カルマンフィルタとファジー推論を用いた単眼視時系列画像からの
3次元人体運動の再構成*天谷 賢治^{*1}, 原 裕二^{*2}, 青木 繁^{*1}Recontstruction of 3D Human Movement from Single View Images
Using Kalman Filter and Fuzzy Reasoning

Kenji AMAYA, Yuji HARA and Shigeru AOKI

A system which reconstructs 3D human movement from a stream of 2D input images is developed. The system finds the 3D shape and motion of the human body by fitting a simple skeleton model to the notable regions found in the 2D image. Surface markers which robust for obstruction were applied. The fitting is done using a nonlinear optimization approach. This optimization is essentially ill-conditioned because depth information can not be obtained from 2D image. In order to overcome this problem, Kalman filter using Fuzzy reasoning which takes account of following a priori information is applied: 1 The versatile skeleton model exhibiting a human body. 2 Postures at each keyframe for the type of human movement. We reconstructed human movement from a stream of Video images by this method in order to demonstrate its applicability.

Key Words: Human Interface, Inverse Problem, Bio-Motion, Motion Capture, A Priori Information, Kalman Filter, Fuzzy Reasoning, Ill-Condition, Single View Image, Keyframe

1. 緒言

近年, 人体のリアリティのある動きをコンピュータに
入力する技術としてモーションキャプチャシステムが
注目されている。(1)-(4) その中でもイメージベースの
システムは, データグローブなどの装備を用いるシステ
ムに比べ, 非接触で計測が行なえるため有用であり, 複
眼視の時系列画像を用いたシステムは実用化されてい
る。しかし, 複眼視の時系列画像を得るのは, 経済的負担
および撮影労力の負担などがかかるため単眼視の時系
列画像を入力とするシステムの開発が望まれている。

単眼視の時系列画像から3次元人体運動を再構成す
るには, 3次元の人体モデルを利用し, 人体モデルの投影
図と入力画像が一致するように人体モデルのパラメー
タを同定する。この同定問題は, 単眼視の時系列画像に
は奥行き方向の情報が含まれていないため, 本質的に悪
条件な問題となる。

そこで, 本研究ではこの悪条件を克服するために先験
情報を利用し, ファジー推論を用いたカルマンフィルタ
を適用することにより, 単眼視の時系列画像から効率的
に3次元人体運動を再構成することを目的とする。

筆者等は, すでに点のマーカを利用した時系列濃淡
画像を入力とするシステムの開発を行った。(5) 点のマー
カーは重なりによる隠れが生じやすいため運動の制約

Fig. 1 Full color images with some surface markers

が大きい。そこで, 今回は, 図1のような面を持つマー
カーを利用したカラー画像を入力とする隠れに強い新
しいシステムを開発した。

2. 人体のモデル化

本問題は2次元画像を入力としているので, 奥行き方
向の情報の欠如は避けられない。本研究では同様にし
て人体の3次元モデルを利用して運動の再構成をする。

実際には, 80自由度以上ある人体の姿勢を22個の関節
角度および骨盤の3次元絶対座標値を成分とする25次
元のパラメータベクトル θ_t としてモデル化した。ここ
で, 画像は離散的に与えられるので, 添字 t は離散的な
時間を表す。図2に人体モデルを示す。各関節は図に示
す自由度を持ち, 身体各部の形状は既知である。

3. 再構成の手順

3.1 モデルフィッティング 次式のように入力画像か
ら抽出した特徴量 $\bar{g}_t$ (3-2節参照) とモデルから求めた

* 原稿受付 1997年6月6日。

*1 正員, 東京工業大学情報環境学専攻(〒152-8552 東京都目黒
区大岡山2-12-1)。

*2 東京工業大学学部生。

Fig. 2 Skeleton model

Fig. 3 The procedure of the reconstruction

特徴量 $g(\theta_t)$ (3-3 節参照) が一致するようにパラメータを同定する。(図 3 参照)

$$\bar{g}_t = g(\theta_t) + v_g \quad (1)$$

ここで、 $\bar{g}_t$ および $g(\theta_t)$ は 20 次元のベクトルである。また、 v_g はガウス性を持つ 20 次元の誤差ベクトル (平均 0, 分散は成分によって異なる) である。 v_g の標準偏差で式 (1) の両辺を除算して正規化をする。

3.2 入力画像の特徴量 $\bar{g}_t$ の抽出 入力画像としては、図 1 のような 5ヶ所の注目領域がそれぞれ単一色を持つカラー画像を用いる。入力画像にはノイズが含まれているので、平滑化を行い人体の 5ヶ所の注目領域 S を抽出する。⁽⁶⁾ 各注目領域 S の 0,1,2 次モーメント m_n を次式によって計算する。

$$m_n = m_{(i,j)} = \int_S x^i y^j dS \quad (n = i + j) \quad (2)$$

1. X coordinates of the center point g : S_X
2. Y coordinates of the center point g : S_Y
3. Slope angle of the major axis : S_θ
4. Aspect ratio (=a/b) : S_λ

Fig. 4 The characteristic values of the notable region

ここで、 i, j は自然数である。(2 次モーメントについては、座標原点を領域 S の図心 (S_X, S_Y) に移して計算する。) A マトリクスを次のように定義する。

$$A = \frac{1}{m_{(0,0)}} \begin{pmatrix} m_{(2,0)} & m_{(1,1)} \\ m_{(1,1)} & m_{(0,2)} \end{pmatrix} \quad (3)$$

A マトリクスの固有値を λ_1, λ_2 ($\lambda_1 > \lambda_2$), λ_2 に対応する固有ベクトルを $a = (a_X, a_Y)$ (領域の長軸方向) とおくと図 4 に示すような注目領域 S の特徴量 $\bar{g}_t$ の各成分は、次式によって求められる。

$$S_X = \frac{m_{(1,0)}}{m_{(0,0)}} \quad (4)$$

$$S_Y = \frac{m_{(0,1)}}{m_{(0,0)}} \quad (5)$$

$$S_\theta = \arctan\left(\frac{a_Y}{a_X}\right) \quad (6)$$

$$S_\lambda = \frac{\lambda_1}{\lambda_2} \quad (7)$$

3.3 モデルの特徴量 $g(\theta_t)$ の計算 モデルの形状は既知であるので、パラメータ θ_t から順運動学を用いて、モデルの各関節部の 3 次元絶対座標値が計算できる。⁽⁷⁾ さらにグラフィックス解析の手法を適用すると、モデルの各関節部のスクリーン上の投影点の 2 次元座標値が計算できる。注目領域 S の各特徴量 (図心、傾き、アスペクト比) は、注目領域 S の両端の関節部の投影点の 2 次元座標値を用いれば、パラメータ θ_t の関数 $g(\theta_t)$ として解析的に表すことができる。 $\frac{\partial g(\theta_t)}{\partial \theta_t}$ は容易に求めることができるので式 (1) を数値的に解く際に用いることができる。

Fig. 5 Normal velocity of each segment

4. 先験情報の利用

式(1)は、奥行き方向の情報を含んでいないので本質的に悪条件である。⁽⁵⁾ 本研究では、あらかじめ対象運動のおおまかな様子が分かっているものとし、それを先験情報として与えることにより悪条件を克服することを考える。先験情報として与える情報とは、キーフレーム(8)における人体モデルのパラメータ値 θ_k である。キーフレームは、図5に示すように、入力画像から抽出した各注目領域 S の図心速度の総和 V_{sum} が極小点になるフレームを選んだ。図心速度の総和 V_{sum} は、次式で表される。

$$V_{sum} = \sum_{i=1}^5 \sqrt{(\dot{S}_{Xti})^2 + (\dot{S}_{Yti})^2} \quad (8)$$

ここで、添字 i は各注目領域を表し、 t は離散的な時間を表す。図5の例では、全体の118フレーム中10フレーム(図中○印)がキーフレームとして選ばれる。

キーフレームにおける人体モデルのパラメータ値 θ_k は、3Dグラフィックスのための人体アニメーションツールであるLife form⁽⁹⁾を用いて作成する。Life formはあるフレームでの人体のポーズを作成した後、別のフレームで別のポーズを作成すると自動的にフレーム間のポーズを補間してアニメーションを作成する。ここで作成されるポーズはあくまでも補助情報として利用するためのポーズであるため、入力画像と完全に一致するようなポーズである必要はなく、一般的な対象運動が表現されていれば十分である。

キーフレームにおける人体モデルのパラメータ値 θ_k を線形補間して全フレームについての先験情報パラメータ値 $\theta_{p,t}$ を得る。 $\theta_{p,t}$ を用いれば、全フレームにおける人体モデルの主要箇所13点(図6参照)の x 座標(奥行き方向) $\bar{d}_t$ を前もって計算できる。従って、これら13点の x 座標をパラメータ θ_t の関数 $d(\theta_t)$ として表せ

Fig. 6 Principal points of the skeleton model

ば、 θ_t は、次式を満たす。

$$\bar{d}_t = d(\theta_t) + v_d \quad (9)$$

ここで、 v_d は全時間に均一なガウス性を持つ誤差(平均0, 分散 $10^5[mm^2]$)である。実際には $\theta_{p,t}$ の誤差は図7に示すように時間 t がキーフレームから離れるほど大きくなる。これを考慮するために、次に説明するカルマンフィルタの外乱ベクトルを求める際に再び $\theta_{p,t}$ を利用する。

v_d の標準偏差で式(9)の両辺を除算して正規化をする。式(1)および(9)を以下に説明するカルマンフィルタの観測方程式とする。

5. ファジー推論を用いたカルマンフィルタの適用

本研究で扱うカルマンフィルタは、以下のような状態方程式(式(10))と観測方程式(式(11))によって構成される。

$$\theta_{t+1} = \theta_t + w_t \quad (t = 0, 1, \dots) \quad (10)$$

$$z_t = h(\theta_t) + v_t \quad (t = 1, 2, \dots) \quad (11)$$

$$z_t = \begin{bmatrix} \bar{g}_t \\ \bar{d}_t \end{bmatrix}, h(\theta_t) = \begin{bmatrix} g(\theta_t) \\ d(\theta_t) \end{bmatrix}, v_t = \begin{bmatrix} v_g \\ v_d \end{bmatrix} \quad (12)$$

ここで、添字 t は離散的な時間を表し、 w_t は25次元の外乱ベクトルを表す。

カルマンフィルタのアルゴリズムに従えば、次式を用いて反復計算を行うことにより、 θ_t の推定値 $\hat{\theta}_t$ が得られる。⁽¹⁰⁾

$$\bar{\theta}_{t+1} = \bar{\theta}_t + \bar{w}_t \quad (13)$$

$$\hat{\theta}_t = \bar{\theta}_t + K_t(z_t - h(\bar{\theta}_t)) \quad (14)$$

ここで、 $\bar{\theta}_t$ は状態方程式による推定値、 w_t は外乱ベクトルの平均値、 K_t はカルマンゲインを表す。

本研究では、 w_t の平均と分散を先験情報パラメータ $\theta_{p,t}$ の誤差を考慮したファジー推論を用いて与える。先

Fig. 7 Error of $\theta_{p,t}$

験情報パラメータ値 $\theta_{p,t}$ の誤差は時間 t がキーフレームから離れるほど大きくなる。一方、パラメータ $\hat{\theta}_t$ は次のキーフレームにおいて再び $\theta_{p,t}$ に近づくのだから、キーフレーム間の $\theta_{p,t}$ の平均速度に近い速度で変化している。従って外乱ベクトルの平均値 $\bar{w}_t$ は次のように与えられる。

If "このフレームはキーフレームから近い" then

"パラメータ $\hat{\theta}_{t+1}$ の値は先験情報パラメータ $\theta_{p,t+1}$ の値に近い" ($\bar{w}_t \simeq \theta_{p,t+1} - \hat{\theta}_t$)

If "このフレームはキーフレームから遠い" then

"パラメータ $\hat{\theta}_t$ の傾きは先験情報パラメータ $\theta_{p,t}$ の傾きに近い" ($\bar{w}_t \simeq \theta_{p,t+1} - \theta_{p,t}$)

これらの情報に対するメンバーシップ関数を図8に示す。⁽¹¹⁾ 2つの後件部のメンバーシップ関数の論理和から得られたメンバーシップ関数 f_m の平均と分散を w_t の平均と分散として与える。本研究では、メンバーシップ関数 f_m をガウス分布であると仮定し、以下のようにカルマンゲイン K_t を求める。

$$K_t = P_t H_t^T R_t^{-1} \quad (15)$$

$$P_t = (M_t^{-1} + H_t^T R_t^{-1} H_t)^{-1} \quad (16)$$

$$= M_t - M_t H_t^T (H_t M_t H_t^T + R_t)^{-1} H_t M_t$$

$$M_{t+1} = P_t + Q_t \quad (17)$$

$$H_t = \frac{\partial h(\theta_t)}{\partial \theta_t} \quad (18)$$

ここで、 Q_t, R_t はそれぞれ w_t, v_t の共分散行列である。

6. 再構成例

本研究で用いた入力画像の例を図9に示す。右下の数字は、フレーム番号を表す。装着した5種類のカラータイツを注目領域とした。ラジオ体操をしている人体を実際に撮影した。撮影はデジタルビデオカメラにて行い、118枚の総フレームを320×240pixelの解像度およ

Fig. 8 Fuzzy reasoning

び30frames/secのフレームレートで計算機に入力した。図5に基づいてキーフレームを選択し、図8に示すメンバーシップ関数を作成した。

再構成結果を図11および図12に示す。参考のために斜め方向から見た実際のビデオ画像を図10に示す。なお、斜め方向のビデオ画像の情報は再構成に一切用いていないことを注記する。正面から見た投影図も斜めから見た投影図も良く再構成されていることが分かる。右ひじ関節における先験情報パラメータ値 $\theta_{p,t}$ と推定値との差を図13に示す。再構成された姿勢は、先験情報と画像情報の両方を含んでいるので、先験情報として与えた姿勢よりも入力画像に近い姿勢になっている。(図14参照)

7. 結言

本研究では面積を持つマーカーを利用した時系列カラー画像を入力とすることにより、隠れに強いモーションキャプチャシステムを開発した。本問題における奥行き情報の欠如からくる悪条件性をキーフレームにおけるモデルパラメータ値を先験情報として与え、ファジー推論を用いたカルマンフィルタを適用することにより克服した。以下にその手順を示す。

1. 入力画像に画像処理を施した後、人体の注目領域を抽出し、特徴量を求めた。
2. 各注目領域の図心速度の総和が極小点になるフレームをキーフレームとして選択し、メンバーシップ関

Fig. 9 Input images

Fig. 11 Front view of the reconstructed motion

Fig. 10 Reference images

Fig. 12 Side view of the reconstructed motion

数を作成した。

3. 人体モデルによって作成されたキーフレームにおけるパラメータ値を線形補間して全フレームにおける先験情報パラメータ値を求めた。
4. 先験情報パラメータ値を用いて人体の主要箇所13点の奥行き方向の座標を計算した。
5. ファジー推論を用いたカルマンフィルタを適用し、人体モデルのパラメータを推定した。

実際にビデオ画像から再構成を行い本手法の有効性を示した。なお、本研究は文部省科学研究費奨励研究Aの補助を受けて行なった。

Fig. 13 difference between $\theta_{p,t}$ and $\hat{\theta}_t$

Fig. 14 Each pose in the 75th frame

参考文献

- (1) Blake, Isard, *Computer graphics proceedings, annual conference series*, (1994)185-192.
- (2) Yamato, Ohya, Ishii, *IEEE Trans. on pattern analysis and machine intelligence*, (1992)379-385.
- (3) Wise, Gardner, Sabelman, Valainis, *Journal of Rehabilitation Research and Development*, 27-4(1990)411-424.
- (4) Long, Yang, *International Journal of Pattern Recognition and Artificial Intelligence*, 5-3(1991)439-458.
- (5) 天谷, 原, 青木, 機論 63-608, C(1997), 1167-1171.
- (6) 柴山, X11による画像処理基礎プログラミング (1994) 技術評論社.
- (7) 広瀬, ロボット工学 (1987) 裳華房.
- (8) 末永, 上原, 顧, 角所, 馬場口, 北橋, 信学技報 PRU94-113, (1995-01), 47-54.
- (9) Calvert, Bruderlin, Dill, Schiphorst, Welman, *IEEE Computer Graphics and Applications*, May (1993).
- (10) 加藤, 最適制御入門 (1987) 東京大学出版会.
- (11) 寺野, 浅居, 菅野, 応用ファジィシステム入門 (1989) オーム社.