

装着型能動視覚センサを用いた人物の運動推定

杉本晃宏[†] 長友渉[‡] 松山隆司[‡]

[†] 国立情報学研究所 [‡] 京都大学情報学研究科

sugimoto@nii.ac.jp

2 台の装着型能動カメラをそれぞれ独立に注視点制御 (2 眼独立注視点制御) し, それによって得られる点対応および直線対応を利用して, 3 次元世界における装着者の回転運動と並進運動を逐次的に推定する手法を提案する. 本手法では, 各カメラを独立に注視点制御するので, 2 台のカメラは視野を共有する必要がない. さらに, カメラが注視している 2 点からみると, 人物を対象としてステレオ視をしている状況であると解釈できるので, 空間的に離れた点をカメラに注視させることにより, 装着した 2 台のカメラの基線長に依存せず, 高精度の運動推定が可能となる.

Estimating Human's Motion by Wearable Active Cameras

Akihiro Sugimoto[†], Wataru Nagatomo[‡] and Takashi Matsuyama[‡]

[†]National Institute of Informatics

[‡]Graduate School of Informatics, Kyoto University

We propose a method for incrementally estimating motion of a human wearing two active cameras. In our method, we apply 'fixation control' to each active camera. That is, while a human moves we control the two active cameras independently so that each fixates its viewing line to its own fixation point. The correspondence of a fixation point over two frames as well as the correspondence of lines nearby the fixation point gives us sufficient constraints to determine human's motion. Two cameras do not have to share the viewing field in our method whereas in the stereo vision they have to do. This property leads to accurate motion estimation independent of the baseline length of the two cameras.

1 はじめに

コンピュータの急速な高性能化, 小型化, 低価格化にともない, コンピュータは我々の生活により身近なものになってきている. そして, そう遠くない将来に, コンピュータを装着したまま日常生活することが現実味をおびつつある. 近年, そのようなウェアラブルコンピュータの実用化を目指した研究が盛んに行われている [1, 2, 5, 6, 9].

しかし, これらの研究では, コンピュータは, 主体である人間に対して従であると位置づけられ, 人間の活動を支援するための道具として捉えられている. ウェアラブルコンピュータをより自然に扱うための技術として, 力学メディアを含むマルチメディアを活用した PUI (perceptual user interface) が提唱されているが, これも, やはり, 人間とコンピュータとは主従の関係にあるという立場, すなわち, 人間にとってコンピュータは道具であるというコンセプトに立脚している.

これに対し, 21 世紀の日常生活環境におけるコン

ピュータは, それ自身が主体性を有し, 人間のパートナーとして存在するという立場に立脚するべきである. そしてそのような立場にたつて, 人間と共生するコンピュータの在り方を探るべきである. 人間と共生するコンピュータは, コンピュータ自身が自発的・自律的に, 人間がおかれている周囲の状況, 人間の行動や意図を理解し, それに基づいて人間との動的なインタラクションを実現する.

このようなコンピュータの実現を目指して, 我々は, 視線測定装置とコンピュータ制御可能な 2 台の首振りカメラで構成される装着型能動視覚センサを開発した (図 1). そして, 見ている点の奥行きが変わるような現実的な環境下でも, 正確に装着者の視線情報を検出する手法を提案した [11]. 本稿では, このセンサを構成する 2 台の装着型能動カメラを用いて, 人物の運動を推定する手法を述べる. そこでは, 各カメラを独立に注視点制御する 2 眼独立注視点制御を用いる (図 2). そして, その制御を通じて得られる点対応, および, 注視点周辺で得られる情報としての直線対応を利用すること



図 1: 装着型能動視覚センサ

で、3次元世界における装着者の回転運動と並進運動を逐次的に推定する。

ロボットビジョンの分野では、移動ロボットの位置や運動を推定するために、ステレオ視を利用する手法が一般的である [3, 7, 8]。しかし、ステレオ視を利用する場合、2台のカメラの視野を共有させる必要がある。そして、運動推定のためには、2台のカメラで撮影された画像において複数の特徴対応を得る必要があり、その処理が実時間性を損う。さらに、運動の推定精度は2台のカメラの基線長に依存する。これに対し、2眼独立注視点制御では、各カメラを独立に制御して空間中の点を注視させればよい(注視する点を2台のカメラで一致させる必要はない)ので、2台のカメラの視野を必ずしも共有させる必要がない。さらに、カメラが注視している2点からみると、人物を対象としてステレオ視をしている状況であると解釈できるので、その場合の基線長はカメラが注視している2点間の距離に一致する。すなわち、空間的に離れた点をカメラに注視させることにより、装着した2台のカメラの基線長に依存せず、高精度の運動推定が可能になる。このように、2眼独立注視点制御によって、高精度に人物の運動を推定することが可能になる。

2 装着型能動視覚センサの意義

コンピュータが人間がおかれている状況を理解するためには、その人の周囲の情報を獲得するセンサが必

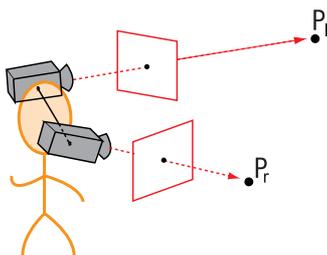


図 2: 2眼独立注視点制御

要となる。人の視線は、意識するしないにかかわらず、その人の興味や注意を強く反映していると考えられるため、人間と入力情報を共有するという点においても、また、その情報量の多さという点でもカメラは有力なセンサである。

一方、人間がおかれている周囲の状況やその人の行動などを理解するために、天井や壁面に設置された複数のセンサ(磁気センサやカメラなど)を用いることが考えられる。しかし、環境に埋め込まれたセンサからの入力情報、すなわち、客観的な視点からの入力情報は、行動しているその人の意図や興味等をとらえるという観点からは、不十分である。本人の視点、すなわち、主観的な視点からの入力情報が必要であると考えられる。これは、人間同士でコミュニケーションする際に、視点空間的に大きく異なる場合、意思の疎通が上手くいかないことがたびたびあることから容易に推測できる。人間の行動や意図の理解という観点からは、その人と視点を共有し、さらに、入力情報を共有する装着型視覚センサが不可欠である。これにより、コンピュータが、人間と同じ視点に立ち、自分の周りの空間や自分の動きを理解することが可能になる。

さらに、コンピュータによって視覚センサを能動的に制御することができれば、装着者の視線とは独立して、コンピュータが自発的・自律的に周囲の情報を獲得することができる。すなわち、状況に応じて、(1)装着者と視線を共有することで、装着者の視点からの情報を得ること(主観的情報の獲得)と、(2)装着者の視線とは独立して、自発的・自律的に周囲の情報を得ること(客観的情報の獲得)とを、適宜、切り替えて、コンピュータは対応することができる。これにより、視線固定のカメラを用いた装着型視覚センサでは実現できない機能を付与することができ、人間の行動・意図の理解のために多角的に情報を獲得することが可能になる。

このように、装着型能動視覚センサは、人間の行動や意図を理解するための有効な手段であると考えられる。

3 注視点制御

3.1 注視点の検出

空間中のある点に常にカメラの視線が向くようにカメラの視線を制御することを注視点制御 [4] とよぶ。注視点制御を実現するためには、カメラは、人物の運動を推定するのに適した、3次元空間中の点を自律的に注視点として検出し、その点にカメラの視線を向けるようにカメラを制御する必要がある。静的な3次元空間を

前提とすると、カメラが注視する点には、以下の要件が求められる。これらの要件を満たす点が注視点として適切であり、そのような点を検出する必要がある。

実在性 3次元空間中に実在する点であること。空間中で視点からの奥行きが異なる2つの物体が、カメラからは重なって撮影された際に、物体の境界の部分は実在しないにもかかわらず、画像上では、点として観測される場合がある。このような点は、人物の運動にともなって3次元空間中を移動するので注視点として適切でない。

同定の容易さ 撮影される画像上において、同定が容易な点であること。注視点の同定が困難であると、設定した注視点を見失ってしまう危険性があり、注視点制御を正確に実行することができない。

物理的制御の安定性 カメラの視線方向の物理的制御上、人物の運動により、その点がカメラの視野から容易に外れないこと。すなわち、カメラの可動範囲に、ある程度の余裕をもったカメラの視線方向に存在する点であること。

空間的距離の確保 他方のカメラの注視点と可能な限り空間的に離れた点であること。これは、既に述べたように、2眼独立注視点制御は、注視している2点から人物を対象としてステレオ視している状況であると解釈できるので、2つの注視点間の距離が大きいほど、運動の推定精度がよくなることが期待されるからである。

3.2 テンプレートマッチングを用いた注視点制御

注視点を設定したとき、注視点制御を実現するためには、検出された注視点が必要な時刻に撮影されるフレーム内のどこに移動したかを求め、移動後の点の方向にカメラの視線を向ける必要がある。これは、テンプレートマッチングを用いることによって実現される。

まず、注視点を中心とするテンプレートを参照テンプレートとして切り出し、次に、これを用いて、次の時刻に撮影された画像とテンプレートマッチングを行う。もっともマッチングの評価値の高い点を求めることによって、移動後の注視点を見つけ出すことができる。見つけ出された注視点が必要となるように能動カメラを制御することによって、注視点制御が実現される。なお、テンプレートマッチングでは、対応する画素ごと

の明度差を参照テンプレートにわたって加算し、その逆数をその点におけるマッチングの評価値とすればよい。

3.3 注視点の更新

3次元空間中における人物の運動を推定するためには、2眼独立注視点制御による人物の運動推定が、途切れることなく継続する必要がある。しかし、3次元空間中における人物の運動によっては、カメラが注視点に視線を物理的に向けることができなくなる場合が生じる。その際には、人物の運動推定に適した別の注視点が必要となる。つまり、注視点を更新しなければならない。

新たに注視点として設定される点も、やはり、第3.1節に述べた要件を満たさなければならない。ただし、注視点制御は2台のカメラで独立に行っているため、注視点としての要件をカメラごとに独立に適応し、新しい注視点を選択すればよい。

実装する際には、人物が空間中を運動することによって、カメラが注視点に視線を物理的に向けることができなくなる前の段階で、予め、次に注視点とする点を保持しておく必要がある。そして、カメラが現在の注視点に視線を物理的に向けることができなくなったとき、その点を新たな注視点とし、新たな注視点に対して、注視点制御を実行すればよい。これにより、人物が3次元空間中を自由に運動しても、途切れることなくその運動を推定することが可能になる。

4 人物の運動推定

2眼独立注視点制御によって得られる情報を利用して人物の運動を推定することを考える。本稿では、能動視覚センサ装着者の3次元空間中での運動を、便宜上、その人物が装着しているカメラの投影中心の運動としてとらえ、以後はカメラ運動と記述する。なお、以下では、2台のうち一方のカメラを右カメラ、他方を左カメラとよぶことにし、右カメラを基準に考える。すなわち、カメラ運動とは、右カメラの投影中心の運動を意味するとする。

それぞれのカメラの内部パラメータに加え、2台のカメラの相対的な位置関係も校正されているとする。すなわち、左カメラのカメラ座標系を、右カメラのカメラ座標系に一致させる並進ベクトルは、左カメラのカメラ座標系において T_{in} で表され、回転行列は、右カメラのカメラ座標系において R_{in} で表されるとしたとき、 T_{in} 、 R_{in} は、ともに既知であるとする。

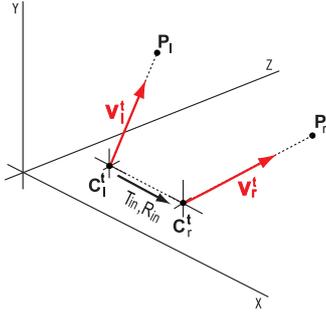


図 3: 時刻 t におけるカメラと注視点との関係

4.1 注視点対応に基づく制約

注視点制御によって、3次元空間中に存在する点に向かう視線方向の時系列フレーム間での対応（注視点対応とよぶことにする）が得られる。この注視点対応からカメラの回転運動及び並進運動に関する制約式を導出する。

時刻 t において、人物が装着している左右のカメラの投影中心は、それぞれ、3次元空間中の点 C_l^t, C_r^t にあるとする。また、左右のカメラは、それぞれ、3次元空間中の点 P_l, P_r を注視点として注視しているとする。点 C_r^t から注視点 P_r に向かう単位方向ベクトルは、時刻 t における右カメラ座標系では、 v_r^t であるとし、点 C_l^t から注視点 P_l に向かう単位方向ベクトルは、時刻 t における左カメラ座標系では、 v_l^t であるとする（図3）。

まず、基準としている右カメラに注目して考える。いま、時刻 t から時刻 $t+1$ の間の人物の運動の結果、右カメラの投影中心が、点 C_r^t から点 C_r^{t+1} に移動したとする（図4）。そして、このときの右カメラの回転および並進運動は、時刻 t における右カメラ座標系での回転行列 R 、および、ワールド座標系での並進ベクトル T で表せるとする。なお、時刻 t における右カメラ座標系の姿勢は、回転行列 R_0^{-1} を作用させることによって、ワールド座標系の姿勢に一致させることができるとしておく。

このとき、 λ, λ' を非零定数として、右カメラでの注視点対応からは次式が得られる：

$$\lambda R_0 v_r^t = \lambda' R_0 R v_r^{t+1} + T.$$

そしてこれは、次のように書き換えられる。

$$\det \begin{bmatrix} R_0 v_r^t & R_0 R v_r^{t+1} & T \end{bmatrix} = 0. \quad (4.1)$$

一方、左カメラで得られる v_l^t は、時刻 t における右カメラ座標系では、 $R_{in} v_l^t$ と表される。また、時刻 t が

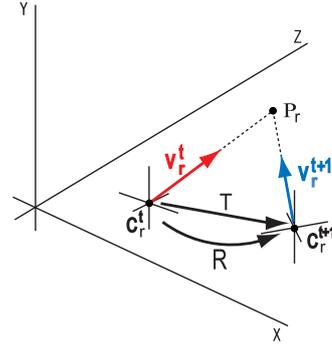


図 4: 右カメラの注視点対応に基づく幾何学的制約

ら時刻 $t+1$ への右カメラの回転運動 R によって、左カメラはワールド座標系で、 $-R_0(R-I)R_{in}T_{in}$ だけ平行移動することになる。したがって、左カメラでの注視点対応からは、

$$\det [R_0 R_{in} v_l^t \quad R_0 R R_{in} v_l^{t+1} \quad T - R_0(R-I)R_{in}T_{in}] = 0. \quad (4.2)$$

が得られる。ここに、 I は単位行列を表す。

式 (4.1), (4.2) が、注視点対応によって得られる3次元空間中でのカメラ運動の制約式である。時刻 t までのカメラ運動を推定できているとすると、 R_0 は既知としてよいので、式 (4.1), (4.2) において、未知数は R と T である。そして、式 (4.1), (4.2) は、それぞれ、 R と T に関して、同次2次の制約式になっている。

4.2 直線対応に基づく制約

カメラ運動のうち、回転運動には3自由度あり、並進運動にも3自由度ある。これに対し、注視点対応によって得られるカメラ運動の制約式は、2個である（式 (4.1), (4.2)）。したがって、注視点対応から得られる制約式だけでカメラ運動を推定することはできない。

そこで、注視点の周辺にある直線の対応を利用することで、カメラ運動の制約式の不足を補うことを考える。直線には、

- (i) 壁と天井の境界、建造物の窓やドアなどその存在が豊富である、
- (ii) Hough 変換等によって実時間性を損わず容易に検出可能である、
- (iii) 空間的な広がりをもっているため、対応づけが容易である、

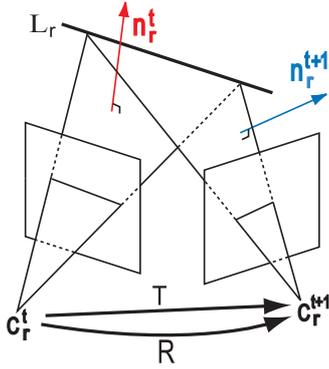


図 5: 右カメラの直線対応に基づく幾何学的制約

という特長があるからである。加えて、下で述べるように、直線の対応から導かれるカメラ運動の制約式は、回転運動のみに依存するという利点がある。

まず、基準としている右カメラに注目して考える。時刻 t において、人物が装着している右カメラの投影中心が、3次元空間中の点 C_r^t にあるとする。そして、右カメラが、3次元空間中に存在する直線 L_r の対応をとっているとする（以下では、直線 L_r を対応直線とよぶ）。対応直線 L_r の単位方向ベクトルをワールド座標系で表すと L_r であるとする¹。また、点 C_r^t （観測時刻におけるカメラの投影中心）と対応直線 L_r とで張られる平面（以下では、この平面を対応直線 L_r の解釈面とよぶ）の単位法線ベクトルを、時刻 t における右カメラ座標系で表したとき、 n_r^t であるとする（図 5）。

以上の状況下で、ワールド座標系の姿勢と、時刻 t および時刻 $t+1$ における右カメラ座標系の姿勢との関係をそれぞれ考えると、ベクトル n_r^t, n_r^{t+1} は、ワールド座標系では $R_0 n_r^t, R_0 R n_r^{t+1}$ と表されることがわかる。そして、 $R_0 n_r^t$ と L_r は直交し、また、 $R_0 R n_r^{t+1}$ と L_r も直交している。これより、右カメラでの直線対応に基づくカメラ運動の制約式が得られる：

$$\mu_r L_r = (R_0 n_r^t) \times (R_0 R n_r^{t+1}). \quad (4.3)$$

ここに、 μ_r は、直線に依存して決まる非零定数である。

同様に、左カメラでの直線対応に基づくカメラ運動の制約式が得られる：

$$\mu_l L_l = (R_0 R_{in} n_l^t) \times (R_0 R R_{in} n_l^{t+1}). \quad (4.4)$$

ここに、 L_l は左カメラでの対応直線の単位方向ベクトルをワールド座標系で表したものであり、 μ_l は直線に依存する非零定数である。また、 n_l^t は、左カメラにお

¹ ここでは、対応直線のワールド座標系での単位方向ベクトルは既知であるとしておく。しかし、これは、式 (4.3) の関係を用いて、運動推定の段階で導出することができる。

ける対応直線 L_l の時刻 t における解釈面の単位法線ベクトルを、時刻 t における左カメラ座標系で表したものである。

式 (4.3) が、右カメラでの直線対応によって得られる3次元空間中のカメラ運動の制約式であり、式 (4.4) が、左カメラでの直線対応によって得られる3次元空間中のカメラ運動の制約式である。これらからわかるように、直線対応に基づく制約式には、カメラ運動のうち並進運動成分が含まれていない。

時刻 t までのカメラ運動を推定できているとすると、 R_0 は既知としてよいので、式 (4.3)、式 (4.4) おいて、未知数は R と非零定数である。そして、この制約式は、 R と非零定数に関して、同次1次式となっている。

4.3 回転運動と並進運動の推定

直線対応に基づく制約式 (4.3)、(4.4) は回転運動のみに依存するので、カメラ運動の推定を回転運動の推定と並進運動の推定とに分離することができる。

まず、カメラの回転運動の推定を考える。 n 本の直線対応が得られているとき、式 (4.3)、(4.4) の未知数は $n+3$ となる。一方、制約式は、3次元のベクトル方程式なので、 $3n$ 個である。したがって、2本以上の対応直線が得られれば、回転運動を求めることができる。

実際に回転運動を求めるには、式 (4.3)、(4.4)、および、直交性の制約 ($RR^T = I$) を連立させ、非線形連立方程式を解けばよい。非線形連立方程式には、一般に、解が複数個存在するので、2本の直線対応から回転運動を導出すると、実際のカメラの回転運動とは異なる解に陥ってしまうことがある。しかし、これは、直線対応を冗長にとることによって回避することができる。なぜなら、直線対応によって得られる回転運動の制約式は未知数に関して1次の制約式となっているので、冗長性をもたせることによって、にせの解を排除することができるからである²。

回転運動が求まれば、未知数は並進運動のみとなる。回転運動が既知のとき、式 (4.1)、(4.2) は、並進運動に関して同次1次方程式になる。したがって、2点の注視点対応から、線形計算によって、スケール倍の不定性を除いて並進運動を求めることができる³。

² これに加えて、冗長性をもたせる理由がもう一つある。カメラ運動が回転のみの場合や、カメラの投影中心がそのカメラで観測している対応直線の解釈面上を運動する場合には、その対応直線を観測して得られる直線対応から導かれる制約式は恒等式になる。すなわち、その直線対応からは、カメラ運動の独立な制約式が得られない。冗長性をもたせることによってこのような状況の発生を防ぐことができる。

³ 2フレーム間でのカメラの並進運動を推定するときにこの不定性が残る。しかし、3枚の画像間の対応点に対して成立する幾何学的

以上により, 2 点の注視点対応とその周辺情報としての 2 本以上の対応直線から, カメラの回転運動および並進運動を推定することができることがわかる. すなわち, 内部および外部パラメータを校正した 2 台の能動カメラを独立に注視点制御することにより, 3 次元空間中でのカメラの回転運動と並進運動を推定することができる.

5 アルゴリズム

これまでの議論に基づいて, 2 眼独立注視点制御により, 3 次元空間中でのカメラの運動を推定するアルゴリズムを記述する.

Step 0: 各カメラでそれぞれの注視点を検出し, 注視点周辺の対応直線を選択する. $t = 1$ とする.

Step 1: $v_r^t, v_\ell^t, n_r^t, n_\ell^t$ を求める.

Step 2: $i = r, \ell$ に対して, 以下を実行する.

- (a) カメラ i を注視点制御することによって, v_i^{t+1}, n_i^{t+1} を求める. カメラ i の注視点をカメラ i の視野に捉えることができなくなった場合のみ, Step (b) に進む.
- (b) 新たな注視点を検出し, 新たな対応直線を選択し, Step 1 に戻る.

Step 3: 式 (4.3), (4.4), および, $RR^T = I$ を連立させて, カメラの回転運動を推定する.

Step 4: 式 (4.1), (4.2) からカメラの並進運動を推定する.

Step 5: $t = t + 1$ として, Step 2 に戻る.

6 実環境下でのカメラ運動推定実験

提案手法を用いて, 実環境下でカメラ運動を推定する実験を行った. その結果, おおむね正しくカメラの運動軌跡を推定することができた.

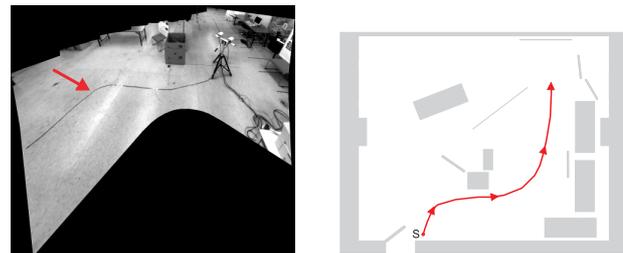
実験には, 以下の装置を用いた.

- 能動カメラ: Sony 製 EVI-G20 \times 2 台
- 雲台, 三脚
- PC (PentiumIII 750MHz) 1 台

拘束 trilinear constraints[10] を用いることで, 異なる 2 フレーム対の間のスケールを整合させることができる. その詳細は付録を参照のこと.



図 6: 擬似装着型能動視覚センサ



(a) 広角視野でみた運動軌跡 (b) 上からみた運動軌跡

図 7: 実環境下でのカメラの運動の軌跡

EVI-20 は, 内部にモータを 2 個搭載していて, パン (水平回転角度 $\pm 30^\circ$), チルト (垂直回転角度 $\pm 15^\circ$) 方向に, カメラの撮像系全体を物理的に回転させることができる. また, 駆動系の回転中心とカメラの投影中心とが一致するように設計されている.

2 台のカメラを, 約 27cm の基線長を保ち, 三脚付きの雲台に設置し, 2 眼独立注視点制御を行う擬似装着型能動視覚センサを構築した (図 6). さらに, Zhang [12] の手法によって, 2 台のカメラの内部・外部パラメータを校正した. 左右のカメラで撮影された画像サイズは, それぞれ, 640×320 (画素) であった.

擬似装着型能動視覚センサを実環境下で移動させた. その際の, 右カメラの移動の軌跡を図 7 に示す. 軌跡の長さは, およそ 6m であった. また, 移動中, その軌跡上の 35ヶ所にマーキングを施し, マーキングを施した各地点を移動中のサンプル地点として 2 眼独立注視点制御を実行した. なお, 入力画像から得たエッジ画像に Hough 変換を施し, 2 本の直線が交わる点を見つけ, そのうち実在する点を目視で選んで注視点として設定した. また, 注視点の更新は 8 回行った. 移動軌跡上の, あるマーキング地点における左右のカメラからの入力画像は図 8 のようになっている. 視野の共有はほとんどなかった (図 8 では, 2 眼独立注視点制御中の注視点と対応直線を重ねがきしている).



(a) 左カメラの画像 (b) 右カメラの画像

図 8: 左右のカメラの入力像

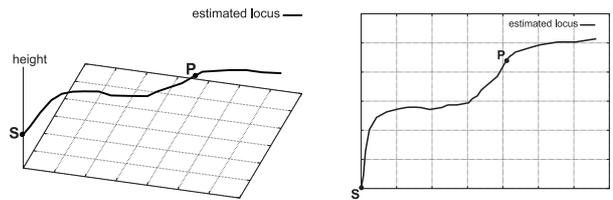
以上の状況の下で、マーキングを施した各地点で右カメラの運動を推定した。カメラ運動の推定のために、各カメラで2本の対応直線を用いた(したがって、対応直線は全部で4本用いた)。画像からの直線の検出にはHough変換を用いた。マーキングを施した各地点で推定された運動をつなぎ合わせて得られたカメラ運動の軌跡を図9に示す(点Sは運動開始地点を表す)。

推定されたカメラ運動の軌跡と実際のカメラ運動の軌跡とを比較すると、床面からの高さは全域にわたってほぼ正しく推定されていることがわかる。これに対し、床面に平行な成分に関しては、運動の最初の段階(図9の点Sから点Pまで)では、両者の軌跡はおおよそ一致しているが、点P以降は、推定されたカメラ運動の軌跡と実際のカメラ運動の軌跡とに大きくズレが生じている。これには、主に二つの原因が考えられる。一つ目は、点P付近のマーキング位置での推定が正しく行われなかったことであり、二つ目は、そこでの誤差がその後の推定に大きく影響を与えたことである。すなわち、逐次的に運動を推定する手法であるため、誤差が累積すると同時に、ひとたび推定に誤りがあるとその後の推定はその誤りで生じた誤差の影響を大きく受けて、推定が正しく行われなくなってしまう。推定誤差が生じた理由としては、画像上の注視点対応の誤差や対応直線の検出誤差が考えられる。また、2台のカメラのパラメタの校正時における誤差によることも考えられる。

以上より、提案手法によっておおむね運動軌跡は正しく推定されているが、逐次型の推定手法であるがゆえに、誤差の累積が推定精度を下げている結果になっているといえる。

7 まとめ

本稿では、自発的・自律的に人間の行動や意図を理解し、それに基づいて人間と動的にインタラクションするコンピュータの実現を目指して、装着型能動カメラを用いた3次元空間中における人物の運動推定手法



(a) 3次元的に表示 (b) 床面上の運動として表示

図 9: 推定されたカメラ運動の軌跡

を提案した。そこでは、2台のカメラに対して注視点制御をそれぞれ独立に実行する2眼独立注視点制御を行う。そして、それを通して得られる点对応、および、注視点周辺で得られる情報としての直線対応を利用することで、3次元世界における人物の回転運動と並進運動を逐次的に推定する。

2眼独立注視点制御では、各カメラを独立に注視点制御すればよい(注視する点は2台のカメラで一致させる必要はない)ので、2台のカメラの視野を必ずしも共有させる必要がない。カメラが注視している2点からみると、2眼独立注視点制御は、人物を対象としてステレオ視をしている状況であると解釈できるので、その場合の基線長はカメラが注視している2点間の距離に一致する。すなわち、空間的に離れた点をカメラに注視させることにより、装着した2台のカメラの基線長に依存せず、高精度の運動推定が可能となる。

逐次推定に起因する累積誤差を排除し、推定精度を向上させることが今後の課題である。また、今後は、人物の注視点とカメラの注視点を一致させ、人物の内面にある主観的な情報を抽出することを目指す予定である。

謝辞 本研究は、科学研究費補助金 特定領域研究 13224051、および、基盤研究 14380161 の補助を受けて行った。

参考文献

- [1] 青木: ウェアブル・コンピュータ向けリアルタイム Personal Positioning System, 情報処理学会論文誌, Vol.41 (2000), 9, pp. 2404-2412.
- [2] B. Clarkson, K. Mase, A. Pentland: *Recognizing User's Context from Wearable Sensors: Baseline System*, Vision and Modeling Technical Report, TR-519, Media Lab. MIT, 2000.
- [3] A. J. Davison and D. W. Murray: Mobile Robot Localisation Using Active Vision, *Proc. of ECCV*, Vol. 2, pp. 809-825, 1998.

- [4] 出口, 鏡, 嵯峨, 本谷: 能動カメラによる運動物体追跡と実時間形状復元システム, 計測自動制御学会論文集, Vol. 35 (1999), 5, pp. 675–683.
- [5] M. Kouroggi, T. Kurata and K. Sakaue: A Panorama-Based Method of Personal Positioning And Orientation And Its Real-Time Applications for Wearable Computers, *Proc. of Int. Symposium on Wearable Computers*, pp. 107–114, 2001.
- [6] 興相正克, 蔵田武志, 坂上勝彦, 村岡洋一: パノラマ画像群を位置合わせに用いた注釈つき映像の実時間提示システム—パノラマベースト・アノテーション—, 信学技報, Vol. PRMU99-64, pp. 1–8 (1999).
- [7] N. Molton and M. Brady: Practical Structure and Motion from Stereo When Motion is Unconstrained *Int. J. of Computer Vision* Vol. 39 (2000), 1, pp. 5–23.
- [8] D. W. Murray, I. D. Reid and A. J. Davison: Steering and Navigation Behaviours Using Fixation, *Proc. of British Machine Vision Conference*, 1996.
- [9] 榊原誠之, 大隈隆史, 竹村治雄, 横矢直和: ビデオシースルー型拡張現実感のための実時間ステレオ画像合成, 電子情報通信学会論文誌 D-II, Vol. J82-D-II, No. 10, pp. 1775–1783 (1999).
- [10] 杉本: 幾何学的不変量とその応用, コンピュータビジョン: 技術評論と将来展望 (松山, 久野, 井宮編) 第7章, 新技術コミュニケーションズ, pp. 97–122, 1998.
- [11] A. Sugimoto, A. Nakayama and T. Matsuyama: Detecting a Gazing Region by Visual Direction and Stereo Cameras, *Proc. of Int. Conf. on Pattern Recognition*, 2002, to appear.
- [12] Z. Zhang: A Flexible New Technique for Camera Calibration, *IEEE Trans. PAMI*, Vol. 22 (2000), 11, pp. 1330–1334.

付録 (推定された並進ベクトルのスケールの整合化)

時刻 t と $t+1$ に撮影された画像から推定された, 時刻 t から $t+1$ への右カメラの回転運動行列を R^t , 並進運動ベクトルを T^t とし, 時刻 $t+1$ と $t+2$ に撮影された画像から推定された, 時刻 $t+1$ から $t+2$ への右カメラの回転運動行列を R^{t+1} , 並進運動ベクトルを T^{t+1} とする (回転行列はその運動開始前の時刻 (t や

$t+1$) における右カメラ座標系で, 並進ベクトルはワールド座標系で表現されているとする). 各々の並進運動ベクトルの推定には定数倍の不定性が存在するので実際の並進ベクトルは, $k^t T^t, k^{t+1} T^{t+1}$ と表される. ここに k^t, k^{t+1} は, それぞれ時刻 $t, t+1$ におけるスケールを表す非零定数である. 並進ベクトルのスケールを整合させることは, k^t と k^{t+1} との比を求めることに相当する. 以下では, $k^t = 1$ としたときの k^{t+1} の値を k とし, k を求めることにする. なお, これらの時刻においては, 3次元空間中の点 Q (その同次座標は X) を注視点として右カメラを注視点制御しているとする.

以下では, 解析を簡単にするため, 時刻 t における右カメラ座標系がワールド座標系に一致しているとする (時刻 t における右カメラ座標系を基準に考える). このとき, 時刻 $t, t+1, t+2$ における右カメラの透視変換行列 P^t, P^{t+1}, P^{t+2} は, K を右カメラの内部パラメータを表す行列として,

$$\begin{aligned} P^t &= K \left[I \mid \mathbf{0} \right], \\ P^{t+1} &= K \left[R^t \mid -R^t T^t \right], \\ P^{t+2} &= K \left[R^{t+1} R^t \mid -R^{t+1} R^t T^t - R^{t+1} k T^{t+1} \right] \end{aligned}$$

とかける.

時刻 $t, t+1, t+2$ における点 Q の像の同次座標を, それぞれ, x^t, x^{t+1}, x^{t+2} とおくと,

$$\begin{bmatrix} P^t & x^t & 0 & 0 \\ P^{t+1} & 0 & x^{t+1} & 0 \\ P^{t+2} & 0 & 0 & x^{t+2} \end{bmatrix} \begin{bmatrix} X \\ -\lambda^t \\ -\lambda^{t+1} \\ -\lambda^{t+2} \end{bmatrix} = 0$$

が得られる. ここに $\lambda^t, \lambda^{t+1}, \lambda^{t+2}$ は非零定数である. 点 Q が存在するとき, $X \neq \mathbf{0}$ なので, 上式を $[X \quad -\lambda^t \quad -\lambda^{t+1} \quad -\lambda^{t+2}]^T$ に関する方程式とみなすと, 点 Q が存在するための必要十分条件は 9×7 の係数行列のランクが 6 以下になることである. すなわち, 9×7 の係数行列の任意の 7×7 の小行列を考えたとき, その行列式の値は零になる. この行列式には, $K, R^t, T^t, R^{t+1}, T^{t+1}, k, x^t, x^{t+1}, x^{t+2}$ が現れるが, このうち k のみが未知である. また, 行列式は k に関して 1 次式になっている. したがって, 線形演算で k を求めることができる.

以上のようにして, 異なる 2 フレーム対の間のスケールを整合させることができる. これを連鎖的に実行することによって, カメラ運動の並進ベクトルのスケールを運動全体にわたって整合させることができる. 最終的には, 運動全体にわたるスケールの不定性がひとつ存在することになる.