

T 5 動画像を使った人間動作予測\*

大田 佳人・速水 康晴・川田 聡・山本 正信†

新潟大学 工学部 情報工学科‡

1 まえがき

人間動作の動画像追跡は、遠隔会議における臨場感通信や人間と機械との円滑なコミュニケーション、等において重要な技術となってきた。人間の動作追跡では、手や足が胴体などによって隠される、いわゆるオクルージョンが問題となっている。カメラの数を増し、多数の視点から観測することも解決策 [1] にはあるが、人間が複数ともなると、多数の視点からでも死角がおきる。隠れの問題は人間の動画像追跡において、避けて通ることのできない問題である。本論文では、人間の動作予測問題を対象とする。予測が正確ならば、隠れが起ころともその位置を推定することができる。また、見えている物体の追跡においても、運動パラメータの計算を頻繁に行う必要がなくなる。これまでの研究では、人体が慣性運動をしていると仮定したものほとんどであった [2]。その場合、速度・角速度は一定であるが、人体の動きは必ずしもそうではない。例えば歩行時の腕の振りや、単擬子とみなした方が正確で、その角速度は一定ではない。本論文では、人間の動きを関節角をパラメータとし AR モデルで表現し、カルマンフィルタを使って関節角の予測を行う。

2 人間動作の解析

人間の体を多関節ロボットでモデル化する。ロボットモデルと動画像中の実際の人間を 1 フレーム毎に時空間勾配法により照合し、ロボットモデルの運動パラメータを推定する。

2.1 時空間勾配法

人体上の点は図 1 のように画像面上の点に原点を中心として中心投影されるとする。人体上の 3 次元速度ベクトル  $(v_x, v_y, v_z)$  は、時空間勾配法により次の様にな

る [1]。

$$\frac{E_x}{z}v_x + \frac{E_y}{z}v_y - \frac{1}{z}(XE_x + YE_y)v_z + E_t = 0 \quad (1)$$

ここで、 $X, Y$  は人体上の点の画像上での投影点の値、 $(E_x, E_y)$  は画面上での空間勾配、 $E_t$  は時間勾配である。

また、人体を多関節ロボットアームでモデル化した時、人体上の位置  $p = (x, y, z)^t$  は、アームの関節角などのパラメータ  $q = (q_1, q_2, \dots, q_n)^t$  を用いて、次のように表現できる。

$$p = f(q) \quad (2)$$

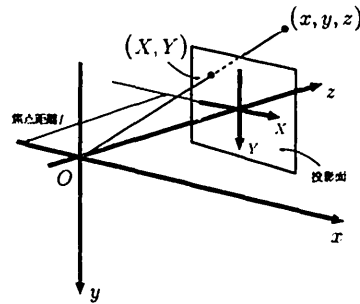


図 1: カメラのモデル

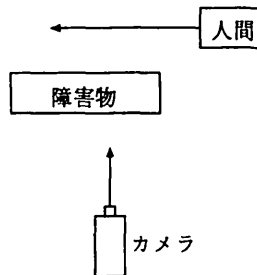


図 2: オクルージョンの発生

\*Prediction of human body in motion using image sequences  
†Yoshito OTA, Yasuharu HAYAMI, Satoshi KAWADA and Masanobu YAMAMOTO

‡Faculty of Engineering, Niigata University

さらに、人体上の速度ベクトル  $\dot{p} = (v_x, v_y, v_z)^t$  とロボットの運動パラメータ  $\dot{q}$  は、ヤコビ行列  $J = \partial f / \partial q$  によって、次の様に関係付けられる。

$$\dot{p} = J\dot{q} \quad (3)$$

この式を式 (1) に代入することによって運動パラメータ  $\dot{q}$  を推定する式が得られる。ロボットモデル上で得られた推定式を連立させ最小二乗法により運動パラメータを推定する。

以上の方法により人間の追跡を行うが、 $n$  フレーム目以降、例えば図2のようなオクルージョンが発生した場合、追跡が不可能となる。このような時、時刻  $n$  以降のフレームの予測を行わなければならないことになる。

### 3 ARモデル

推定する運動パラメータを次のようなARモデル (autoregressive model)、すなわち、過去のパラメータの値で表現できると仮定する。時刻  $n$  の、ある1つの運動パラメータの値を  $y_n$  とすれば、

$$y_n = \sum_{i=1}^m a_i y_{n-i} + v_n \quad (4)$$

ここで、 $a_i$  は自己回帰係数、 $v_n$  は平均0、分散  $\sigma$  の正規分布にしたがう白色雑音である。追跡することができた運動パラメータ値より、各運動パラメータについて、それぞれのARモデルの次数をレビンソン法、最小二乗法、PARCOR法のいずれかによって推定する。

## 4 カルマンフィルタ

### 4.1 状態空間モデル

カルマンフィルタを用いるためにアームパラメータなどの時系列を、状態空間モデルにあてはめて考える必要がある。状態空間モデルは次のようなものである。

$$x_n = F_n x_{n-1} + G_n v_n \quad (5)$$

$$y_n = H_n x_n + w_n \quad (6)$$

ここで、各角速度パラメータが式 (4) のようなARモデルである時、状態ベクトル  $x_n$ 、 $F$ 、 $G$ 、 $H$  は、次のようになる。また、 $v_n$  は平均ベクトル0、分散共分散行列  $Q_n$  の  $m$  次元の正規白色雑音である。式 (6) の  $w_n$  は0となる。

$$x_n = (y_n, y_{n-1}, \dots, y_{n-m+1})^t \quad (7)$$

$$F = \begin{bmatrix} a_1 & a_2 & \dots & a_m \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix}, G = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (8)$$

$$H = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \quad (9)$$

この様な状態空間モデルを各運動パラメータについて求める。

### 4.2 予測

状態空間モデルとして表現したパラメータは、次のような逐次的なアルゴリズムにより、追跡により求めたパラメータの値  $y_1, \dots, y_n$  が与えられた下での状態  $x_n$  の条件付き平均  $x_{n|n}$  を求めることができる。

$$x_{n|n-1} = F_n x_{n-1|n-1} \quad (10)$$

$$V_{n|n-1} = F_n V_{n-1|n-1} F_n^t + G_n Q_n G_n^t \quad (11)$$

$$(12)$$

$$K_n = V_{n|n-1} H_n^t (H_n V_{n|n-1} H_n^t + R_n)^{-1} \quad (13)$$

$$x_{n|n} = x_{n|n-1} + K_n (y_n - H_n x_{n|n-1}) \quad (14)$$

$$V_{n|n} = (I - K_n H_n) V_{n|n-1} \quad (15)$$

さらに、求めた  $x_{n|n}$  を初期値として、時間  $n$  以降の状態ベクトルの予測値  $x_{n+i|n}$  を次のように求めることができる。

$$x_{n+i|n} = F_{n+i} x_{n+i-1|n} \quad (i = 1, 2, \dots) \quad (16)$$

$x_{n+i|n}$  より、各運動パラメータの予測値  $y_{n+i|n}$  は次のように求められる。

$$y_{n+i|n} = H_{n+i} x_{n+i|n} \quad (i = 1, 2, \dots) \quad (17)$$

このようにして、再び追跡できるフレームまでの間、各運動パラメータの値を予測することで、ロボットモデルの位置と姿勢の予測値が求められる。この予測値を元に再び追跡を開始する。

## 5 あとがき

時空間勾配法により人間動作を追跡する際、各運動パラメータの値をARモデルで表わし、カルマンフィルタにより人間の動作を予測する方法を提案した。現在の段階ではそれぞれの運動パラメータについては独立に扱い、個々に予測値を推定している。今後の課題としては、運動パラメータを関連付けて予測することが考えられる。

## 参考文献

- [1] 川田、佐藤、大崎、山本：人間動作のマルチカメラによる追跡と仮想世界での再現、信学技報、RPU95-96,1995
- [2] 亀田、美濃、池田：差分画像を利用した人体の動作認識、信学技報、RPU95-98,1995