

動作者の視点からのカメラ画像によるモーションキャプチャとロボット作業教示への応用

横小路 泰義[†] 北岡 佑輝[†] 吉川 恒夫[†]

本論文では、将来の家庭内でのヒューマノイドの応用を視野に入れ、動作者の視点からのカメラ映像を用いた一種の「実演による教示」の手法を提案する。実演者の動作は、実演者の目の近くに位置する頭部装着型のステレオカメラでとらえられる。実演者の手先および作業環境内に取り付けられたマーカを同時に追跡することにより、実演者の手先の動きだけでなく、実演者の頭部の動きも推定することができる。頭部の動きから推定される注視点の情報は、能動カメラシステムが搭載されたロボットにおいて動作を再生する際に有用な情報となりうる。実験結果から、提案手法の有効性が示される。

Motion Capturing from Demonstrator's Viewpoint and Its Application to Robot Teaching

YASUYOSHI YOKOKOHI,[†] YUKI KITAOKA[†] and TSUNEO YOSHIKAWA[†]

In this paper, we propose a kind of "teaching by demonstration" method, aiming at its application to humanoid robots at home in the future. The demonstrator's motion is captured by a pair of stereo cameras mounted on his/her head, locating very close to his/her eyes. By tracking the landmarks attached to the demonstrator's hand and the working environment, one can estimate not only the demonstrator's hand motion but also his/her head motion, which can be used for the active vision system. Experimental result shows the effectiveness of the proposed framework.

1. はじめに

近年ヒューマノイドロボットの研究がさかんに行われており、将来の民生応用も期待されている。ヒューマノイドは、その形態ゆえ人間の住環境との親和性が高いだけでなく、ユーザによる「実演による教示」(teaching by demonstration)での作業教示のしやすさが大きな特徴である。すなわちユーザの実演をキャプチャできれば、関節角レベルにおいてもほぼそのままの形でロボットに適用でき、教示中の実演者の手、足、頭などの動きすべてがその作業において重要な意味を持ちうる。このような作業教示のしやすさは、将来家庭にロボットが普及するために非常に重要である。図1に、家庭でのヒューマノイドロボット利用のイメージ図を示す。

人間の動きを認識・計測する手法として、機械式のもの(ゴニオメータ)やカメラなどの視覚センサ、磁気センサ、ジャイロなどの各種センサを用いたものが

あるが、計測の簡便さや正確さの点から視覚センサを用いたものが広く用いられている。実際にビデオゲームの格闘技などにおけるキャラクタの動きをより自然なものにする手法として、高速ビデオカメラによるモーションキャプチャが実用化されている^{1),2)}。しかしながらこれらは主に全身運動をキャプチャするためのものであり、8台程度のカメラを実演者の周りを取り囲むように配置する。市販のモーションキャプチャのように特殊なマーカを用いることなく人間の動きを推定する方法の例としては、山本ら^{3),4)}が多リンクモデルと見なした人間の動きを単眼または多視点のカメラ映像からとらえられた自然シーンから時空間勾配法により推定する方法を示している。一方、ロボットの動作を「実演による教示」で与える試みには國吉ら⁵⁾や池内ら⁶⁾の研究がある。これらの例でも実演者の動きや作業対象物は複数台のカメラやレンジファインダによりキャプチャされるが、これらの視覚センサは環境に固定されていた。

図1中のヒューマノイドロボットのモデルは、新エネルギー・産業技術総合開発機構の委託を受けて「人間協調・共存型ロボットシステム研究開発」で開発されたものである。

[†] 京都大学
Kyoto University



図 1 家庭でのヒューマノイドロボット利用のイメージ図
Fig. 1 Humanoid robot in household.

しかしながらこのような環境に固定したカメラ配置は、以下の理由により一般家庭内での「実演による教示」の目的にはそぐわない。第 1 に複数台のカメラを家庭内のあらゆる場所に配置するのは、システムの規模が大きくなりすぎ現実的でなく、第 2 に家庭内でヒューマノイドに行わせる作業は、格闘技のような全身運動よりはむしろ、主に上半身動作またはより細かい手元動作が中心になると予想され、たとえ壁や天井各所にカメラを配置しても撮影距離やオクルージョンの問題から精度良くモーションキャプチャできる保証はない。図 1 に示した例では、風呂桶の掃除以外は主に上半身の動作であり、特に図 1 (a) の例では、背後からのカメラでは動作をほとんどとらえることができないことが分かる。

そこで本研究では視覚センサであるステレオカメラを実演者の目の位置の近くに装着することを提案し、これによりシステムの規模を小さくできるとともに人間が作業をしやすい視点を自然に選ぶことによりオクルージョンの減少が期待できること、さらには教示動作として腕や手の動きだけでなく、作業中に実演者がどの点を注視していたかも同時に教示できることを示す。以下本論文では、2 章において基本的な定式化について述べたあと 3 章において我々が行った実験結果について述べ、考察を加える。最後に 4 章で本論文のまとめを示す。

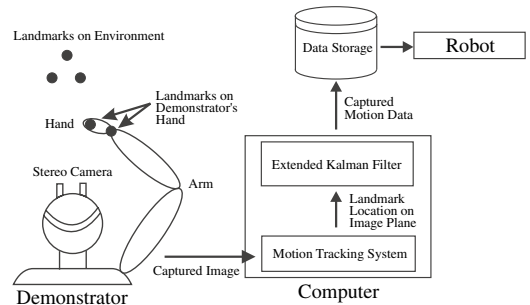


図 2 システム構成
Fig. 2 System configuration.

2. 頭部搭載カメラによるモーションキャプチャ

2.1 システムの概要

図 2 に本論文で扱うシステムの概要を示す。作業教示者は自分の目の位置付近に CCD カメラ 2 台を装着しステレオ視とする。環境および手先にはマーカが取り付けられ、カメラからの映像は動画像処理装置に送られて画像中のマーカが追跡される。各フレームで得られたマーカの画面上での位置から頭(カメラ)の位置・姿勢および手先の位置・姿勢が拡張カルマンフィルタにより推定され、得られた教示データが保存される。その後、得られた教示データをもとにロボットに作業を行わせる。

バーチャルリアリティ(VR)やオーグメンテッドリアリティ(AR)の分野では、映像提示のデバイスとしてヘッドマウントディスプレイ(HMD)が用いられるが、ユーザの頭部運動に応じた映像提示をするためにユーザ頭部の動き(姿勢 3 成分もしくは並進も含めた 6 成分)を正確に追跡することが重要である。頭部の運動をトラッキングする方法としては磁気センサを用いる方法が簡便であるが、ユーザ頭部に搭載されたカメラから環境に固定された複数のマーカをとらえ、これよりユーザ頭部の位置と姿勢を推定する画像ベースの方法もよく用いられる^{7),8)}。すばやい頭部の運動を正確にトラッキングするためには、30 Hz のビデオレートは遅く latency も問題となるので、筆者らの手法を含め画像とジャイロや加速度計などの他のセンサを相補的に用いる手法がとられる場合もある^{9),10)}。

環境に配置した十分な数のマーカが頭部搭載のカメラによりとらえられていれば、VR におけるヘッドトラッキングと同様に、作業者の頭部運動を推定することは可能である。このとき、カメラの視野内に実演者自身の手先に配置したマーカが同時にとらえられていれば、頭部搭載カメラから見た相対的な手先の位置姿

勢も同様に推定できる．環境に固定したマーカにより環境に対する頭部搭載カメラの位置姿勢は推定できているので，結局手先の動きはユーザ頭部からの相対運動としても，環境からの絶対運動としても推定できることになる．また作業者の頭部運動の推定値からは，近似値ながら作業者の作業中の注視点の位置も推定できる．

図 2 で示したように，提案するシステムではステレオカメラとしているが，単眼カメラでもトラッキングは可能である．たとえば，3次元空間内での頭部運動を推定するには，単眼でも最低限一直線上にない3点のマーカがトラッキングできればよい．しかしながら，カメラ画像面上に投影されたマーカの位置は，カメラの光軸方向（奥行き方向）の動きに対して鈍感である．このことは，画像ベースのトラッキングでは，カメラの光軸方向の精度がそれと直交する他の2方向に比べて悪いことを意味する．これは実映像と仮想映像との画像面上での重ね合わせ精度のみが問題となるARではさほど問題とならないものの，3次元空間内での絶対的な運動追跡の精度が問題となるモーションキャプチャでは無視できない．頭部搭載カメラをステレオ視とすることにより，このカメラの奥行き方向の精度の向上が期待できる．

2.2 頭部装着型カメラの形式

人間実演者の頭部に装着したカメラ映像から実演者の手先動作および作業中の注視点を教示データとして獲得するために必要な頭部装着型カメラの形式として，以下の2種類が考えられる．

- カメラのみを頭部に装着するタイプ（HMC（Head-Mounted Camera）型）
- ビデオスルー型のヘッドマウントディスプレイ（VST-HMD型）

VST-HMD型は，カメラがとらえた映像そのものが実演者に提示されるので，実演者が見ているものは確実にカメラでとらえられていることが保証される．ただし，一般にHMDの表示系の視野角は狭く，この視野角に合わせたカメラを装着すると，実演者の視野角はHMD非装着時に比べて大きく制限されることとなる．またHMDの重量が重ければ，HMDの装着自体が負担となる．

一方，HMC型の場合は，カメラのみの装着となるので，小型軽量のカメラを選定すれば，実演者への装着の負担も少ない．また実演者の視野を妨げることもないので，実演者は通常感覚で実演をすることが可能である．ただし，カメラの視野角が小さいと，実演者の不注意で動作獲得のために重要な映像がカメラの

視野角から外れる恐れがある．またカメラを装着するごとに，カメラのレンズ中心が実演者の視線を正面に向けたときと一致するようなキャリブレーションが必要である点もこのタイプの欠点である．

本論文の実験システムでは，検討の結果HMC型を採用した．詳細は3.2節で述べる．ただし以下の定式化は，頭部装着型カメラの形式を問わずに適用可能である．

2.3 ステレオカメラでの透視変換

図3に示すように2台のカメラを平行に配置し，左カメラ座標系を Σ_l ，右カメラ座標系を Σ_r とする． Σ_l, Σ_r はそれぞれ z 軸をカメラの奥行き方向に， x 軸を2台のカメラのベースライン方向にとり， y 軸はそれらと右手系をなすようにとる．さらに Σ_l および Σ_r と座標軸の方向が同じで，ベースラインの midpoint に原点を持つカメラ座標系 Σ_c を定義する．また両カメラの画像平面上に，図3に示すように， x 軸と y 軸がそれぞれカメラ座標系のそれらと同じ向きになるようにスクリーン座標系を設定する．

2台のカメラ間のベースライン距離を D とすれば， Σ_c から見て ${}^c\hat{p} = ({}^c\hat{p}_x, {}^c\hat{p}_y, {}^c\hat{p}_z)^T \in \mathbb{R}^3$ （以下 \mathbb{R}^n は n 次元ベクトルの集合， $\mathbb{R}^{n \times m}$ は $n \times m$ 行列の集合とする）の位置にある点 P は左右のスクリーン上においてそれぞれ以下の点に投影される¹¹⁾．

$${}^l u_x = \frac{f({}^c\hat{p}_x + D/2)}{s_x {}^c\hat{p}_z}, \quad {}^l u_y = \frac{f {}^c\hat{p}_y}{s_y {}^c\hat{p}_z} \quad (1)$$

$${}^r u_x = \frac{f({}^c\hat{p}_x - D/2)}{s_x {}^c\hat{p}_z}, \quad {}^r u_y = \frac{f {}^c\hat{p}_y}{s_y {}^c\hat{p}_z} \quad (2)$$

ここで， f はカメラの焦点距離， s_x, s_y はピクセル比を示す．

2.4 イメージャコピアン の導出

ある1つのマーカに関してその画像平面上の位置が

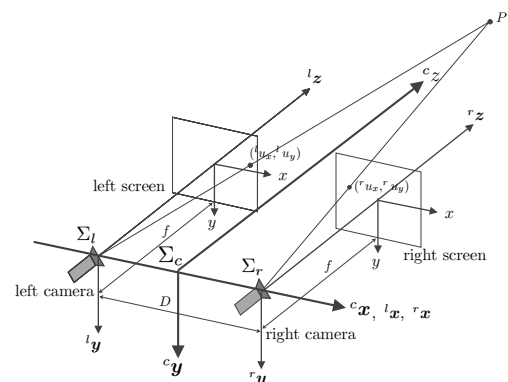


図3 カメラ座標系の定義

Fig. 3 Stereo cameras and their coordinate frames.

$$\mathbf{u} = \begin{pmatrix} {}^l u_x & {}^l u_y & {}^r u_x & {}^r u_y \end{pmatrix}^T \in \mathbb{R}^4 \quad (3)$$

であるとし、画像面上でのマーカの速度とカメラ座標系におけるマーカの速度をそれぞれ $\dot{\mathbf{u}}$ 、 ${}^c \dot{\mathbf{p}}$ とすると次式が成り立つ。

$$\dot{\mathbf{u}} = \mathbf{J}_c({}^c \hat{\mathbf{p}}) {}^c \dot{\mathbf{p}} \quad (4)$$

ここで $\mathbf{J}_c({}^c \hat{\mathbf{p}}) \in \mathbb{R}^{4 \times 3}$ はイメージヤコビアンとして知られている行列であり、ここでは次式で与えられる。

$$\mathbf{J}_c({}^c \hat{\mathbf{p}}) = \begin{pmatrix} \frac{f}{s_x} \frac{c}{c \hat{p}_z} & 0 & -\frac{f({}^c \hat{p}_x + D/2)}{s_x} \frac{c}{c \hat{p}_z^2} \\ 0 & \frac{f}{s_y} \frac{c}{c \hat{p}_z} & -\frac{f}{s_y} \frac{c}{c \hat{p}_z^2} \frac{c}{c \hat{p}_y} \\ \frac{f}{s_x} \frac{c}{c \hat{p}_z} & 0 & -\frac{f({}^c \hat{p}_x - D/2)}{s_x} \frac{c}{c \hat{p}_z^2} \\ 0 & \frac{f}{s_y} \frac{c}{c \hat{p}_z} & -\frac{f}{s_y} \frac{c}{c \hat{p}_z^2} \frac{c}{c \hat{p}_y} \end{pmatrix} \quad (5)$$

図4に示すように、環境に基準座標系 Σ_0 を固定し、教示者の手先にはハンド座標系 Σ_h を固定する。 Σ_0 から見た Σ_c の位置・姿勢を ${}^0 \mathbf{p}_c \in \mathbb{R}^3$ 、 ${}^0 \mathbf{R}_c \in \mathbb{R}^{3 \times 3}$ とし、同様に Σ_h の位置・姿勢を ${}^0 \mathbf{p}_h \in \mathbb{R}^3$ 、 ${}^0 \mathbf{R}_h \in \mathbb{R}^{3 \times 3}$ とする。ただし、 ${}^0 \mathbf{R}_c$ 、 ${}^0 \mathbf{R}_h$ は 3×3 の姿勢行列である。また環境に固定された複数個の基準マーカ位置を ${}^0 \mathbf{p}_{m_i} \in \mathbb{R}^3$ ($i = 1, 2, \dots$) と表すと、カメラ座標系から見たこれら基準マーカの位置 ${}^c \hat{\mathbf{p}}_{m_i} \in \mathbb{R}^3$ は、

$${}^c \hat{\mathbf{p}}_{m_i} = {}^0 \mathbf{R}_c^{-1} ({}^0 \mathbf{p}_{m_i} - {}^0 \mathbf{p}_c) \quad (i = 1, 2, \dots) \quad (6)$$

で与えられる。

教示者の手先にも複数個の手先マーカが配置されており、 Σ_h の原点からこれら手先マーカへの位置ベクトルを ${}^h \mathbf{r}_j$ ($j = 1, 2, \dots$) と表す。カメラ座標系から見たこれら手先マーカの位置 ${}^c \hat{\mathbf{p}}_{h_j} \in \mathbb{R}^3$ は、

$${}^c \hat{\mathbf{p}}_{h_j} = {}^0 \mathbf{R}_c^{-1} ({}^0 \mathbf{p}_h - {}^0 \mathbf{p}_c + {}^0 \mathbf{R}_h {}^h \mathbf{r}_j) \quad (j = 1, 2, \dots) \quad (7)$$

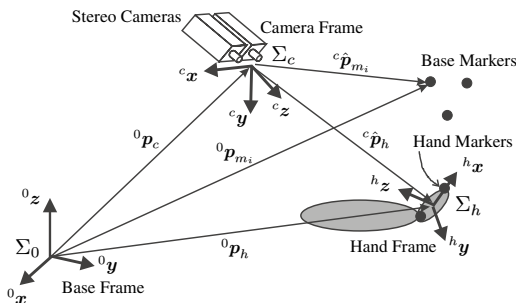


図4 基準座標系と手先座標系の定義

Fig. 4 Base coordinate frame and hand coordinate frame.

と表される。これら式(6)、(7)を時間微分して以下のように速度間の関係式を得る。

$${}^c \dot{\hat{\mathbf{p}}}_{m_i} = {}^c \boldsymbol{\omega}_0 \times \{ {}^0 \mathbf{R}_c^{-1} ({}^0 \mathbf{p}_{m_i} - {}^0 \mathbf{p}_c) \} - {}^0 \mathbf{R}_c^{-1} {}^0 \dot{\mathbf{p}}_c \\ = -{}^c \boldsymbol{\omega}_c \times {}^c \hat{\mathbf{p}}_{m_i} - {}^c \mathbf{v}_c \quad (i = 1, 2, \dots) \quad (8)$$

$${}^c \dot{\hat{\mathbf{p}}}_{h_j} = {}^c \boldsymbol{\omega}_0 \times \{ {}^0 \mathbf{R}_c^{-1} ({}^0 \mathbf{p}_h - {}^0 \mathbf{p}_c + {}^0 \mathbf{R}_h {}^h \mathbf{r}_j) \} \\ + {}^0 \mathbf{R}_c^{-1} \{ {}^0 \dot{\mathbf{p}}_h - {}^0 \dot{\mathbf{p}}_c + {}^0 \boldsymbol{\omega}_h \times ({}^0 \mathbf{R}_h {}^h \mathbf{r}_j) \} \\ = -{}^c \boldsymbol{\omega}_c \times {}^c \hat{\mathbf{p}}_{h_j} + {}^c \mathbf{v}_h - {}^c \mathbf{v}_c + {}^c \boldsymbol{\omega}_h \times {}^c \mathbf{r}_j \quad (j = 1, 2, \dots) \quad (9)$$

ここに、 ${}^c \boldsymbol{\omega}_0 \in \mathbb{R}^3$ はカメラ座標系 Σ_c から見た基準座標系 Σ_0 の回転速度ベクトルであり、逆に ${}^c \mathbf{v}_c \in \mathbb{R}^3$ 、 ${}^c \boldsymbol{\omega}_c \in \mathbb{R}^3$ は Σ_0 に対する Σ_c の並進と回転の速度ベクトルである。また左肩の添え字 c はこれらのベクトルをカメラ座標系 Σ_c で表現していることを意味する。相対関係から ${}^c \boldsymbol{\omega}_0 = -{}^c \boldsymbol{\omega}_c$ である。同様に ${}^c \mathbf{v}_h \in \mathbb{R}^3$ 、 ${}^c \boldsymbol{\omega}_h \in \mathbb{R}^3$ は、基準座標系に対するハンド座標系 Σ_h の並進と回転の速度をカメラ座標系で表現したベクトルを表す。

式(3)に依って基準マーカおよび手先マーカがカメラの画像平面上に投影された位置をそれぞれ4次元ベクトル $\mathbf{u}_{m_i} \in \mathbb{R}^4$ ($i = 1, 2, \dots$)、 $\mathbf{u}_{h_j} \in \mathbb{R}^4$ ($j = 1, 2, \dots$) で表すとすると、式(4)から次式を得る。

$$\dot{\mathbf{u}}_{m_i} = \mathbf{J}_c({}^c \hat{\mathbf{p}}_{m_i}) {}^c \dot{\hat{\mathbf{p}}}_{m_i} \quad (i = 1, 2, \dots) \quad (10)$$

$$\dot{\mathbf{u}}_{h_j} = \mathbf{J}_c({}^c \hat{\mathbf{p}}_{h_j}) {}^c \dot{\hat{\mathbf{p}}}_{h_j} \quad (j = 1, 2, \dots) \quad (11)$$

ここで、今後の定式化のためにカメラ座標系とハンド座標系の速度をすべて基準座標系で表現することとし、

$${}^0 \hat{\mathbf{v}} = \begin{pmatrix} {}^0 \mathbf{v}_c^T & {}^0 \boldsymbol{\omega}_c^T & {}^0 \mathbf{v}_h^T & {}^0 \boldsymbol{\omega}_h^T \end{pmatrix}^T \in \mathbb{R}^{12} \quad (12)$$

を用いて式(8)、(9)、(10)、(11)をまとめると以下の関係式を得る。

$${}^i \dot{\mathbf{u}}_{m_i} = \begin{pmatrix} -\mathbf{J}_c({}^c \hat{\mathbf{p}}_{m_i}) {}^0 \mathbf{R}_c^{-1} & \mathbf{J}_c({}^c \hat{\mathbf{p}}_{m_i}) [{}^c \hat{\mathbf{p}}_{m_i} \times] {}^0 \mathbf{R}_c^{-1} \\ \mathbf{O}_3 & \mathbf{O}_3 \end{pmatrix} {}^0 \hat{\mathbf{v}} \\ = \mathbf{J}_{m_i} {}^0 \hat{\mathbf{v}} \quad (i = 1, 2, \dots) \quad (13)$$

$${}^i \dot{\mathbf{u}}_{h_j} = \begin{pmatrix} -\mathbf{J}_c({}^c \hat{\mathbf{p}}_{h_j}) {}^0 \mathbf{R}_c^{-1} & \mathbf{J}_c({}^c \hat{\mathbf{p}}_{h_j}) [{}^c \hat{\mathbf{p}}_{h_j} \times] {}^0 \mathbf{R}_c^{-1} \\ \mathbf{J}_c({}^c \hat{\mathbf{p}}_{h_j}) {}^0 \mathbf{R}_c^{-1} & -\mathbf{J}_c({}^c \hat{\mathbf{p}}_{h_j}) [{}^c \mathbf{r}_j \times] {}^0 \mathbf{R}_c^{-1} \end{pmatrix} {}^0 \hat{\mathbf{v}} \\ = \mathbf{J}_{h_j} {}^0 \hat{\mathbf{v}} \quad (j = 1, 2, \dots) \quad (14)$$

ここに $\mathbf{O}_3 \in \mathbb{R}^{3 \times 3}$ は 3×3 の零行列であり、 $[\cdot \times]$ は、ベクトル外積に等価な歪対称行列である。このように

して得られた $J_{m_i} \in \mathbb{R}^{4 \times 12}$, $J_{h_j} \in \mathbb{R}^{4 \times 12}$ はそれぞれ画像平面上の基準マーカ速度, 手先マーカ速度とカメラおよび手先の基準座標系から見た速度との関係を表すイメージャコピアンである.

2.5 拡張カルマンフィルタ

以下では後に説明する実験での条件に合わせて手先の運動は水平面内に限られるものとする. したがって手先座標系 Σ_h の位置姿勢は平面内での並進 2 次元と平面の法線周りの回転 1 次元で表すものとする. ここまでの定式化では一般性を持たせるために基準マーカや手先マーカの本数は任意としてきたが, 以下では基準マーカの本数は 3 次元空間内のカメラ座標系の運動を推定するために最低必要な 3 個とし, 手先マーカの本数は平面内に限られる手先の運動を推定するために最低必要な 2 個とする.

システムの離散型の状態方程式は, k をサンプル時刻を表す整数として以下で与えられる.

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{G}_k(\mathbf{x}_k)\mathbf{w}_k \quad (15)$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{d}_k \quad (16)$$

ここに $\mathbf{x}_k \in \mathbb{R}^{18}$ は, 基準座標系から見たカメラ座標系の位置・姿勢 (ロール・ピッチ・ヨー角) および手先座標系の水平面での位置・姿勢とそれぞれの速度・角速度の計 18 個の要素からなる状態変数ベクトルであり次式で定義される.

$$\mathbf{x}_k = \begin{pmatrix} {}^0\mathbf{p}_c^T(t_k) & {}^0\phi_c^T(t_k) & {}^0\mathbf{v}_c^T(t_k) & {}^0\boldsymbol{\omega}_c^T(t_k) \\ {}^0p_{h_x}(t_k) & {}^0p_{h_y}(t_k) & {}^0\phi_{h_z}(t_k) \\ {}^0\dot{p}_{h_x}(t_k) & {}^0\dot{p}_{h_y}(t_k) & {}^0\dot{\phi}_{h_z}(t_k) \end{pmatrix}^T \quad (17)$$

また $\mathbf{y}_k \in \mathbb{R}^{20}$ は, 基準マーカ 3 個と手先マーカ 2 個について式 (3) に相当する画面上のマーカ位置ベクトルを縦に並べた計 20 個の要素からなる観測ベクトルであり次式で定義される.

$$\mathbf{y}_k = \begin{pmatrix} \mathbf{u}_{m_1}^T(t_k) & \mathbf{u}_{m_2}^T(t_k) & \mathbf{u}_{m_3}^T(t_k) \\ \mathbf{u}_{h_1}^T(t_k) & \mathbf{u}_{h_2}^T(t_k) \end{pmatrix}^T \quad (18)$$

また式 (15), 式 (16) において $\mathbf{w}_k \in \mathbb{R}^{18}$ はシステム雑音, $\mathbf{d}_k \in \mathbb{R}^{20}$ は観測雑音である. 式 (15) 中の $\mathbf{f}_k(\mathbf{x}_k)$ は各位置・姿勢が等速で変化するものとして以下のように定める.

$$\mathbf{f}_k(\mathbf{x}(t_k)) = \begin{pmatrix} {}^0\mathbf{p}_c(t_k) + {}^0\mathbf{v}_c(t_k)\Delta t \\ \phi_c(t_k) + \mathbf{T}(\phi_c(t_k))^{-1}\boldsymbol{\omega}_c(t_k)\Delta t \\ \mathbf{v}_c(t_k) \\ \boldsymbol{\omega}_c(t_k) \\ {}^0p_{h_x}(t_k) + {}^0\dot{p}_{h_x}(t_k)\Delta t \\ {}^0p_{h_y}(t_k) + {}^0\dot{p}_{h_y}(t_k)\Delta t \\ {}^0\phi_{h_z}(t_k) + {}^0\dot{\phi}_{h_z}(t_k)\Delta t \\ {}^0\dot{p}_{h_x}(t_k) \\ {}^0\dot{p}_{h_y}(t_k) \\ {}^0\dot{\phi}_{h_z}(t_k) \end{pmatrix} \quad (19)$$

ここに $\Delta t = t_{k+1} - t_k$ はサンプリング間隔を表し, 状態変数はビデオレート (30 [Hz]) で更新されるので $\Delta t = 33.3$ [msec] となる. $\mathbf{T}(\phi_c(t_k)) \in \mathbb{R}^{3 \times 3}$ は, カメラ座標系のロール・ピッチ・ヨー角 $\phi_c = (\phi \ \theta \ \psi)^T \in \mathbb{R}^3$ の変化速度 $\dot{\phi}_c$ を角速度 $\boldsymbol{\omega}_c$ に変換する行列であり, 次式で与えられる.

$$\mathbf{T}(\phi_c) = \begin{pmatrix} 0 & -\sin \phi & \cos \phi \cos \theta \\ 0 & \cos \phi & \sin \phi \cos \theta \\ 1 & 0 & -\sin \theta \end{pmatrix} \quad (20)$$

また式 (16) 中の $\mathbf{h}_k(\mathbf{x}_k)$ はシステムの状態を観測値に投影する関数であり, 式 (1), (2) と等価なものである. $\mathbf{G}_k(\mathbf{x}_k) \in \mathbb{R}^{18 \times 18}$ はシステム雑音と状態変数の関係を表す一般に正則な行列であるが, ここでは後述するように教示者の頭と手先のランダムな加速度に起因する変動分をシステム雑音として扱い, \mathbf{G}_k はつねに単位行列となる.

得られた状態方程式 (15), (16) を線形化し, 拡張カルマンフィルタを求めると以下ようになる.

フィルタ方程式:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k[\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1})] \quad (21)$$

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}_k(\hat{\mathbf{x}}_{k|k}) \quad (22)$$

カルマンゲイン:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{Q}_{d_k}]^{-1} \quad (23)$$

誤差の共分散行列:

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} \quad (24)$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k\mathbf{P}_{k|k}\mathbf{F}_k^T + \mathbf{G}_k\mathbf{Q}_{w_k}\mathbf{G}_k^T \quad (25)$$

ただし, 初期値は

$$\hat{\mathbf{x}}_{0|-1} = \bar{\mathbf{x}}_0, \quad \mathbf{P}_{0|-1} = \mathbf{P}_{x_0} \quad (26)$$

であるとする. また,

$$\mathbf{F}_k = \frac{\partial \mathbf{f}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k} \Big|_{\mathbf{x}_k = \hat{\mathbf{x}}_{k|k}},$$

$$\mathbf{H}_k = \frac{\partial \mathbf{h}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k} \Big|_{\mathbf{x}_k = \hat{\mathbf{x}}_{k|k-1}} \quad (27)$$

である．これ以外にここで用いた記号を以下に示す．

$\hat{\mathbf{x}}_{k|k}$: $\mathbf{y}_0 \sim \mathbf{y}_k$ が与えられたときの \mathbf{x}_k の推定値

$\hat{\mathbf{x}}_{k+1|k}$: $\mathbf{y}_0 \sim \mathbf{y}_k$ が与えられたときの \mathbf{x}_{k+1} の推定値

$\mathbf{P}_{k|k}$: $\hat{\mathbf{x}}_{k|k}$ の推定誤差の共分散行列

$\mathbf{P}_{k+1|k}$: $\hat{\mathbf{x}}_{k+1|k}$ の推定誤差の共分散行列

\mathbf{Q}_{w_k} : w_k の共分散行列

\mathbf{Q}_{d_k} : d_k の共分散行列

$\mathbf{H}_k \in \mathbb{R}^{20 \times 18}$ はその定義から式 (13) の \mathbf{J}_{m_i} および式 (14) の \mathbf{J}_{h_j} をそれぞれマーカの数 ($i = 1, 2, 3$, $j = 1, 2$) だけ縦に並べたものから ${}^0\dot{\mathbf{v}}$ のうち手先の平面内の並進速度 2 成分と角速度の 1 成分に対応する列のみを抜き出したものを用いた．システム雑音はある確率密度を持った白色ガウス性の加速度に起因するものとし、 $\mathbf{G}_k \mathbf{Q}_{w_k} \mathbf{G}_k^T$ のうち \mathbf{G}_k は前述のとおり単位行列とする． \mathbf{Q}_{w_k} は、時刻 k における式 (20) の変換行列 $\mathbf{T}(\phi_c(t_k))$ を \mathbf{T}_k と略記すると

$$\mathbf{W}_{11} = \begin{pmatrix} a_c \mathbf{I}_3 & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \alpha_c \mathbf{T}_k^{-1} \mathbf{T}_k^{-T} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & a_h \mathbf{I}_2 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \alpha_h \end{pmatrix} \quad (28)$$

$$\mathbf{W}_{12} = \begin{pmatrix} a_c \mathbf{I}_3 & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \alpha_c \mathbf{T}_k^{-1} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & a_h \mathbf{I}_2 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \alpha_h \end{pmatrix} \quad (29)$$

$$\mathbf{W}_{22} = \begin{pmatrix} a_c \mathbf{I}_3 & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \alpha_c \mathbf{I}_3 & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & a_h \mathbf{I}_2 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \alpha_h \end{pmatrix} \quad (30)$$

を用いて以下のように定められる¹²⁾．

$$\mathbf{Q}_{w_k} = \begin{pmatrix} \frac{1}{3} \mathbf{W}_{11} \Delta t^3 & \frac{1}{2} \mathbf{W}_{12} \Delta t^2 \\ \frac{1}{2} \mathbf{W}_{12}^T \Delta t^2 & \mathbf{W}_{22} \Delta t \end{pmatrix} \quad (31)$$

ここに \mathbf{I}_2 , \mathbf{I}_3 はそれぞれ 2×2 , 3×3 の単位行列であり、また \mathbf{O} は適当なサイズの零行列である．また a_c はカメラ座標系の運動に加えられる並進加速度 ${}^0\dot{\mathbf{v}}_c$ を白色ガウス性の雑音と見なしたときの確率密度である．したがってこの並進加速度を連続時間で考えたとき、その共分散行列は次式で与えられる．

$$E\{\dot{\mathbf{v}}_c(t) \dot{\mathbf{v}}_c^T(t')\} = a_c \delta(t - t') \mathbf{I}_3 \quad (32)$$

ただし、 $E\{\cdot\}$ は平均を表し、 $\delta(\cdot)$ はデルタ関数である．同様に α_c は回転加速度の確率密度、 a_h は手先の運動に加えられる並進加速度雑音の確率密度、 α_h は回転加速度の確率密度である．式 (31) で与えられる共分散行列は、白色ガウス性の加速度を積分して得られるブラウン運動の共分散行列と考えることができる¹³⁾．文献 12) では姿勢の表現に瞬間回転軸周りの微小回転ベクトルを用いているのに対してここではロールピッチヨー角を用いているため \mathbf{T}_k^{-1} や \mathbf{T}_k^{-T} が一部含まれている．

本節で定式化したように、ステレオカメラによって取り込まれたイメージプレーン上での個々のマーカ位置はステレオマッチングによって 3 次元位置に復元されることなくそのまま拡張カルマンフィルタの観測量として扱われる．これによりたとえステレオカメラの左右どちらかのみでオクルージョンが発生したりしてトラッキングに失敗しても、もう一方のマーカ位置の情報は有効に利用される．トラッキングに失敗したマーカがある場合は、観測ベクトル \mathbf{y}_k から対応する要素を抜くと同時に、式 (13) および式 (14) のイメージヤコビアン \mathbf{J}_{m_i} , \mathbf{J}_{h_j} から対応する列を抜けばよい．多視点からの画像をマッチングなどの前処理をせずにそのまま推定に利用する考え方は Rizzi ら¹⁴⁾ などによって指摘されており、VR の分野でも Welch ら¹⁵⁾ によって提案されている．

3. 実験

3.1 実験内容

今回の実験では日常的な手先を使う作業の 1 つとして、スライド式の扉を有する CD ラックの扉を開けて複数の CD の中から聞きたい CD を取り出し、その後扉を閉める、という作業を設定した．そして得られた教示データのうち手先に関するものだけを用いて、平面 3 自由度のスカラ型ロボットアームに実際に作業を行わせた．実験で用いた座標系の定義を図 5 に示す．図 5 に示すように手先マーカとして教示者の右手の手首の関節付近と親指の付け根付近の 2 カ所にマーカを取り付けた．また環境マーカとして CD ラック上面に 3 点取り付けた．基準座標系の原点は、CD ラックの上面の基準マーカのうち、教示者から向かって一番左側のマーカの作業台面上への射影点にとり、手先座標系の原点は 2 つの手先マーカの中心にとった．基準座標系の向きを教示者が正面を向いたときのカメラ座標系の向きと一致するようにとったため、図 4 に示した座標系の向きと異なることに注意されたい．これにより式 (17) で抜き出される手先位置姿勢の成分も変わ

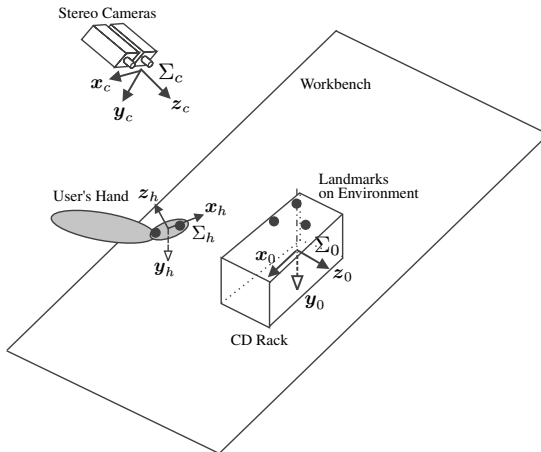


図 5 作業環境と座標系の定義

Fig. 5 Task environment and definition of coordinate frames.

るが本質的ではない．同様に手先座標系の向きも教示者が右手を真横に向けたときに基準座標系のそれと一致するようにとった．マーカのトラッキングが容易となるように，CD ラックを含む環境は白紙で覆い，教示者の右手には白手袋を装着した．

先述したように，実験における手先動作は再生に用いるロボットがスカラ型のため，作業台上の 2 次元面内 ($x-z$ 平面内の並進 2 自由度と y 軸回りの回転 1 自由度) に限定して行った．ただし手先の姿勢は作業台面の外向き法線 ($-y$ 軸) 回り反時計方向を正とした．実演者の頭の動きには拘束はなく，頭部に装着されたカメラがつねにマーカをとらえる限り，3 次元空間内を自由に動かしてもよい．

3.2 実験システム

実験で用いたステレオ CCD カメラの概観を図 6 に示す．小型 CCD カメラ (東芝製 IK-UM40) 2 台をそれぞれ教示者の右目・左目のやや上の位置にくるようにベルトで固定されている．ベースラインの距離は $D = 70$ [mm] である．CCD の有効画素数は水平 768 画素，垂直 494 画素である．カメラには焦点距離 $f = 5.5$ [mm] で水平と垂直の画角がそれぞれ約 50 度，約 38 度の小型レンズ (東芝製 JK-L55U) が装着されている．ピクセル比はカメラの視野角を実測して求め， $s_x = s_y = 7.64 \times 10^{-3}$ [mm/pixel] とした．

2.2 節でも議論したように，頭部装着型カメラには 2 種類の形式が考えられる．本研究においても実験に先立ち，文献 10) でも使用した現有の HMD (島津製作所製のシースルービジョン STV-E) を改造して製作したビデオシースルー型 HMD にカメラを追加してステレオ型としたものを最初に検討したが，水平で 48



(a) 概観



(b) 装着した様子

図 6 頭部装着型ステレオカメラ

Fig. 6 Head-mounted stereo cameras.

度の視野角ではやはり不十分で，実演者は通常感覚で実演をすることは難しかった．また HMD の重量も約 1.4 kg と重く，長時間の実演では実演者には大きな負担であったため，実演者の視野を妨げず装着の負担も少ない，図 6 の HMC 型を採用することとした．

左右のカメラからの映像は NTSC 信号として出力され，フィールド多重化回路 (カスタム電子製 FC-55) により 1 つの映像に合成された後に富士通製の動き追跡処理装置トラッキングビジョン (VME バス仕様のビデオモジュール VMDL-2，トラッキングモジュール TMDL-2 の 2 枚より構成される) に送られる．

本実験で用いる CCD カメラの映像出力方式はインターレース方式であり，まず最初の 1/60 秒で奇数番目の水平走査線 (奇数フィールド) を走査し，次の 1/60 秒で偶数番目の水平走査線 (偶数フィールド) を走査することにより 1 つのフレーム (1/30 秒) を構成する．通常 2 つの別々の動画像を処理するには 2 台の動画像処理装置を必要とするが，図 7 に示すようにフィールド多重化回路を用いて片方のカメラからの奇数フィールドともう一方のカメラからの偶数フィールドから 1 つのフレームを構成することにより，1 台のトラッキングビジョンで 2 台のカメラからの動画像を処理することができる．本来奇数フィールドと偶数フィールドではシャッタータイミングが違うので，厳密には同時に観測されたマーカ位置とはいえないが¹⁰⁾，本実験での教示動作は比較的遅いので，両フィールドでのシャッタータイミングのずれの影響は無視した．

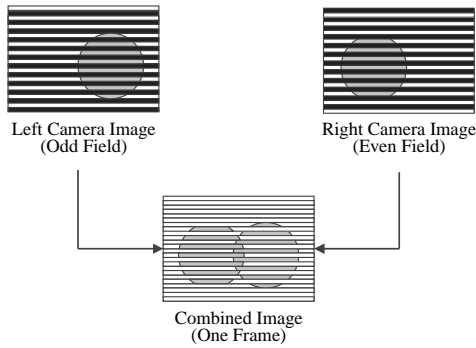


図 7 フィールドの多重化

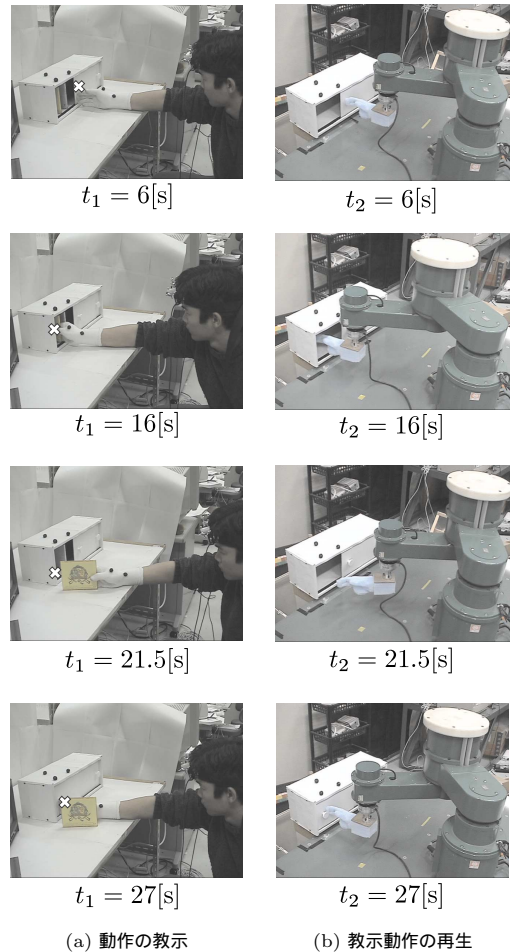
Fig. 7 Conversion to the field sequential format.

トラッキングビジョンのビデオ RAM のサイズは 512 ピクセル \times 512 ピクセルなので、カメラからの画像の周辺部が一部トリミングされる。カメラからの映像はカラーであるが、トラッキングビジョンに取り込まれる段階で白黒画像となる。そしてトラッキングビジョンにより得られたマーカの画像平面上の位置情報を入力として拡張カルマンフィルタにより頭の位置・姿勢および手先位置・姿勢を推定し、その推定値を時系列データとして保存する。カルマンフィルタにおける並進加速度の確率密度は、頭部運動、手先運動ともに $a_c = a_h = 50 \text{ [m}^2/\text{s}^3]$ とし、回転加速度の確率密度は $\alpha_c = \alpha_h = 50 \text{ [rad}^2/\text{s}^3]$ とした。観測雑音の共分散行列 Q_{d_k} は、対角成分が 100 [pixel²] の対角行列とした。

3.3 実験結果

実験における教示の様子を図 8 (a) に、得られた教示データを図 9 に示す。紙面の都合上、図 8 では多くの写真を掲載できないので、教示した動作シーケンスを簡単に説明する。まず CD ラックの扉を開き、中に入っている 2 枚の CD のうち 1 枚を取り出す。教示者から見て CD ケース左側のラベルを見て、これが聞きたい CD ではないことが分かったので CD を元に戻す。次にもう一方の CD を取り出し、同様にラベルを見て聞きたい CD であることを確認してからこれを CD ケースの手前にそのまま立てて置く。最後に CD ラックの扉を再び閉じる。以上が教示した動作シーケンスである。

ここで図 9 において、まず教示者の手先の位置姿勢のデータを見てみよう。手先の x 座標は、CD ラックの扉の開閉方向であり、 $t_1 = 4.5 \text{ [s]}$ 付近の極小ピークから $t_1 = 6 \text{ [s]}$ 付近の極大ピークまでのストローク L_1 が最初に扉を開ける動作に対応している。また $t_1 = 25 \text{ [s]}$ 付近の極大ピークから $t_1 = 27 \text{ [s]}$ 付近の極小ピークまでのストローク L_2 は、最後に扉を閉じ



(a) 動作の教示

(b) 教示動作の再生

図 8 実験の様子

Fig. 8 Overview of the experiment.

る動作である。CD ラックの実際の扉のストロークは 16.8 [cm] であるが、 L_1, L_2 から手先の姿勢を考慮してストロークを計算すると、約 19 [cm] とやや大きめの値となった。これは x 座標のピーク値が必ずしも扉の開閉の始点終点に対応しているとは限らないこと、また教示者の指の曲げ方の違いなどが影響しているようである。また扉の開閉動作両端でのオフセット L_3, L_4 は、扉のノブや指先の厚みによって生じるものである。また手先の z 座標は、CD ラックの奥行き方向に対応し、 $t_1 = 6 \text{ [s]}$ から $t_1 = 18 \text{ [s]}$ にかけての 3 つのピークは CD を取り出したり戻したりするためにラックの中に 3 回手先を入れた動作に対応している。グラフ中の L_5 は 14.5 [cm] であり、CD ケースの横幅 14.2 [cm] からみて妥当な数値といえる。次に教示者の頭部の位置姿勢のデータを見てみると、頭部位置の x 座標と z 座標がそれぞれ手先の x 座標と z 座標と

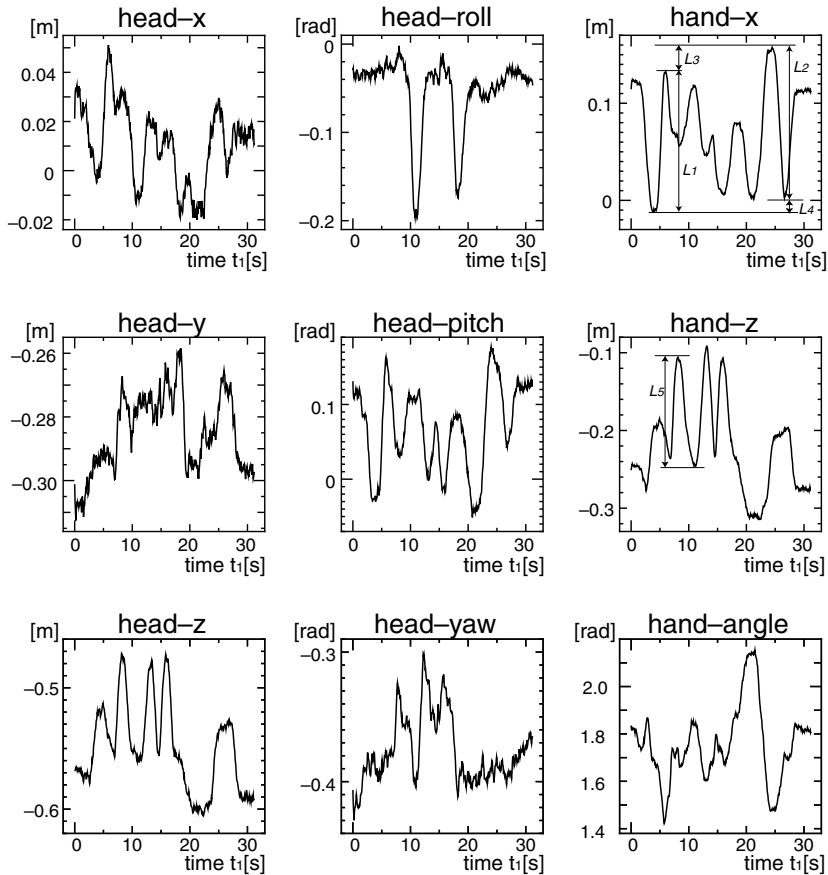


図 9 頭および手先の位置・姿勢

Fig.9 Captured head pose and hand pose.

運動して変化していることが分かる．また頭部姿勢成分のうちロール軸（ z 軸周りの回転）には大きなピークが 2 回現れているが、これは教示者が首を左に傾けて CD ケースのラベルを見る動作に対応している．またピッチ軸成分（ロール軸回転後の y 軸周りの回転）は、再び手先の x 座標の変化とよく連動している．

このようにキャプチャされた動作は、実際の教示動作をよく表現しているが、これらの中から手先に関するものを取り出してこれをロボットに再生させた．この教示データの再生の様子を図 8 (b) に示す．図に示すように、再生に用いたロボットには CD ラック内のコンパクトディスクを把持できるような機構がないので、代わりに梱包などに用いられる発泡プラスチック材で手先の形状を模したものを取り付けられている．図からも分かるように、キャプチャされた手先のデータを用いて扉を開閉させることに成功した．

カメラ（頭部）の位置・姿勢のデータに関してはそのデータをもとにカメラ座標系の z 軸の方向を算出

した．図 8 (a) の各時刻の写真内には、推定した視線方向と CD ラックの扉のある垂直面との交点に相当する大まかな位置に白い \times 印をつけてある．これより推定された視線方向は、主観的な判断ではあるがほぼ教示者の視線と等しいものとなることが分かった．このことから注視点の教示も十分に行えているものと考えられる．

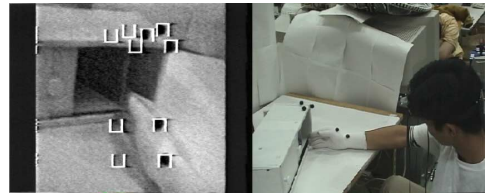
3.4 考 察

今回の実験では教示データに関しては真値を知ることができないので、得られたデータを定量的に評価することはできない．CD ラックの扉を開閉する際に発泡プラスチック材の手先が約 1 cm 程度たわむときがある．これが手先位置の推定誤差に起因するものなのか実演者の手先とロボットに取り付けた手先モデルとの形状の違いなどの他の要因によるものなのかは不明であるが、少なくとも扉を開閉するには十分な精度で手先位置・姿勢の教示データが得られているものと思われる．

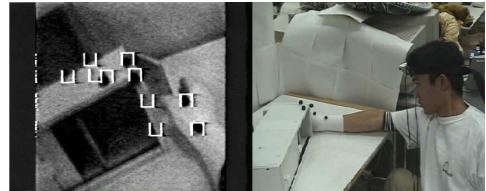
教示データの真値を得る方法としては、人間教示者の代わりに手と頭部を持ったヒューマノイドロボットを用いることができれば最も確実である。また教示者の腕の代わりに今回再生で用いたロボットアームに既知の動作をさせ、それを教示者が観察してどれだけ正確にロボットアームの動きがキャプチャされたかを評価することもできる。人間教示者による場合でも、頭部の動きはパッシブリンクで計測する方法が考えられる。手先の運動については、連続的な評価はできないが、既知の場所のスイッチなどに触れる動作から推定値の精度を評価することができる。3.3節で、得られたデータから扉のストローク長を算出しようとした試みもこれに相当する。また必ずしも真値とはいえないが、環境にカメラを複数台別途設置し、そこから教示者の頭部や手先の動きを観測する方法も考えられる。しかしながらこのとき図 8(a) の $t_1 = 27$ [s] のように、設置するカメラの場所によってはマーカの一部が隠れる場合があることに注意すべきであり、逆にこれは提案手法の有効性を示しているともいえる。今後以上述べた方法のどれかを用いて教示データの真値を測定し、教示データの精度の定量的な評価を行うことが課題となる。

また注視点の推定も、実演者がつねに正面を注視しているという仮定のもとでの推定であり、厳密には眼球の動きを追跡する必要がある。現在実演者頭部に装着可能な小型軽量のものが市販されているので、頭部装着型カメラと組み合わせることは可能であろう。本実験では、動作再生に用いたロボットにカメラを持つ頭部がないため、注視点のデータは用いられなかった。作業環境がまったく同一で、手先の動きが正確に再生できれば、注視点のデータは必要がないともいえる。しかし動作教示時と動作再生時で作業環境や対象物の位置にある程度の変動がある場合や、変形する対象物を扱うような複雑な作業によっては作業中の視覚フィードバックが必須となる場合が考えられ、得られた注視点データはどの位置からどの方向にカメラを向ければよいかを決定する際に用いることができる。

注視点データの有効性を示す例として、図 10 に CD を取り出す別の作業例を示す。図の左側は頭部搭載カメラの左側のカメラからの映像であり、トラッキングビジョンに取り込まれた画像から再びフィールドシーケンシャルコンバータを介して左側カメラの映像を復元したため解像度が劣化している。画像中には基準マーカと手先マーカのトラッキングのためのウィンドウがあり、フィールドシーケンシャルコンバータでの画像復元時に残った右側カメラ用のトラッキングウイ



(a) 通常の視線で CD ラックを見た場合



(b) CD ラックで隠れていた CD が見えるように視線を移した場合

図 10 視線を大きく変更して行った作業の一例

Fig. 10 An example task that needs a large change of the viewing direction.

ンドウも見える。またこの実験では、基準マーカの位置を先の実験から変更している。この作業では目的の CD がラックの右側に隠れて置かれており、図 10(a) の通常の視線方向では CD が見えないので図 10(b) のように頭部を大きく右に移動してラックの右側に回り込むことにより初めて目的の CD を見ることができる。CD を把持する際の位置調整にビジュアルフィードバックが必要ならば、通常の頭部位置に固定したカメラでは、この作業は実行できないことになる。

本実験で用いた CCD カメラの視野角は、本来はビデオスルー型 HMD を構成するために現有の HMD の表示系の視野角（水平 48 度）となるべく一致するように選ばれたものであり、実演者が注意しないと動作獲得のために重要な映像がカメラの視野角から外れる場面がしばしばあった。本実験のためには、より視野角の広いカメラを使用するほうが望ましい。また HMD も年々小型軽量で広視野角のものが市販されつつあるので、再び頭部装着型カメラの形式についても検討する余地がある。

本論文で提案した手法は、原理的には手先位置はまず頭部搭載カメラからの相対位置として求められ、その後別途得られた頭部の位置姿勢の情報を用いての座標変換の後、絶対座標系での位置を得ることができると解釈できる。このため、カメラから見た手先位置の推定誤差に、頭部の位置姿勢の推定誤差が累積し、最終的に大きな推定誤差を生みそうであるが、カルマンフィルタでは手先と頭部の位置姿勢は同列に扱われており、実際このような誤差の累積はないようである。

厳密には、数学的な誤差解析をする必要があり、今後の課題の1つとしたい。また、手先の推定精度も想像以上に良い結果となったが、手先の運動がテーブル面上の2次元平面内という拘束があったためとも解釈できる。今後は、手先の運動も拘束なく自由に3次元空間内を動かせる場合への適用や、ステレオ視と単眼視での性能比較なども行っていきたい。

また本論文での定式化では、実演者の頭部と手先はまったく独立に運動するものとして扱っていたが、実際には実演者の頭と手先は首と腕でつながっている。このように追跡対象の幾何学的モデルに基づいたトラッキング^{3),16)}も今後検討していきたい。これ以外にも今後の課題として、マーカ数の増加による作業範囲の拡大、各種センサの併用によるトラッキング精度の向上、ロボット側の手先への把持機構の付加、手先だけでなく手指の動きのキャプチャ、机上動作だけでなく全身運動への拡張、などがあげられる。

4. おわりに

本論文ではモーションキャプチャによる作業教示の際に視覚センサであるカメラを人間の目の位置に配置することにより、単にその作業だけでなく注視点も同時に教示する手法を提案した。そして実際に教示者の視点に配置したステレオカメラから送られてくる映像のみをもとに頭および手先の位置・姿勢を満足できる精度でキャプチャできた。さらに得られた教示データのうち手先に関するものを用いてロボットアームに簡単な作業を行わせることに成功した。考察でも述べたように、何らかの方法で教示データの真値を得て、提案手法の推定精度を定量的に評価することが今後の課題の中でも重要なものの1つである。

従来のように教示者の周りに複数台のカメラを配置する方法とは違い、本手法では少ないカメラにもかかわらず手元作業などの細かい作業もオクルージョンの影響を受けにくい。これは、教示者が自然に自分が見やすいように自らの視線を選んでいる特性をうまく利用している。考え方によっては作業を実演している本人の目に近い位置が、対象とする動作を観察するのに最も適したカメラ位置であるということもできよう。このように本手法は比較的簡単なシステム構成で、手先の運動だけでなく頭部の位置姿勢すなわち視線方向や作業中の注視点もキャプチャすることができ、将来の家庭内で用いられるヒューマノイドロボットの動作教示法として有効であろうと考えている。ただし、工場内のような環境とは違って、家庭内では教示した動作データを単純に再生しただけでつねに作業が実行可

能になるとは限らない。したがって、今後は教示データをベースにセンサフィールドバックの機能を付加していくことも必要となり、ここで得られた注視点データなどもビジュアルフィードバック機能を付加する際に有用となる。

謝辞 本論文での実験を行うにあたり、(株)東芝研究開発センターの松日楽信人氏より実験装置の一部についてご協力をいただいた。ここに記して感謝の意を表します。本研究は、「人間協調・共存型ロボットシステム」の研究開発(通称 HRP またはヒューマノイドロボットプロジェクト)として新エネルギー産業技術総合開発機構(NEDO)から委託された製造科学技術センターからの再委託により実施した研究の一部である。

参考文献

- 1) <http://www.vicon.com/>
- 2) <http://www.motionanalysis.com/>
- 3) 山本正信, 川田 聡, 近藤拓也, 越川和忠: ロボットモデルに基づく人間動作の3次元画像追跡, 電子情報通信学会論文誌, Vol.J79-D-II, No.1, pp.71-83 (1996).
- 4) 佐藤明知, 川田 聡, 大崎喜彦, 山本正信: 多視点画像からの人間動作の追跡と再構成, 電子情報通信学会論文誌, Vol.J80-D-II, No.6, pp.1581-1589 (1997).
- 5) 國吉康夫, 井上博允, 稲葉雅幸: 人間が実演して見せる作業の実時間視覚認識とそのロボット教示への応用, 日本ロボット学会誌, Vol.9, No.3, pp.295-303 (1991).
- 6) 池内克史, 末広尚士: 視覚による組立作業理解のための作業モデルとそれに基づく動作生成, 日本ロボット学会誌, Vol.11, No.2, pp.281-290 (1993).
- 7) Uenohara, M. and Kanade, T.: Vision-Based Object Registration for Real-Time Image Overlay, *Computers in Biology and Medicine*, Vol.25, No.2, pp.249-260 (1995).
- 8) State, A., Hirota, G., Chen, D.T., Garrett, W.F. and Livingston, M.A.: Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking, *Proc. SIGGRAPH '96*, pp.429-438 (1996).
- 9) You, S., Neumann, U. and Azuma, R.: Hybrid Internal and Vision Tracking for Augmented Reality Registration, *Proc. IEEE VR99*, pp.260-267 (1999).
- 10) 横小路泰義, 菅原嘉彦, 吉川恒夫: 画像と加速度を用いたHMD上での映像の正確な重ね合わせ, 日本バーチャルリアリティ学会論文誌, Vol.4, No.4, pp.589-598 (1999).
- 11) Maru, N., Kase, H. and Yamada, S.: Ma-

nipurator Control by Visual Servoing with the Stereo Vision, *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.1865–1870 (1993).

- 12) Gennery, D.B.: Visual Tracking of Known Three Dimensional Objects, *Int. J. Computer Vision*, Vol.7, No.3, pp.243–270 (1992).
- 13) Maybeck, P.S.: *Stochastic Models, Estimation, and Control (Volume 1)*, Academic Press, New York (1979).
- 14) Rizzi, A.A. and Koditschek, D.E.: An Active Visual Estimator for Dextrous Manipulation, *IEEE Trans. Robotics and Automation*, Vol.12, No.5, pp.697–713 (1996).
- 15) Welch, G. and Bishop, G.: SCAAT: Incremental Tracking with Incomplete Information, *Proc. SIGGRAPH '97*, pp.333–344 (1997).
- 16) Nickels, K. and Hutchinson, S.: Model-Based Tracking of Complex Articulated Objects, *IEEE Trans. Robotics and Automation*, Vol.17, No.1, pp.28–36 (2001).

(平成 13 年 8 月 29 日受付)

(平成 14 年 3 月 8 日採録)

(担当編集委員 加藤 晃市)



横小路泰義

昭和 36 年生。昭和 63 年京都大学大学院工学研究科精密工学専攻博士課程中途退学。昭和 63 年より京都大学工学部助手。平成 4 年より京都大学工学部助教授。平成 6 年より京都大学大学院工学研究科助教授。ロボット工学の研究に従事。博士(工学)。システム制御情報学会, 日本ロボット学会, 計測自動制御学会, 日本機械学会, 日本バーチャルリアリティ学会, バイオメカニズム学会, IEEE, ACM 等の会員。



北岡 佑輝

昭和 53 年生。平成 12 年京都大学工学部物理工学科卒業。平成 14 年京都大学大学院工学研究科修士課程修了。ロボット工学および流体工学の研究に従事。現在松下産業機器株式会社に勤務。



吉川 恒夫

昭和 16 年生。昭和 44 年京都大学大学院博士課程修了。昭和 44 年より京都大学工学部助手。昭和 45 年より京都大学助教授。昭和 61 年より京都大学工学部教授。平成 6 年より京都大学大学院工学研究科教授。ロボット工学および力覚人工現実感の研究に従事。工学博士。システム制御情報学会, 日本ロボット学会, 計測自動制御学会, 日本機械学会, 日本バーチャルリアリティ学会等の会員。米国 IEEE のフェロー。