

論 文

キーフレーム拘束を利用した単眼動画像からの人間動作の追跡

大田 佳人^{†*} 山際 貴志^{†**} 山本 正信[†]

Keyframe Tracking of Human Body in 3D Motion from a Monocular Image Sequence

Yoshito OHTA^{†*}, Takashi YAMAGIWA^{†**}, and Masanobu YAMAMOTO[†]

あらまし 動画像から人体を3次元的に追跡する手法は大きく分けて、フレームごとにモデルを人体に照合させる方法と、初期フレームで照合させた後、変位分を推定し累積する手法に大別される。後者は、計算量が少なくすむが、誤差も累積され長時間の追跡で外れが問題となっている。本論文では、画像とモデルとの照合を最終フレームでも行い、この照合により与えられた位置と姿勢の拘束を他のフレームに伝搬させることによりずれの解消を行う。この伝搬は、隠れが起こったとしてもその間の動きの推定を可能とする。但し、弛緩法による拘束の伝搬を行っているため、動きの初期推定値を必要とする。隠れが起こった区間の初期推定値は、隠れが起こっていないフレーム区間から予測により推定する。この予測にカルマンフィルタを用いる。モデルと人体を照合をするフレームをキーフレームとし、この追跡方法をキーフレーム追跡法と呼ぶことにする。

キーワード キーフレーム、時空間こう配法、滑らか拘束、カルマンフィルタ、ARモデル、弛緩法

1. まえがき

人間動作の動画像追跡は、コンピュータアニメーションやテレビゲーム等でのキャラクタの動作の生成において重要な技術となってきた。

動画像追跡の手法では、人体のモデルを画像中の人体像に照合させるのが一般的である。これまで提案されてきた手法の多くは、モデルと人間との照合をフレームごとに行っている[1],[2],[5]~[7],[10],[15],[16],[18]。この方法の問題点として、画像中から人体を検出することの困難さや、探索過程を含む照合に膨大な計算量を必要とすること等が挙げられている。

これに対して、時間のかかるモデルと人体像との照合を追跡開始フレームだけで行い、位置・姿勢の変位分を推定し累積することにより追跡を行う手法が提案されている[4],[9],[11],[17],[19]。特に、時空間こう配法を用いれば位置と姿勢の変位(運動パラメータ)を直接的に推定することができる[17],[19]。得られた変位に基づきモデルを移動させれば、次のフレームで

もモデルは人体と一致している。変位の推定とモデルの移動を繰り返しながら追跡を行う。しかしながら、運動パラメータの推定誤差も累積されるため、長時間の追跡ではモデルと人体とのずれが拡大することが問題とされてきた。

本論文では、画像とモデルとの照合を最終フレームでも行い、この照合により与えられた位置と姿勢の拘束を他のフレームに伝搬させることによりずれの解消を行う。この伝搬は、隠れが起こったとしてもその間の動きの推定を可能とする。但し、弛緩法による拘束の伝搬を行っているため、動きの初期推定値を必要とする。隠れが起こった区間の初期推定値は、隠れが起こっていないフレーム区間から予測により推定する。この予測にカルマンフィルタを用いる。モデルと人体を照合をするフレームをキーフレームとし、この追跡方法をキーフレーム追跡法と呼ぶことにする。

次章では、準備として、運動している人体とそれを観測しているカメラ系についてモデル化する。3.では、運動パラメータを決定するための拘束条件を導入する。4.では拘束条件を満たす解の計算法について述べる。5.は隠れが起こったときの追跡法、6.では、複雑で早い動作の追跡結果を示す。

[†] 新潟大学工学部情報工学科, 新潟市
Niigata University, 8050, Ikarashi2-nochou, Niigata-shi, 950-2181
Japan

* 現在, 株式会社ナムコ

** 現在, 東芝マイクロエレクトロニクス株式会社

2. モデル

人間・カメラ系のモデルを導入する。

2.1 カメラモデル

人間の動きを観測するカメラシステムは、シーンに固定されているとする。シーンにはワールド座標系 Σ_w が付加されている。カメラのモデルは中心投影モデルとする。すなわち、カメラ座標系を直交座標系 (x, y, z) とし、投影面を z 軸に垂直な平面 $z = f$ とする。 f はカメラの焦点距離であるが、ここでは簡単のため $f = 1$ とする。投影面座標系 (X, Y) は z 軸と投影面の交点を原点とし、座標軸 X, Y をそれぞれ x, y 軸に平行になるように設定する。このとき、人体上の点 (x, y, z) の画面上への投影点 (X, Y) は次式で与えられる。

$$X = \frac{x}{z}, \quad Y = \frac{y}{z} \quad (1)$$

人体上の点の速度ベクトル (v_x, v_y, v_z) とその投影面上での速度ベクトル (V_X, V_Y) は、式 (1) を時間微分することにより次のように得られる。

$$\begin{cases} V_X = \frac{1}{z}(v_x - Xv_z) \\ V_Y = \frac{1}{z}(v_y - Yv_z) \end{cases} \quad (2)$$

2.2 人体モデル

人体を多関節構造でモデル化する。人間のパーツ(胴体, 頭部, 腰部, 左右上腕, 下腕, 上脚, 下脚)を多面体で近似したモデルを, CAD モデラで作成する。それぞれのパーツを, 上腕の親は胴体, 下腕の親は上腕といった親子関係で結ぶ。このとき, 胴体を階層構造の頂点に置き, 番号 $i = 1$ で表す。親が動けばそれに伴って子も動く。各パーツには物体固有の座標系が付加されている。腕, 足などのパーツは, この座標系の原点を人体の関節にあたる位置におく。第 j フレームにおいて, 物体 i の座標系 $\sum_{i,j}$ で表された位置 $p_{i,j}$ は, ワールド座標系 \sum_w では次式により, 位置 $p_{w,j}$ として表される。

$$p_{w,j} = R_{i,j}p_{i,j} + T_{i,j} \quad (3)$$

ここで, $R_{i,j}$ はワールド座標系 \sum_w に対する, 座標系 $\sum_{i,j}$ の回転を表す変換行列, $T_{i,j}$ は, 原点の位置を表す変換ベクトルである。

2.3 キーフレーム拘束

キーフレームでモデルを人体像に照合させることにより, 人体の位置と姿勢を知ることができる。得られた姿勢, 位置は, 初期フレームでは, $R_{i,0}, T_{i,0}$, 最終フレームでは, $R_{i,n}, T_{i,n}$ とする。

モデルと人体像との照合法 [1],[5]~[7],[10],[12],[15],[16],[18] は種々提案されているが, そのほとんどが多視点画像からの照合である。ここでは, 手動による照合を行った [19]。この照合は, 単眼画像に対する照合であるため正確な 3次元照合は難しいが, 正しい位置・姿勢から大きく外れることはない [19]。

モデルを人体像に一致させたとき, カメラから観測したモデルより, 人体の各パーツを表すラベル画像と距離画像を作成することができる [17]。ラベル画像は, 3.1 での時空間拘束式の導入, 5.3 での隠れの検知に用いる。距離画像は, 時空間拘束式 (9) の 3次元速度ベクトルに関する線形化に用いる。

2.4 運動モデル

物体 i の $j-1$ フレームから j フレームへの運動を, 物体座標系 $\sum_{i,j}$ 自身の原点回りの回転を表す行列 $Q_{i,j}$ と, ワールド座標系 \sum_w での平行移動を表すベクトル $S_{i,j} = (\alpha_{i,j}, \beta_{i,j}, \gamma_{i,j})^T$ によって表すことにする。回転行列 $Q_{i,j}$ は, x, y, z 軸回りの回転量 $\phi_{i,j}, \psi_{i,j}, \phi_{i,j}$ が微小ならば, 次のように近似できる。

$$Q_{i,j} = \begin{pmatrix} 1 & -\phi_{i,j} & \psi_{i,j} \\ \phi_{i,j} & 1 & -\phi_{i,j} \\ -\psi_{i,j} & \phi_{i,j} & 1 \end{pmatrix} \quad (4)$$

我々の人体モデルでは, 胴体 ($i = 1$) の動きのみ並進と回転の 6 自由度を有するが, 他の部分 (腕や足) の動きは, 関節回りの回転のみとする。物体 i の $j-1$ フレームから j フレームへの位置, 姿勢の変位をベクトル x_j として, それぞれの物体の回転, 並進量により,

$$x_j = (x_{(1,j)}, x_{(2,j)}, \dots, x_{(m,j)})^T \quad (5)$$

とする。但し,

$$x_{(i,j)} = \begin{cases} (\alpha_{i,j}, \beta_{i,j}, \gamma_{i,j}, \phi_{i,j}, \psi_{i,j}, \phi_{i,j})^T & i = 1 \\ (\phi_{i,j}, \psi_{i,j}, \phi_{i,j})^T & i \neq 1 \end{cases}$$

これを人体の運動を表す運動パラメータとする。

第 a フレームでの物体座標系 $\sum_{i,a}$ から b フレームでの物体座標系 $\sum_{i,b}$ への姿勢の変換行列 $R_{i,a \rightarrow b}$

は、次式で表される。

$$R_{i,a \rightarrow b} = \prod_{l=a+1}^b Q_{i,l} \quad (6)$$

また、物体座標系 $\sum_{i,a}$ の原点から b フレームでの物体座標系 $\sum_{i,b}$ の原点へのワールド座標系 \sum_w 上での移動量 $T_{i,a \rightarrow b}$ は

$$T_{i,a \rightarrow b} = \sum_{l=a+1}^b S_{i,l} \quad (7)$$

と表すことができる。

3. 動きの拘束条件

本論文の目標は人体の運動パラメータ x_j を観測した動画像から精度良く推定することである。そのために、本章では運動パラメータの拘束条件として、時空間拘束 [19]、終端拘束、滑らか拘束の三つを導入する。

3.1 時空間拘束

投影面上での位置 (X, Y) 、時刻 t フレームでの画像の明るさを $E(X, Y, t)$ とする。このとき画面上の速度ベクトル (V_X, V_Y) は、近似的に次の 1 次式に拘束されることが知られている [3]。

$$E_X V_X + E_Y V_Y + E_t = 0 \quad (8)$$

ここで、 (E_X, E_Y) は画面上での空間こう配、 E_t は時間こう配である。

式 (2) の (V_X, V_Y) を式 (8) に代入すれば、人体上の 3 次元速度ベクトル (v_x, v_y, v_z) の拘束式が得られる [19]。

$$\frac{E_X}{z} v_x + \frac{E_Y}{z} v_y - \frac{1}{z} (X E_X + Y E_Y) v_z + E_t = 0 \quad (9)$$

画像上の点 (X, Y) が属している人体上のパーツとその 3 次元位置は、2.3 で与えられたラベル画像と距離画像から知ることができる。そのとき、人体は多関節構造でモデル化されているので、人体上の位置 $p = (x, y, z)^T$ は、関節角などをパラメータ $q = (q_1, q_2, \dots, q_n)^T$ とした非線型関数として次式のように表現できる [13]。

$$p = f(q) \quad (10)$$

更に、人体上の速度ベクトル $\dot{p} = (v_x, v_y, v_z)^T$ と人体の姿勢を表すパラメータの時間変化率 \dot{q} は、ヤコビ

行列 $J = \partial f / \partial q$ によって、次の線形式で関係づけられる。

$$\dot{p} = J \dot{q} \quad (11)$$

この式を式 (9) に代入することによって、姿勢を表すパラメータの時間変化率 \dot{q} の拘束式が得られる。カメラから人体までの距離 z は、距離画像から与えられているのでこの拘束式は線形拘束式である。ヤコビ行列の具体的な計算法については論文 [19] を参照されたい。

動きが早くないとき、姿勢を表すパラメータの時間変化率は、運動パラメータ x_j で近似することができる (時間間隔を単位時間とおく)。このとき、運動パラメータの拘束式を係数行列 A_j 、定数ベクトル b_j として、次の線形方程式で表すことにする。

$$A_j x_j = b_j \quad (12)$$

3.2 終端拘束

初期フレームから最終フレームまでの運動パラメータ $x_j (j = 1, \dots, n)$ を使って、最終フレームでの位置と姿勢を計算することができる。最終フレームでの位置と姿勢を求めたとき、これが、与えられた最終フレームでの位置 $T_{i,n}$ と姿勢 $R_{i,n}$ に一致しなくてはならない。これを終端拘束と呼ぶ。すなわち、運動パラメータ $x_j (j = 1, \dots, n)$ の累積により得られた最終フレームでの物体 i の位置を $\hat{T}_{i,n}$ 、姿勢を $\hat{R}_{i,n}$ とすれば、

$$\hat{R}_{i,n} = R_{i,n} \quad (13)$$

$$\hat{T}_{i,n} = T_{i,n} \quad (14)$$

でなければならない。

胴体以外のパーツは、その親と関節でつながっている。従って直接の親の座標系を基準とすれば、そのパーツの動きは姿勢のみで表すことができる。すなわち、パーツ i の親のパーツを k とすると、親を基準としたときの初期フレーム 0 での姿勢 $R'_{i,0}$ は、

$$R'_{i,0} = R_{k,0}^{-1} R_{i,0} \quad (15)$$

であるので、式 (6) と式 (15) より、親を基準にした自身の最終姿勢 $\hat{R}'_{i,n}$ は、

$$\hat{R}'_{i,n} = R_{k,0}^{-1} R_{i,0} \prod_{j=1}^n Q_{i,j} \quad (16)$$

として計算される。また、最終フレームで与えられる親と自身の姿勢 $R_{k,n}$ と $R_{i,n}$ から、親を基準にした自身の最終姿勢 $R'_{i,n}$ は、次式のように表される。

$$R'_{i,n} = R_{k,n}^{-1} R_{i,n} \quad (17)$$

このとき、式(13)を次のように書き換えることができる。

$$\hat{R}'_{i,n} = R'_{i,n} \quad (18)$$

胴体は親がないので、形式的に $R_{k,0}$, $R_{k,n}$ は単位行列とする。

位置の終端拘束(14)は、胴体についてのみ考えればよい。胴体はすべての親であるので、式(7)より $\hat{T}_{1,n}$ は、

$$\hat{T}_{1,n} = \sum_{j=1}^n S_{1,j} \quad (19)$$

となる。

3.3 滑らか拘束

フレーム間で推定された運動パラメータは、運動が急変することがなければ、時間と共に滑らかに変化する。この条件は次のように表すことができる。

$$\mathbf{x}_j \cong \mathbf{x}_{j-1} \quad (20)$$

4. 運動パラメータの逐次推定

時空間拘束(12)、終端拘束(14)、(18)、滑らか拘束(20)を同時に満たす運動パラメータは、一般に各拘束式にノイズが含まれているので得ることができない。そこで、各拘束の左辺と右辺の差のノルムから得られる、次の評価式を最小とする解を求めることにする。

$$\begin{aligned} L = & \sum_{j=1}^n \|A_j \mathbf{x}_j - \mathbf{b}_j\|^2 \\ & + \frac{\lambda}{2} \sum_{i=1}^m \|\hat{R}'_{i,n} - R'_{i,n}\|^2 \\ & + v \|\hat{T}_{1,n} - T_{1,n}\|^2 \\ & + \sum_{j=2}^n (\mathbf{x}_j - \mathbf{x}_{j-1})^\top \text{diag} \left(\frac{\nu}{2} \right) (\mathbf{x}_j - \mathbf{x}_{j-1}) \end{aligned} \quad (21)$$

ここで、スカラー λ , v , およびベクトル ν の各要素は、各拘束にかかる重みであり、負でない実数とする。また、 m は人体モデルを構成するパーツの個数とする。

4.1 弛緩法

評価式(21)は運動パラメータに関する非線型関数である。最小化には、適当な初期値からの逐次最適化を行う[14]。時空間拘束では係数行列 A_j , \mathbf{b}_j は、今のフレーム以前の運動パラメータが得られなければ、モデルの位置を知ることができず、値を確定することができない。従って、まず、時空間拘束のみで、最終フレームまでの運動パラメータを求め、これを初期解とする。この後、評価式(21)が最小となるように解を逐次修正していく。

この修正はニュートン法による。すなわち、評価式(21)を、運動パラメータで $\mathbf{x}_j (j=1, \dots, n)$ で微分し、 $\mathbf{0}$ と等しくした連立方程式を解くことにより解の修正量を得る。しかし、この連立方程式の係数行列は非常に大きなものとなる。そこで、初期フレームからフレーム順に解を求めることにする。

評価式(21)を運動パラメータ \mathbf{x}_j で微分し結果を $\mathbf{0}$ とおく。

$$\frac{\partial L}{\partial \mathbf{x}_j} = \mathbf{0} \quad (22)$$

この方程式の中で、時刻 j 以外の運動パラメータ値に初期解を代入する。このとき、式(22)は、次の線形推定式となる。

$$\begin{aligned} (A_{t,j} \mathbf{x}_j - \mathbf{b}_{t,j}) + \lambda (E_{r,j} \mathbf{x}_j - \mathbf{b}_{er,j}) \\ + v (E_{t,j} \mathbf{x}_j - \mathbf{b}_{et,j}) + (A_{s,j} \mathbf{x}_j - \mathbf{b}_{s,j}) = \mathbf{0} \end{aligned} \quad (23)$$

但し、

$$\begin{aligned} A_{t,j} &= A_j^\top A_j \\ \mathbf{b}_{t,j} &= A_j^\top \mathbf{b}_j \end{aligned}$$

であり、係数行列 $E_{r,j}$, $E_{t,j}$, $A_{s,j}$ は、それぞれ付録の式(A.1), (A.3), (A.5), および定数ベクトル $\mathbf{b}_{er,j}$, $\mathbf{b}_{et,j}$, $\mathbf{b}_{s,j}$ は、式(A.2), (A.4), (A.6)で与えられている。

この推定式(23)の未知数は、 \mathbf{x}_j のみである。この推定式を \mathbf{x}_j について、初期フレームから最終フレームまで逐次解いた結果、近似解 $\mathbf{x}^{(1)}$ を得る。初期解をこの近似解により更新し、再度同様な手続きを行い解 $\mathbf{x}^{(2)}$ を得る。この手続きを繰り返すことにより正しい解に収束させる。

4.2 弛緩法での問題

4.1で述べた方法では、それぞれの拘束にかかる重みの大きさが解に影響を及ぼす。例えば、終端拘束に

かかる重み λ , ν が大きくなると、式 (23) の第 2 項、第 3 項の終端拘束のみを解いた解に近くなる。そのとき、ある一つのフレーム間の動きだけで終端条件を満たそうとする。その結果、前回の解よりも新しい解の方がキーフレーム拘束から大きくずれ、解の振動や発散を起こす場合もある。

一方、この重みが小さければ、キーフレーム姿勢に収束する速度が遅くなる。従って、重みの決定が重要な問題となる。しかし、実際の追跡において、終端拘束の重みを最適にするのは難しい問題である。ここでは、収束を速くするために終端拘束の重みを大きくしても、解が振動や発散を起こさないように推定式 (23) を修正することにする。

4.3 推定式の修正

終端拘束の重み λ , ν を大きくすると、推定式 (23) は終端拘束のみから得られる次の推定式で近似することができる。

$$\lambda(E_{r,j}x_j - b_{er,j}) + \nu(E_{t,j}x_j - b_{et,j}) = 0 \quad (24)$$

この推定式の解は容易に得られる。なぜならば、各パーツ間の運動パラメータは互いに独立であり、線形推定式の係数行列が単位行列であるからである。解は、

$$x_{ej} = b_{er,j} + b_{et,j} \quad (25)$$

となる。

問題は、この解が全フレーム分の誤差をたった 1 フレーム間で修正できることにある。そこで、解が誤差の一部だけを修正し、前回の解から大きく離れないようにする。すなわち、終端拘束のみで得られる解 x_{ej} と前回の解 x_{oj} との差を

$$\mathbf{Error}_j = x_{ej} - x_{oj} \quad (26)$$

とする。この \mathbf{Error}_j の一部を前回の推定量 x_{oj} に加えた、

$$x'_{ej} = u\mathbf{Error}_j + x_{oj} \quad (27)$$

$$0 \leq u \leq 1$$

が終端拘束のみを使った解になるように、式 (24) の定数ベクトル $b_{er,j}$, $b_{et,j}$ を変更する。ここで、 u は x_{ej} と x_{oj} との間のバランスをとる重みである。この u を、例えば

$$u = \frac{1}{\text{全フレーム数}} \quad (28)$$

のようにすれば、誤差の修正を各フレームで等分に負担するようにできる。

推定式 (24) の変更は次のように行う。式 (27) で得られた x'_{ej} のうち、胴体の並進運動に対応する要素を $b'_{et,j}$ とし、各部の回転運動に対応する要素を $b'_{er,j}$ とする。式 (24) の $b_{et,j}$ を $b'_{et,j}$ に $b_{er,j}$ を $b'_{er,j}$ に置き換える。

$$\lambda(E_{r,j}x_j - b'_{er,j}) + \nu(E_{t,j}x_j - b'_{et,j}) = 0 \quad (29)$$

この推定式の解は、当然 x'_{ej} である。

本来の推定式 (23) の第 2 項と第 3 項を推定式 (29) の第 1 項と第 2 項に置き換えれば、次の推定式が得られる。

$$(A_{t,j}x_j - b_{t,j}) + \lambda(E_{r,j}x_j - b'_{er,j}) + \nu(E_{t,j}x_j - b'_{et,j}) + (A_{s,j}x_j - b_{s,j}) = 0 \quad (30)$$

この新たな推定式を使えば、正しい解に素早く収束させることができる。但し、重み λ , ν が大きくなると、終端拘束のみからの解 x'_{ej} から正しい解への収束が遅くなる。適度な重みの大きさは実験的に求めた。本論文では、 $\lambda = 10.0$, $\nu = 20.0$, ν の要素値を 1.0 とした。

5. 隠れ

これまでは、隠れが起こらないとして議論を進めてきたが、本章では、隠れが起こった場合を問題とする。

5.1 予測

隠れが途中から起こったとすれば、それまで観測した運動パラメータを使って、隠れているときの運動を推定することができる。推定する運動パラメータを過去のパラメータの値で表現できるとし、AR モデル (autoregressive model) でモデル化する。すなわち、時刻 j での、ある一つの運動パラメータの値 y_j は、過去 h フレームのパラメータ値の線形和で表せるとする。

$$y_j = \sum_{i=1}^h a_i y_{j-i} + v_j \quad (31)$$

ここで、 a_i は自己回帰係数である。また、 v_j は平均 0、分散 σ_v の正規分布に従う白色雑音とする。隠されるまでに追跡することができたパーツの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

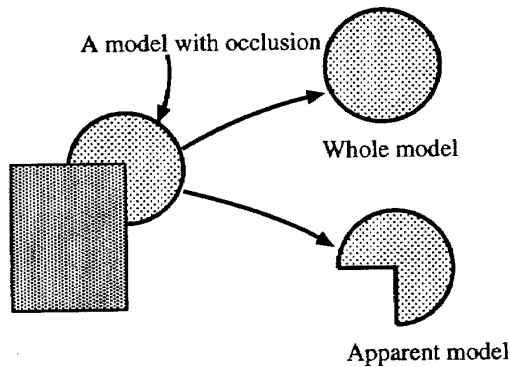


図1 隠れの判別
Fig.1 Detection of occlusion.

モデルの自己回帰係数はレビンソン法[8]によって推定する。

5.2 状態空間モデル

カルマンフィルタを用いるために、運動パラメータ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に定める。

$$s_j = F s_{j-1} + G v_j \quad (32)$$

$$y_j = H s_j + w_j \quad (33)$$

ここで、状態ベクトル s_j 、および行列 F 、 G 、 H は、次のように定義される。

$$s_j = (y_j, y_{j-1}, \dots, y_{j-h+1})^T \quad (34)$$

$$F = \begin{pmatrix} a_1 & a_2 & \dots & a_h \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{pmatrix}, \quad (35)$$

$$G = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^T \quad (36)$$

$$H = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \quad (37)$$

w_j は平均 0、分散 σ_w の正規性白色雑音である。

このような状態空間モデルを各運動パラメータについて当てはめる。状態空間モデルとして表現したパラメータは、カルマンフィルタにより予測値を推定することができる。カルマンフィルタによる予測の概略を付録に示す。このようにして、再び物体が現れ追跡可能となるまでの間、モデルの位置と姿勢を推測する。

5.3 隠れの検知

モデルが画像中の人体と一致しているとすれば、隠れはモデルの投影像を利用して検知することができる。

すなわち、CAD モデラは人体モデルの各パーツにあらかじめラベルを割り振っている。人体モデルを画面上に投影したとき、投影領域の画素にそのラベルを記入しラベル画像[17]を得る。パーツの見える部分の面積は領域内のラベル(画素)の総数として得られる。また、モデルのパーツごとの投影から各パーツ全体の投影像面積も求めることができる。

図1に示されるように、人体の各パーツモデルが部分的に隠されているとき、見える部分の面積に対して、そのパーツモデルの全面積との比があるしきい値より小さくなったとき、隠れが起こったとみなすことにする。

また、今まで見えていたパーツモデルが、次のフレームでは見えなくなる場合も考えられる。このような場合、1フレーム先の予測により移動させたモデルを用いて見え隠れの判別を行う。現在のモデル、予測したモデルのどちらかで、隠れの条件を満たせば、隠れの発生とすることにした。

5.4 隠れを考慮したキーフレーム追跡

時空間拘束のみで運動パラメータの初期解を求める際に、隠れの発生が検知されたならば、その隠れが再び解消されるまで運動パラメータの予測を行う。

このようにして初期解が得られたならば、2回目以降の繰返し演算では、隠れの区間中に隠されたパーツについては、評価式(21)から時空間拘束を除いた評価式

$$\begin{aligned} L' = & \frac{\lambda}{2} \sum_{i=1}^m \|\hat{R}'_{i,n} - R'_{i,n}\|^2 + \nu \|\hat{T}_{1,n} - T_{1,n}\|^2 \\ & + \sum_{j=2}^n (\mathbf{x}_j - \mathbf{x}_{j-1})^T \text{diag} \left(\frac{\nu}{2} \right) (\mathbf{x}_j - \mathbf{x}_{j-1}) \end{aligned} \quad (38)$$

を最適化することにする。全アルゴリズムを図2に示す。

5.5 実験

提案した追跡アルゴリズムの確認のため、歩行動作の追跡を行った。この動作は、カメラの前を右から左に約4歩弱で横切る歩行である。全フレーム数は97フレームである。人体の右足と右手はしばしば胴体に隠される。この実験では、そのうちの右足に着目し追跡を行った。

まず、隠れが起こったとき、予測は行わぬがキーフレーム拘束を用いずに追跡を行った。その結果、85フレ

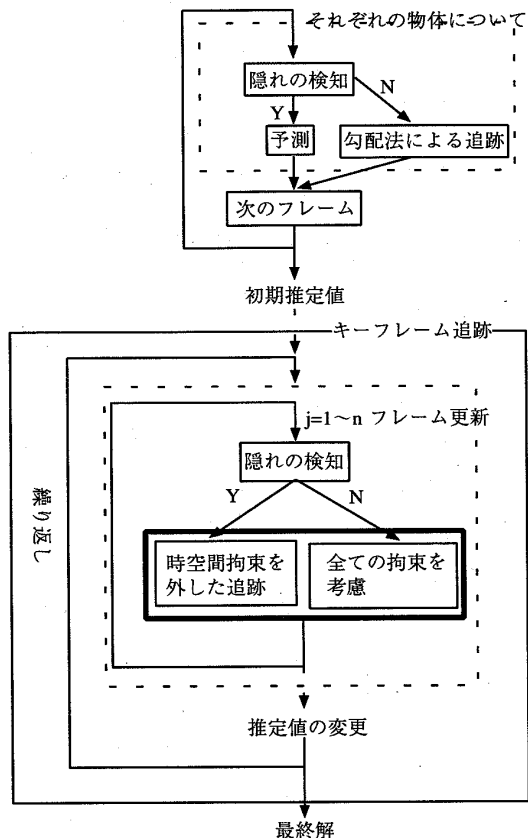


図2 キーフレーム追跡アルゴリズム
Fig.2 Keyframe tracking algorithm.

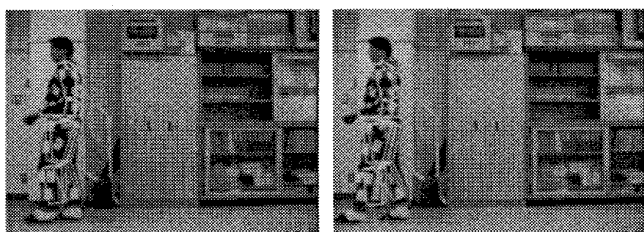


図3 キーフレームを用いない追跡結果. 90 (左), 96 (右) フレーム目
Fig.3 Tracking results without keyframe constraints. Left: 90th frame, and Right: 96th frame.

ムごろから次第にモデルと人体とのずれが大きくなった。図3には、90 (左) と 96 (右) フレームでのずれ具合が、モデルの輪郭を画像に白線で重ねることで表示されている。次に、キーフレーム拘束を考慮した追跡結果を図4に示す。度重なる隠れにもかかわらず、右足は追跡できている。このときの弛緩法の反復回数は30回であった。

6. 区間を分割した追跡

非線型評価関数 (21) の最適化では、初期解が正しい解から大きく外れていると局所解に陥る可能性があ

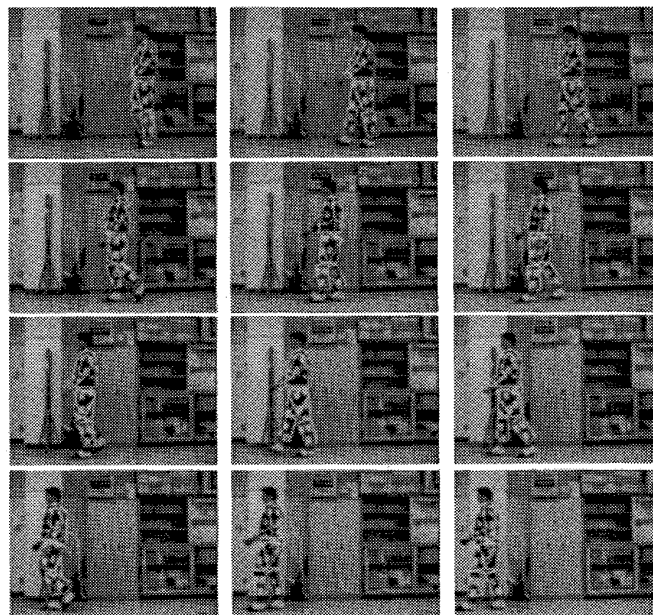


図4 歩行動作のキーフレーム追跡. 追跡結果を8フレームおきに表示

Fig.4 Keyframe tracking of pedestrian. The results are shown at intervals of 8 frames.

る。この原因として、画質が悪く動きが3次元的で速い、隠れが頻繁に起こる、等の場合予測が正確でなく、結果として初期解が正しい解から大きく外れることがある。

このような場合、初期フレームと最終フレーム以外にも、間にいくつかのキーフレームを与える必要がある。新しいキーフレームの候補として、できるだけ隠れが起こらないフレームがモデルと人体の照合には望ましい。従って、隠れが生じたフレームの前後に新しいキーフレームを加えることにした。分割された、それぞれの区間について、提案したキーフレーム追跡を行う。

6.1 追跡実験

円盤投げ競技の投てきシーン^(注1)の動画追跡を試みた。円盤投げは、高速の回転動作であるため、各フレームをフィールド分割し、全部で72フィールドを使用した。それでも、腕や足の隠れが頻繁に起こるため、最初と最後二つのキーフレームだけでは追跡が難しい。

0 (初期フレーム), 71 (最終フレーム) 以外に新たに加えたキーフレームは、15, 26, 41, 52, 61, 66, である。反復演算を1000回行ったときの追跡結果を図5に示す。弛緩法での反復計算につれて、モデルが人体

(注1) : 1996年のアトランタ・オリンピックにおいて、女子円盤投げの優勝者 (I. Wyludda) の優勝投てきシーンの映像。



図5 円盤投げの投てき動作のキーフレーム追跡。追跡結果を6フレームおきに表示

Fig.5 Keyframe tracking of a discus-thrower, I. Wyludda, who is a gold medalist of the woman discus throw at the 1996 Atlanta Olympics. The results are shown at 6 intervals of fields.

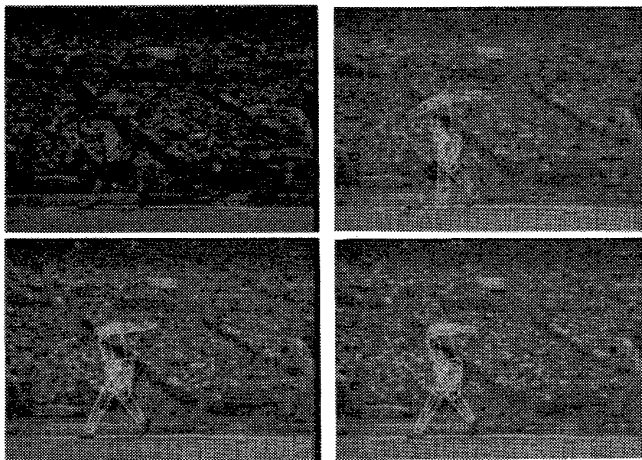


図6 反復計算によるモデルの人体への収束。15フィールド目で示した。左上：原画像，右上：5回，左下：50回，右下：500回

Fig.6 These images depict several steps in the iterative approach of a model to the 15th field of the discus-throw image sequence by the keyframe tracking. Left-upper, right-upper, left-down and right-down images are an original image at the 15th field, 5th, 50th and 500th iterations of convergence, respectively.

に収束していく様子を図6に示す。ここでは、第15フレームに対し、図6の左上に原画像を、右上、左下、右下にそれぞれ5回、50回、500回繰り返したときの結果を示す。

7. むすび

いくつかのキーフレームで与えられる人体の位置と姿勢の拘束を、他のフレームにも伝搬させることにより、ずれの起こらない追跡手法を提案した。

この伝搬過程に必要な動きの初期推定値として、隠れが起こらないときには従来の時空間拘束による推定、隠れが起こったときにはカルマンフィルタによる予測を用いた。

提案した手法を用いて、早くて複雑な動作例として、円盤投げの投てき動作を追跡することができた。このことは、古い映画からでも俳優の動作追跡が可能であることを意味する。追跡動作をCG上に復元することにより、過去の俳優によるパフォーマンスアニメーションやゲームキャラクタの動作生成が可能である。

今後の課題として、現在手作業で行っているキーフレームでのモデル照合の自動化、およびより正確な予測法の確立が挙げられる。

謝辞 本研究は一部、科研費重点領域研究・人工現実感 (No.06212208) の支援を受けた。

文 献

- [1] D.M. Gavrila and L.S. Davis, "3-D model-based tracking of humans in action: A multi-view approach," Proc. of IEEE CVPR'96, pp.73-80, 1996.
- [2] L. Goncalves, E.D. Bernardo, E. Ursella, and P. Peroma, "Monocular tracking of the human arm in 3D," Proc. of 5th ICCV, pp.764-770, 1995.
- [3] B.K.P. Horn and B.G. Schunk, "Determining optical flow," Artificial Intelligence, vol.17, nos.1-3, pp.185-203, 1981.
- [4] 岩井儀雄, 八木康史, 谷内田正彦, "単眼動画画像からの手の3次元運動と位置の推定," 信学論 (D-II), vol.J80-D-II, no.1, pp.44-55, Jan. 1997.
- [5] I.A. Kakadiaris and D. Metaxas, "Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection," Proc. of IEEE CVPR'96, pp.81-87, 1996.
- [6] 亀田能成, 美濃導彦, 池田克夫, "シルエット画像からの関節物体の姿勢推定法," 信学論 (D-II), vol.J79-D-II, no.1, pp.26-35, Jan. 1996.
- [7] 木本伊彦, 梶谷昭彦, 安田靖彦, "スティックモデルに基づく単眼視動画画像からの人体歩行運動の解析の一手法," 信学論 (D-II), vol.J74-D-II, no.3, pp.376-387, March 1991.
- [8] 北川源四郎, "時系列解析プログラミング," 岩波書店, 1993.
- [9] R. Koch, "Dynamic 3-D scene analysis through synthesis feedback control," IEEE PAMI, vol.15, no.6, pp.556-568, 1993.
- [10] 倉掛正治, ラマカント ネバティア, "関節のある動物体の記述と追跡," 信学論 (D-II), vol.J76-D-II, no.4, pp.854-862, April 1993.

[11] H. Li, P. Roivainen, and R. Forchheimer, "3-D motion estimation in model-based facial image coding," IEEE PAMI, vol.15, no.6, pp.545-555, 1993.

[12] J. Ohya and F. Kishino, "Human posture estimation from multiple images using genetic algorithm," 12th ICPR, pp.750-753, 1994.

[13] R.B. Paul, "Robot manipulators, mathematics, programming, and control," The MIT Press, Cambridge Massachusetts, 1991.

[14] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," Nature, vol.317, no.6035, pp.314-319, 1985.

[15] J.M. Rehg and T. Kanade, "DigitEyes: Vision-based human hand tracking," CMU-CS-93-220, 1993.

[16] K. Rohr, "Towards model-based recognition of human movements in image sequences," CVGIP: Image Understanding, vol.59, no.1, pp.94-115, 1994.

[17] 佐藤明知, 川田 聡, 大崎喜彦, 山本正信, "多視点動画像からの人間動作の追跡と再構成," 信学論 (D-II), vol.J80-D-II, no.6, pp.1581-1589, June 1997.

[18] 島田伸敬, 白井良明, 久野義徳, "確率に基づく探索と照合を用いた画像からの手指の3次元姿勢推定," 信学論 (D-II), vol.J79-D-II, no.7, pp.1201-1217, July 1996.

[19] 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次元動画追跡," 信学論 (D-II), vol.J79-D-II, no.1, pp.71-83, Jan. 1996.

付 録

1. 回転運動の終端拘束の偏微分

回転運動の終端拘束を $L_1 = \|\hat{R}'_{i,n} - R'_{i,n}\|^2$ とおく。

$$\begin{aligned}
 L_1 &= \|\hat{R}'_{i,n} - R'_{i,n}\|^2 \\
 &= \|R_{k,0}^{-1}R_{i,0} \prod_{l=1}^n Q_{i,l} - R_{k,n}^{-1}R_{i,n}\|^2 \\
 &= \|(R_{k,0}^{-1}R_{i,0} \prod_{l=1}^{j-1} Q_{i,l})Q_{i,j}(\prod_{l=j+1}^n Q_{i,l}) \\
 &\quad - R_{k,n}^{-1}R_{i,n}\|^2 \\
 &= \|(R_{k,0}^{-1}R_{i,0} \prod_{l=1}^{j-1} Q_{i,l})\{Q_{i,j} \\
 &\quad - (R_{k,0}^{-1}R_{i,0} \prod_{l=1}^{j-1} Q_{i,l})^{-1} \\
 &\quad R_{k,n}^{-1}R_{i,n}(\prod_{l=j+1}^n Q_{i,l})^{-1}\}(\prod_{l=j+1}^n Q_{i,l})\|^2
 \end{aligned}$$

ここで, $(R_{k,0}^{-1}R_{i,0} \prod_{l=1}^{j-1} Q_{i,l})$ と $(\prod_{l=j+1}^n Q_{i,l})$ は正規直交行列であるので,

$$L_1 = \|Q_{i,j} - P_{i,j}\|^2$$

となる。但し,

$$\begin{aligned}
 P_{i,j} &= (R_{k,0}^{-1}R_{i,0} \prod_{l=1}^{j-1} Q_{i,l})^{-1}R_{k,n}^{-1}R_{i,n}(\prod_{l=j+1}^n Q_{i,l})^{-1} \\
 &= \begin{pmatrix} p_1(i,j) & p_2(i,j) & p_3(i,j) \\ p_4(i,j) & p_5(i,j) & p_6(i,j) \\ p_7(i,j) & p_8(i,j) & p_9(i,j) \end{pmatrix}
 \end{aligned}$$

とする。このとき,

$$\begin{aligned}
 L_1 &= (1 - p_1)^2 + (\phi_{i,j} + p_2)^2 + (\psi_{i,j} - p_3)^2 \\
 &\quad + (\phi_{i,j} - p_4)^2 + (1 - p_5)^2 + (\varphi_{i,j} + p_6)^2 \\
 &\quad + (\psi_{i,j} + p_7)^2 + (\varphi_{i,j} - p_8)^2 + (1 - p_9)^2
 \end{aligned}$$

であるので, 運動パラメータ x_j に関して偏微分を行うと,

$$\begin{aligned}
 \frac{\partial L_1}{\partial \alpha_{1,j}} &= 0, \quad \frac{\partial L_1}{\partial \beta_{1,j}} = 0, \quad \frac{\partial L_1}{\partial \gamma_{1,j}} = 0 \\
 \frac{\partial L_1}{\partial \varphi_{i,j}} &= 2(2\varphi_{i,j} - (p_8^* - p_6^*)) \\
 \frac{\partial L_1}{\partial \psi_{i,j}} &= 2(2\psi_{i,j} - (p_3^* - p_7^*)) \\
 \frac{\partial L_1}{\partial \phi_{i,j}} &= 2(2\phi_{i,j} - (p_4^* - p_2^*))
 \end{aligned}$$

となる。ここで, 添字の * は繰り返し一つ前の推定値を示す。従って,

$$\frac{\partial L_1}{\partial x_j} = 4(E_{r,j}x_j - b_{er,j})$$

但し,

$$E_{r,j} = \begin{pmatrix} 0 & & & \\ & I & & \\ & & \ddots & \\ & & & I \end{pmatrix} \tag{A.1}$$

とする。0, I は, それぞれサイズが3のゼロ行列, 単位行列である。また,

$$b_{er,j} = \begin{pmatrix} 0 \\ c_{1,j} \\ \vdots \\ c_{m,j} \end{pmatrix} \tag{A.2}$$

但し,

$$\mathbf{c}_{i,j} = \begin{pmatrix} (p_8^* - p_6^*)/2 \\ (p_3^* - p_7^*)/2 \\ (p_4^* - p_2^*)/2 \end{pmatrix}$$

とし, $\mathbf{0}$ は長さ 3 のゼロベクトルである.

2. 並進運動の終端拘束の偏微分

並進運動の終端拘束 $\mathbf{L}_2 = \|\hat{\mathbf{T}}_{1,n} - \mathbf{T}_{1,n}\|^2$ は, 並進運動パラメータ $\mathbf{S}_{1,j}$ のみの関数であるので,

$$\begin{aligned} \frac{\partial \mathbf{L}_2}{\partial \mathbf{S}_{1,j}} &= 2(\hat{\mathbf{T}}_{1,n} - \mathbf{T}_{1,n}) \\ &= 2(\mathbf{S}_{1,j} - (\mathbf{T}_{1,n} - \hat{\mathbf{T}}_{1,n}^* + \mathbf{S}_{1,j}^*)) \end{aligned}$$

となる. 故に,

$$\frac{\partial \mathbf{L}_2}{\partial \mathbf{x}_j} = 2(E_{t,j} \mathbf{x}_j - \mathbf{b}_{et,j})$$

となる. ここで,

$$E_{t,j} = \begin{pmatrix} \mathbf{I} & & & \\ & \mathbf{0} & & \\ & & \ddots & \\ & & & \mathbf{0} \end{pmatrix} \quad (\text{A.3})$$

とする. \mathbf{I} , $\mathbf{0}$ はサイズが 3 の単位行列, ゼロ行列である. また,

$$\mathbf{b}_{et,j} = \begin{pmatrix} \mathbf{T}_{1,n} - \hat{\mathbf{T}}_{1,n}^* + \mathbf{S}_{1,j}^* \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} \quad (\text{A.4})$$

とする. $\mathbf{0}$ は長さ 3 のゼロベクトルである.

3. 滑らか拘束の偏微分

滑らか拘束を

$$\mathbf{L}_3 = \sum_{j=2}^n (\mathbf{x}_j - \mathbf{x}_{j-1})^\top \text{diag} \left(\frac{\nu}{2} \right) (\mathbf{x}_j - \mathbf{x}_{j-1})$$

とおけば, その偏微分は

$$\frac{\partial \mathbf{L}_3}{\partial \mathbf{x}_j} = 4 \text{diag} \left(\frac{\nu}{2} \right) \left(\mathbf{x}_j - \frac{\mathbf{x}_{j-1}^* + \mathbf{x}_{j+1}^*}{2} \right)$$

となる. 但し, $\mathbf{x}_0^* \equiv \mathbf{x}_1^*$, $\mathbf{x}_{n+1}^* \equiv \mathbf{x}_n^*$ とする. 従って,

$$\frac{\partial \mathbf{L}_3}{\partial \mathbf{x}_j} = 2(\mathbf{A}_{s,j} \mathbf{x}_j - \mathbf{b}_{s,j})$$

となる. 但し,

$$\mathbf{A}_{s,j} = \text{diag}(\nu) \quad (\text{A.5})$$

$$\mathbf{b}_{s,j} = \text{diag}(\nu) \frac{\mathbf{x}_{j-1}^* + \mathbf{x}_{j+1}^*}{2} \quad (\text{A.6})$$

とする.

4. カルマンフィルタによる予測

カルマンフィルタは, 1 フレーム先の予測とフィルタリングを交互に繰り返しながらフレーム順に進められる [8]. すなわち, $j-1$ フレームまでに観測された運動パラメータを使った j フレームの状態予測は,

$$\begin{aligned} \mathbf{s}_{j|j-1} &= \mathbf{F} \mathbf{s}_{j-1|j-1} \\ \mathbf{V}_{j|j-1} &= \mathbf{F} \mathbf{V}_{j-1|j-1} \mathbf{F}^\top + \mathbf{G} \Lambda_v \mathbf{G}^\top \end{aligned}$$

で与えられる. ここで, 添字 $a|b$ は b フレームまでの情報から推測された a フレームでの推測値を意味する. また, \mathbf{V} , Λ_v はそれぞれ状態 \mathbf{s} , ノイズ \mathbf{v} の共分散行列である. なお, ノイズの共分散行列はレビンソンのアルゴリズム [8] により推定した.

j フレームでの運動パラメータ y_j を得たときのフィルタリングは,

$$\begin{aligned} \mathbf{K}_j &= \mathbf{V}_{j|j-1} \mathbf{H}^\top (\mathbf{H} \mathbf{V}_{j-1|j-1} \mathbf{F}^\top + \Lambda_w)^{-1} \\ \mathbf{s}_{j|j} &= \mathbf{s}_{j|j-1} + \mathbf{K}_j (y_j - \mathbf{H} \mathbf{s}_{j|j-1}) \\ \mathbf{V}_{j|j} &= (\mathbf{I} - \mathbf{K}_j \mathbf{H}) \mathbf{V}_{j|j-1} \end{aligned}$$

で与えられる. \mathbf{K} はカルマンゲインであり, Λ_w はノイズ w の共分散行列である.

$j+1$ フレームで隠れが起こったとき, k フレーム先の状態予測は

$$\begin{aligned} \mathbf{s}_{j+k|j} &= \mathbf{F} \mathbf{s}_{j+k-1|j} \\ \mathbf{V}_{j+k|j} &= \mathbf{F} \mathbf{V}_{j+k-1|j} \mathbf{F}^\top + \mathbf{G} \Lambda_v \mathbf{G}^\top \end{aligned}$$

により与えられる. このとき運動パラメータの予測値は,

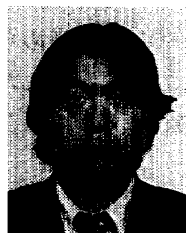
$$y_{j+k|j} = \mathbf{H} \mathbf{s}_{j+k|j}$$

で与えられる.

(平成 9 年 8 月 22 日受付, 10 年 2 月 10 日再受付)

**大田 佳人**

平7新潟大・工・情報卒，平9同大大学院自然科学研究科前期博士課程了，同年，(株)ナムコ入社，在学中は動画像解析の研究に従事。

**山際 貴志**

平7新潟大・工・情報卒，平9同大大学院自然科学研究科前期博士課程了，同年，東芝マイクロエレクトロニクス(株)入社，在学中は，感性情報処理および動画像解析に関する研究に従事。

**山本 正信 (正員)**

昭48九工大・工・制御卒，昭50東工大大学院修士課程了，同年，電総研入所，動画像処理，コンピュータビジョン等の研究に従事，平1~2カナダ国立研究協議会招聘研究員，昭62情処学会研究賞受賞，平4より新潟大工学部情報工学科教授，工博，情報処理学会，IEEE各会員。