機構を持たない単眼カメラのみでディスプレイを観測している際のカメラの位置・姿勢が計測できるため安価なシステムが構築可能である。ディスプレイ枠の計測には位相計算型の勾配ベクトルを用いた Particle Filter を用いており、安定性と高速性の両者を同時に達成している。

### Position and orientation estimation using head-mount camera

Yusuke Nakazaki<sup>†</sup> and Toshikazu Wada<sup>†</sup>

By using 3D position information of human eyes watching computer display, we can realize viewpoint dependent image displaying, i.e., motion parallax generation, which enables us popup displaying of 3D contents. We already developed a 3D displaying system using active stereo camera measuring 3D eye position. However, this system requires expensive equipments, such as, stereo camera, pan-tilt unit, rotary encoders, and so on. For making this 3D displaying method popular, we have to have a cheaper device that measures 3D eye positions. In this research, we use a head-mount camera observing LCD display frame. That is, we can estimate the head-mount camera position and orientation from the deformation of the display frame on the image. This method does not require pan-tilt unit, because the computer operator watches the display from any viewpoints and the head-mount camera captures the display frame. When the operator does not watch the display, we don't have to know the viewpoint position and orientation. In our implementation, we employed phase based gradient estimation for fast and robust implementation of particle filter.

## 1. はじめに

人物の目の3次元位置に依存した3次元物体の表示を行うためには、眼の3次元位置を計測する必要がある。図1のような能動ステレオカメラを用いて眼の位置を計測する方法[1]を提案したが、装置が高価になってしまうため、安価な装置で実現する方法が求められている。

本論文では、頭部に装着した単眼カメラの位置と向きを計測する方法を提案する。カメラをPCで扱うとき、多くの場合において、ディスプレイが周辺に存在することに着目し、カメラでディスプレイの矩形を観測し、この矩形の変形から位置と姿勢を推定する。ディスプレイの形状が既知であれば、撮影するカメラの位置・姿勢を知ることが出来る。この方法では、首振り機構を持たない単眼カメラのみで、ディスプレイを観測している際のカメラの位置・姿勢を観測できるため安価なシステムが構築することが出来る。

カメラが撮影する環境が既知の場合に、カメラの位置・姿勢を推定する方法として、ホモグラフィ行列を用いる方法がある。これは、形状が既知の矩形が画面上に存在すれば、カメラ画像から矩形領域を検出し、矩形の4頂点とその矩形のアスペクト比からホモグラフィ行列を求め、ホモグラフィ行列を分解することにより、カメラの現在位置から目標位置までの並進ベクトルと回転行列を推定する手法がある。しかし、目標を撮影した動画で、1フレーム毎に処理を行うことは困難である。

提案手法では、カメラ画像上のディスプレイの矩形を、位相計算型の勾配ベクトルを利用した Particle Filter[2]を用いて計測し、撮影するカメラの位置・向きを推定する。

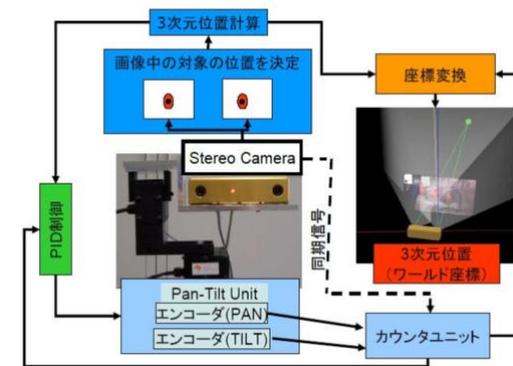


図1 能動ステレオ計測

<sup>†</sup> 和歌山大学大学院システム工学研究科システム工学専攻  
Department of system engineering, Wakayama University

## 2. 関連手法

本章では、前処理として用いる歪み補正について述べたのち、提案手法で用いる Particle Filter についての説明を行う。

### 2.1 歪み補正

本節ではカメラのレンズ歪みの補正について述べる。対象物の形状に関する情報を正しく抽出するためには、レンズ歪みによる画像中の歪みを取り除く必要がある場合がある。内部パラメータには以下のような項目がある。

- ・カメラの焦点距離 :  $f$
- ・画像中心 :  $c_x, c_y$
- ・画素サイズ :  $s_x, s_y$

これらのパラメータによって、世界座標系の点  $P$  は、図 1 のように画像平面上に投影される。

世界座標系における点  $P$  の座標を  $(x, y, z)$  とし、カメラ座標系における点  $P$  の座標を  $P_c(x_c, y_c, z_c)$  とする。理想的なピンホールカメラでは、点  $P$  を画像平面上に透視投影した点を  $(U_u, V_u)$  とすると、 $(U_u, V_u)$  は次式で表される。

$$U_u = f \frac{x_c}{z_c} + c_x, V_u = f \frac{y_c}{z_c} + c_y$$

しかし、レンズ収差によって投影点にずれが生じ、点  $P$  は  $p(U_d, V_d)$  に投影される。補正モデル[3]を用いることで、 $(U_u, V_u)$  は次式のように  $(U_d, V_d)$  で表される。

$$U_u = U_d + (U_d - c_x)(k_1 r^2 + k_2 r^4) + p_1(r^2 + 2(U_d - c_x)^2) + 2p_2(U_d - c_x)(V_d - c_y)$$

$$V_u = V_d + (V_d - c_y)(1 + k_1 r^2 + k_2 r^4) + p_1(r^2 + 2(V_d - c_y)^2) + 2p_2(U_d - c_x)(V_d - c_y)$$

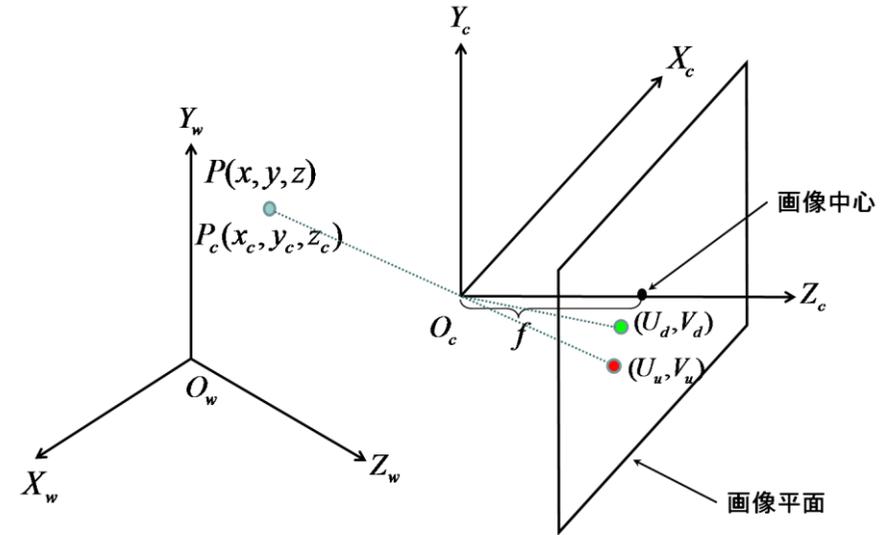


図 2 3次元空間の2次元画像平面への投影

ここで、 $r^2 = (U_d - c_x)^2 + (V_d - c_y)^2$  である。

また、投影された点をデジタル画像で扱うために、座標を離散化する必要がある。

$(U_d, V_d)$  がデジタル画像上で離散化された座標を  $(U_d', V_d')$  とすると、次式のようなになる。

$$U_d' = \frac{U_d}{s_x} + C_u$$

$$V_d' = \frac{V_d}{s_y} + C_v$$

ここで、 $C_u, C_v$  はカメラ座標系におけるデジタル画像の原点の座標である。

以上の式から、デジタル画像上の歪みを含まない点  $(U_u', V_u')$  は、次式で表される。

$$\begin{aligned} X &= U_d' \\ Y &= V_d' \end{aligned}$$

$$\begin{aligned} U_u' &= X + (X - c_x')(k_1 r'^2 + k_2 r'^4) + p_1(r'^2 + 2(X - c_x')^2) + 2p_2(X - c_x')(Y - c_y') \\ V_u' &= Y + (Y - c_y')(k_1 r'^2 + k_2 r'^4) + p_2(r'^2 + 2(Y - c_y')^2) + 2p_1(X - c_x')(Y - c_y') \end{aligned}$$

本研究では Zhang[4] のキャリブレーション法を用いることで、カメラの内部パラメータとレンズの歪み係数  $k_1, k_2, p_1, p_2$  を決定した。また、LUT を作成することで、高速な歪み補正を行う。

## 2.2 Particle Filter

本節ではディスプレイ枠の計測と追跡に用いる Particle Filter について述べる。Particle Filter とは、マルコフ連鎖モンテカルロ法 (MCMC) に基づく追跡アルゴリズムの一つであり、追跡対象の位置・姿勢、形状などの状態の確率分布を、ランダムサンプリングによって時刻毎に推定することで追跡を実現する手法である。

時刻  $t$  における対象の状態ベクトルを  $x_t$  と定義し、同時刻の観測ベクトルを  $z_t$  とし、時刻  $t$  までの観測値の集合を  $Z_t = \{z_1, \dots, z_t\}$  とする。このとき、 $x_t$  の事後分布  $p(x_t | Z_t)$  はベイズの定理から次式のようになる。

$$p(x_t | Z_t) = k t p(z_t | x_t) p(x_t | Z_{t-1}) \quad (1)$$

ただし、 $k t$  は正規化項である。また、 $p(x_t | Z_{t-1})$  は時刻  $t$  における事前分布、 $p(z_t | x_t)$  は状態ベクトル  $x_t$  の尤度を表している。ここで事前分布  $p(x_t | Z_{t-1})$  は、前の時刻  $t-1$

の事後分布  $p(x_{t-1} | Z_{t-1})$ 、および、あらかじめ与えた  $x_t$  の推移確率  $p(x_t | x_{t-1})$  を用いて次式で求められる。

$$p(x_t | Z_{t-1}) = \int_{x_{t-1}} p(x_t | x_{t-1}) p(x_{t-1} | Z_{t-1}) \quad (2)$$

また、 $p(z_t | x_t)$  はある状態  $x_t$  のときに、観測値  $z_t$  を得る確率(尤度)であり、式(1)、(2)より、各時刻における尤度  $p(z_t | x_t)$  を求めることで、事後分布  $p(x_t | Z_t)$  の推定ができる。Particle Filter では、事前分布  $p(x_t | Z_{t-1})$  に従ったランダムサンプリングによる仮説の生成と、その各仮説に対して尤度を求めることで、事後分布を離散的に近似し推定する。実際には、各仮説についての尤度を求め、その比に従った重み付けを行うことで、事後分布を近似的に推定している。各仮説を  $s_t^{(n)}$  とすると、その重み  $\pi_t^{(n)}$  は次式により求められる。

$$\pi_t^{(n)} = \frac{p(z_t | X_t = s_t^{(n)})}{\sum_{i=1}^N p(z_t | X_t = s_t^{(i)})}$$

Particle Filter では、事前分布  $p(x_t | Z_{t-1})$  に従った仮説  $s_t^{(n)}$  の生成、各仮説の重み  $\pi_t^{(n)}$  の計算を、[選択]、[予測]、[観測]の手順によって行う。アルゴリズムを以下に示す。

### (1) 選択

時刻  $t-1$  の仮説の集合  $s_{t-1} = \{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$  から各仮説の重み  $\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}$  の比に従い、仮説の選択を行う。ここで選択した仮説の集合を  $s'_{t-1}$  とする。 $t=0$  のときは、あらかじめ与えた初期分布  $p_0(x)$  に従い、 $s'_0$  を生成する。

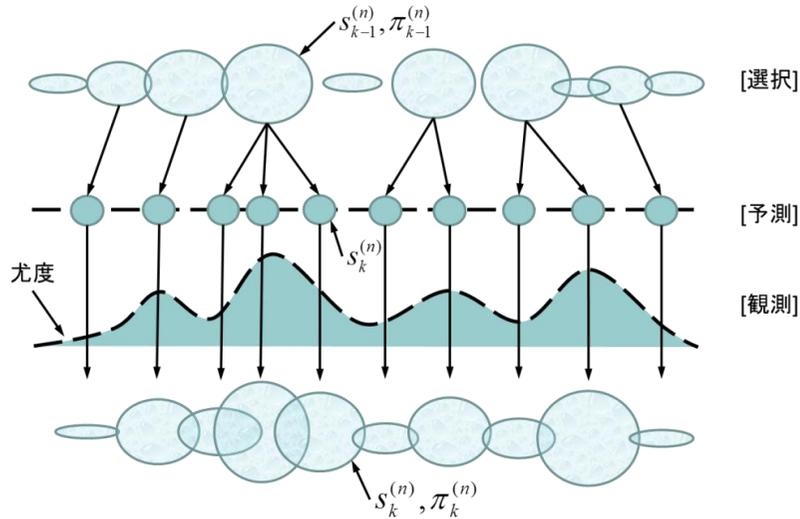


図 3 Condensation の処理手順

## (2) 予測

選択した仮説の集合  $s'_{i-1}$  の各仮説から推移確率  $p(x_t | x_{t-1})$  に従ったサンプリングを行う。

## (3) 観測

$s_t$  の各仮説  $s_t^{(n)}$  について尤度  $p(z_t | X_t = s_t^{(n)})$  を推定し、各仮説  $s_t^{(n)}$  の重み  $\pi_t^{(n)}$  を次式で計算する。

これらの「選択」、「予測」、「観測」の処理を繰り返すことで、各時刻の事後分布  $p(x_t | Z_t)$  が推定される。ある時刻  $t$  における状態ベクトル  $x_t$  の期待値は次式のようにになる。

$$E[x_t] = \sum_{n=1}^N \pi_t^{(n)} s_t^{(n)}$$

## 3. 提案手法

本章では、まず提案手法の基本アイデアを述べ、その後に提案手法の具体的な内容を説明する。

### 3.1 基本アイデア

本論文の基本アイデアは、頭部にカメラを装着した人物がディスプレイを見ているとき、ディスプレイの大きさが既知であれば、そのカメラで撮影された画像中のディスプレイの変化から、カメラの位置・姿勢を推定することができ、人物の頭部の位置と向きがわかることである。

### 3.2 提案手法の概要

本提案手法の概要について述べる。ワールド座標系の原点  $O_w$  をディスプレイの画面の左下の角とし、下端に  $X$  軸、左端に  $Y$  軸、画面の法線方向に  $Z$  軸をとる。また、 $X_w, Y_w, Z_w$  をワールド座標系の  $XYZ$  それぞれ方向ベクトルとする。

回転行列  $R$  と行列並進  $T$  を用いると、ワールド座標系におけるカメラの位置  $C$  は、

$$C = \begin{bmatrix} (R \cdot X_w)^T \\ (R \cdot Y_w)^T \\ (R \cdot Z_w)^T \end{bmatrix}^T \cdot (O_w + T)$$

となる。ここで、

$$X_w = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, Y_w = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, Z_w = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, O_w = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

であるため、 $C$  は以下のように簡略化出来る。

$$C = R \cdot T$$

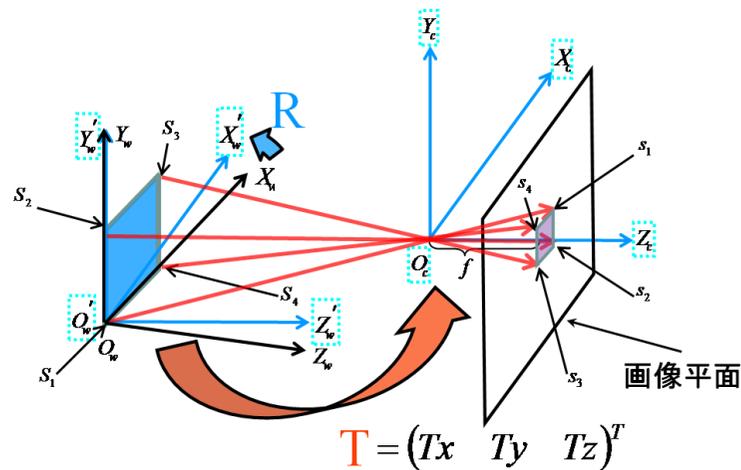


図 4 3次元空間平面の画像平面への透視投影

また、ディスプレイの画面の大きさとカメラの焦点距離が既知であれば、ワールド座標系でのディスプレイの画面の角の座標  $S_1, S_2, S_3, S_4$  が得られ、 $R$  と  $T$  を用いて、 $S_1, S_2, S_3, S_4$  をカメラ座標系で表した点を  $S'_1, S'_2, S'_3, S'_4$  とすると、それぞれ画像平面上の点  $s_1, s_2, s_3, s_4$  に投影される。

したがって、ディスプレイの画面の大きさとカメラの焦点距離が既知であれば、ディスプレイの枠を表す四角形  $S_1S_2S_3S_4$  は、回転行列  $R$  と行列並進  $T$  にのみ依存し

て、図 4 のように画像平面上の四角形  $s_1s_2s_3s_4$  に投影される。

これは、画像平面上のディスプレイの画面の枠を求めることが出来れば、ディスプレイの画面の左下の角からのカメラの位置  $T$  とカメラの姿勢  $R$  を推定できることを示している。

従って、本提案手法では Particle Filter を用いてディスプレイの画面の枠を求め、カメラの位置・姿勢を推定する。

-1	0	1	-1	-2	-1
-2	0	2	0	0	0
-1	0	1	1	2	1

図 5 Sobel オペレータ

### 3.3 Particle Filter によるディスプレイの枠の計測

本提案手法では、撮影した画像上のディスプレイの画面の枠を  $\phi, \rho, \theta, Tx, Ty, Tz$  の 6 パラメータで表し、Particle Filter を用いて追跡を行う。 $\phi, \rho, \theta, Tx, Ty, Tz$  はそれぞれ、ワールド座標上の  $X$  軸まわりの回転角度、 $Y$  軸まわりの回転角度、 $Z$  軸まわりの回転角度、 $X$  軸方向の移動量、 $Y$  軸方向の移動量、 $Z$  軸方向の移動量である。

$\phi, \rho, \theta, Tx, Ty, Tz$  の 6 パラメータを Particle Filter の状態ベクトルとし、6 パラメータによって画像平面上に描かれる四角形と画像平面上に投影されたディスプレイの画面の枠がどれだけ一致しているかを尤度とすると、状態ベクトルの期待値がカメラの位置・姿勢を推定した結果となり、状態ベクトルの期待値によって描かれる四角形によりディスプレイの画面の枠を求めることができる。

## 4. 提案手法の実装

本章では、前章で述べた提案手法の実装について述べる。提案手法を実装するにはディスプレイの画面の枠を求められるような尤度評価の方法が必要である。

撮影するディスプレイが光っており、周辺より明るいという仮定をすると、ディスプレイの画面の枠上の点での勾配方向を決定できる。4.1 節で提案手法に用いた位相計算型の勾配ベクトルの計算方法について述べ、4.2 節で勾配を用いた尤度評価の方法について述べる。

### 4.1 位相計算型の勾配ベクトル

コンピュータビジョンでは勾配を計算するとき、サイズが  $3 \times 3$  の Sobel オペレータ(図 5) を用いることが多い。Sobel オペレータを用いた計算では、基本的には注目画素と

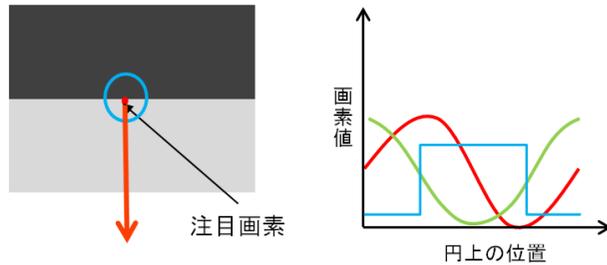


図 6 位相計算型の勾配ベクトル計算



図 7 実験に用いたカメラ

CPU	Intel(R) Xeon 3.06Ghz
メモリ	2G bytes
カメラ	Point Grey DR2-08S2C-EX-CSWW
ディスプレイ	Sharp LL-T1620-B

表 1 使用した PC とカメラとディスプレイ

上下左右と斜めに隣接する 8 近傍の画素しか参照しないため、ノイズの影響を受け易く、ノイズが大きい場合勾配を正確に求めることができない。オペレータのサイズを大きくするとノイズの影響を低減することができるが、計算量は  $O(n^2)$  となってしまう。

ここで、コンピュータビジョンにおける勾配方向とは、画像を曲面とみなした場合には登りとなる方向であり、注目画素より明るくなっている向きを表している。つま



図 8 実験動画のキャプチャ画像 1

り、注目画素の周りを見て、明るくなっている向きがわかれば、勾配方向として決定できる。そこで、注目画素の周りの画素を 1 周求め、正弦波、余弦波をかけあわせて、最も明るくなっている方向を求めることが出来る。この方法では、1 周画素を参照する半径は任意に決定できるので、ノイズに影響を受けにくい勾配方向を求められ、計算量はたかだか  $O(2\pi n)$  である。

#### 4.2 勾配を用いた尤度評価

まず、6 パラメータによって描かれる四角形の各辺の法線ベクトルを求める。次に、法線ベクトルを求めた格辺上にそれぞれ個のサンプリング点とり、各点の勾配ベクトルを計算する。

四角形上の各点の法線ベクトルと勾配ベクトルの内積の和を尤度とすると、ディスプレイの画面の枠を計測出来ているかを表す尤度評価が行うことが出来る。

### 5. 実験

本章では、提案手法を実装した実験結果を示す。

実験では位相計算型の勾配ベクトルを用いた Particle Filter により位値・姿勢が推定出来ていることを確かめる。実験に使用した機材を表 1 と図 8 に、そして、図 8 に実験動画のキャプチャ画像を示す。



図 9 実験動画のキャプチャ画像 2

1 フレーム毎に歪み補正を行い、各仮説について、仮説が持つ 6 パラメータによって描かれる四角形上にサンプル点を取り、各点における勾配ベクトルを計算して尤度を求める。尤度が高い順の 20 仮説について、尤度に基づく重み付けを計算し、期待値によって描かれる四角形を赤色の線で描画したものを示す。

図 8 では、ほぼディスプレイの画面の枠を赤色の四角形で捉えられていて、ディスプレイの計測は行えていると考えられる。しかし、実験ではカメラの位置・姿勢の真値を求められていないため、推定された結果の精度が求められていないという問題がある。

## 6. おわりに

本研究では、画像平面上の四角形を位置・姿勢を表す 6 つのパラメータで表現することで、大きさが既知のディスプレイをカメラで撮影したとき、画像上のディスプレイの画面の枠と四角形を一致させることで、撮影するカメラの位置・姿勢を推定する手法を提案した。

提案手法では、位相計算型の勾配ベクトル決定法を利用した Particle Filter を用いることで、安定かつ高速に、カメラの位置・姿勢が推定される。

位相計算型の勾配ベクトルを用いた Particle Filter によりディスプレイの計測が行えており、カメラの位置・姿勢が推定出来ていることを確認した。

今後の課題として、画像上のディスプレイの画面の枠と四角形を一致しているかの評価が主観的にしか出来ていないため、シミュレーション実験で精度評価を行い、どの位の精度で推定出来ているのか確かめる必要がある。また、図 9 はカメラをディスプレイに近づけたときにディスプレイの画面の枠の追跡が一時的に外れ、計測に失敗している様子を示す。仮説数の増加やランダムサンプリングのパラメータを調整する

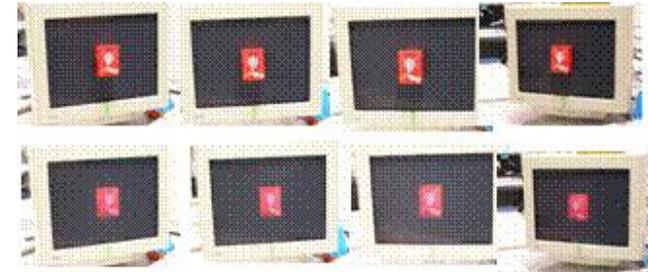


図 10 計測した 3 次元位置から提示した 3 次元物体

ことで、精度が向上すると考えられるが、実時間性が損なわれる可能性があるため、仮説数を増やしても、高速に処理を行うための適切なパラメータの調整が必要がある。

さらに高速かつ高精度にカメラの位置・姿勢が求められれば、図 10 のような能動ステレオカメラを用いて視点の 3 次元位置を計測し、3 次元物体の提示を行った結果が、首振り機構を持たない単眼カメラの位置に合わせた 3 次元物体の提示を行うことで出来るようになると思われる。

## 参考文献1

- 1) 三輪創平, 陳謙, 和田俊和: 実時間視点追従機能を有する立体映像提示法, 画像の認識・理解シンポジウム (MIRU2006), pp. 740-745, (2006)
- 2) M.Isard and A.Blake: "CONDENSATION-Conditional Density Propagation for Visual Tracking", International Journal of Computer Vision, pp.1-36(1998)
- 3) D.C.Brown: "Close Range Camera Calibration", Photogrammetric Engineering, vol.37, No.8, pp.855-866(1971)
- 4) Zhengyou Zhang: "A flexible new technique for camera calibration", IEEE TRANSACTIONS ON PATTERN AND MACHINE INTELLIGENCE, vol.PA-3, No.4, pp.323-344(2000)