

局所的バンドル調整を用いた自己運動推定における累積誤差の軽減

池田 友彦[†] 杉本 晃宏^{††}

逐次型の運動推定手法では、運動が長くなると推定誤差が累積し、運動軌跡の推定精度が次第に悪くなるという問題をかかえている。そこで、本稿では、搭載した2台のアクティブカメラそれぞれを注視点制御することによりカメラの自己運動を推定する手法を例にとり、時々刻々に推定されたカメラ運動のパラメータがそれまでの入力画像系列と幾何学的に整合するように局所的にバンドル調整を適用し、推定結果をそのつど修正する手法を提案する。そして、(1) 局所的バンドル調整によって統計的に有意に誤差が軽減されること、(2) 局所的バンドル調整を3フレームに対して適用すると誤差を最も効果的に軽減することができること、(3) 提案手法によって長い運動に対しても安定な運動推定を実現することができること、を実験によって示す。

Reducing Accumulated Errors in Ego-motion Estimation Using Local Bundle Adjustment

TOMOHIKO IKEDA[†] and AKIHIRO SUGIMOTO^{††}

Incremental motion estimation methods involve a problem that estimation accuracy gradually becomes worse as the motion trajectory becomes longer and longer. This is due to accumulation of estimation errors incurred in each estimation step. To keep estimation accuracy stable even for a long trajectory, we propose to locally apply the bundle adjustment to each estimated motion so that the modified estimation becomes geometrically consistent with time-series frames acquired so far. To demonstrate the effectiveness of this approach, we employ an ego-motion estimation method using the binocular fixation control, and show that (i) our modification of estimation is statistically significant; (ii) in order to reduce estimation errors most effectively, three frames are optimal for applying the bundle adjustment; (iii) the proposed method is effective in the real situation, demonstrating drastic improvement of accuracy in estimation for a long motion trajectory.

1. はじめに

きたるべき装着型コンピューティング環境では、装着者の意図や興味を理解し、有益な情報を自律的に提供するシステム¹⁸⁾の実現が期待され、そのためには、装着者の位置や運動軌跡の情報を知ることが必要不可欠である^{1),5),16)}。また、自動清掃などを目的とした自律移動型ロボットでは、ロボット自身がどのような環境に存在し、どこに位置しているかを認識する技術が必要である^{7),20),28),35)}。特に、ロボットの自己位置姿勢をできる限り正確に同定する機能は、効率のかつ確実な作業遂行を実現するうえで最も基本的かつ重要である。このように、カメラを装着した人間やカメラを搭載したロボットの、3次元空間中における運動およ

び位置を推定する技術を確立することは、ますます重要になりつつある。

移動ロボットの自己位置推定は、自己位置と環境の幾何形状を同時に推定するSLAM (Simultaneous Localization and Mapping) とよばれている問題の中で扱われることが多い^{8),10),21),29)}。ここでは、エンコーダやジャイロ、加速度センサといった内界センサを用い、そのセンサから推定された移動変化量を時間積分することで総移動量を求めるという手法 (デッドレコーディング法) が一般的である³⁾。デッドレコーディング法における位置推定の誤差を軽減するために、カメラやレンジセンサといった外界センサを搭載し、内界センサと外界センサとを併用する手法も提案されている。この最も一般的な手法は、ロボットの自己位置を状態変数とした (拡張) カルマンフィルタを構成し、内界外界センサ両者による推定位置を誤差分散に応じて融合するものである^{14),15),22)}。

一方、装備軽減のためカメラのみを搭載し、カメラ

[†] 千葉大学大学院自然科学研究所

School of Science and Technology, Chiba University

^{††} 国立情報学研究所

National Institute of Informatics

からの入力映像を利用して自己位置を推定する手法も多く提案されている．ここでは，ステレオ視の枠組みを利用した手法が一般的である^{(6),(9),(19)}．具体的には，ステレオ視に基づいて3次元空間中の点を復元し運動前後での対応を求めたり，オプティカルフローを付加的に利用したりすることで運動推定を実現している．これに対し，近年，杉本ら^{(13),(25)~(27)}は，2台のアクティブカメラを独立に注視点制御する2眼独立注視点制御という考え方を導入し，搭載したカメラの基線長に依存しない推定精度を保証する運動推定手法を提案している．

しかし，これらの手法はすべて，時間的に連続する数フレームの画像を用いて逐次的に運動を推定しているため，運動が長くなると推定誤差が累積し，運動軌跡の推定精度が次第に悪くなるという問題をかかえている．これまで，ガウス分布などの確率分布による誤差モデルを導入し（拡張）カルマンフィルタによる状態遷移によって伝搬する誤差の分散を積算することで累積誤差を評価したり^{(2),(23),(33),(34)}，誤差に基づく推定の不確かさを確率モデルによって表現し，ベイズ推定を利用して確率的に累積誤差を論じたりする手法⁽¹¹⁾が提案されている．しかし，実際には，人間やロボットの運動が状態遷移モデルに従わない，誤差が，前提としている確率分布に従わない，などの場合が存在し，そのような場合これらのアプローチは破綻する．

一方，カメラパラメータのキャリブレーション⁽³²⁾では，投影モデルに基づいて推定されたパラメータの値に対して，再投影誤差（パラメータの推定値の下での特徴点の投影位置の予測と実際に観測される投影位置との差）を考え，それを最小化するように推定値を修正するというアプローチがとられている．特に，マルチカメラの外部パラメータのキャリブレーションの際には，順番に選ばれたカメラ対に対して外部パラメータを求め，最後に，すべてのカメラが全体として整合するようにパラメータ修正を行い，最終的な外部パラメータを求めるという手法が一般的である．このような再投影誤差の最小化は写真測量の分野で古くから行われており，バンドル調整^{(12),(31)}とよばれている．

バンドル調整は，幾何学的な拘束を用いて推定された誤差を補正することで累積誤差の軽減をはかるといふ考え方に基づいている．しかし，その補正は，パラメータの推定がすべて終わった後に，最後に一括して全パラメータの修正を実行するもので，バッチ処理となっている．したがって，バンドル調整は，本来，自己運動推定のような逐次的な処理とは相容れないものである．この矛盾を解消し，手法の逐次性を保持しつ

つバンドル調整によるパラメータ修正を行うために，本稿では，局所時間内に得られた推定結果にバンドル調整を適用する^{(36),(37)}という考え方を採用する．そして，推定された運動前後での画像間の特徴点の対応がエピソード拘束を満たすようにバンドル調整を局所的に適用し，運動パラメータを随時修正することで累積誤差の軽減をはかる．エピソード拘束は，運動に依存せず成立する制約であるため，本アプローチは人間やロボットの運動モデルを必要としない．

杉本ら^{(13),(25)~(27)}の2眼独立注視点制御では，カメラ制御を通じて運動前後で注視点の対応が自動的に得られているので，バンドル調整のために特徴点対応を求める新たな処理が不要であり，エピソード拘束を注視点の対応に課する格好の手法となっている．そこで，本稿では，2眼独立注視点制御を用いた運動推定手法に局所的バンドル調整を適用し，誤差が累積しないように推定結果を随時修正する手法を提案する．そして，(1)局所的バンドル調整によって統計的に有意に誤差が軽減されること，(2)局所的バンドル調整を3フレームに対して適用すると誤差を最も効果的に軽減することができること，(3)提案手法によって長い運動に対しても安定な運動推定を実現することができること，を実験によって示す．

2. 2眼独立注視点制御を用いた運動推定

2眼独立注視点制御^{(25),(26)}とは，2台の装着型アクティブカメラを独立に注視点制御するカメラ制御のことである（図1）．このカメラ制御によって，運動前後の画像間で注視点の対応が自動的に得られる．本稿をself-containedとするために，以下では，2眼独立注視点制御から得られる情報と注視点周辺のオプティカルフローから得られる情報とを統合することによってカメラの自己運動を推定する手法^{(13),(27)}を概観する．

左右2台のカメラのうち右カメラを基準に考え，2台のカメラを固定した台座の運動によって生じる右カメラの回転，並進運動をカメラ運動とよぶことにする．したがって，カメラ運動には，注視点制御のために行

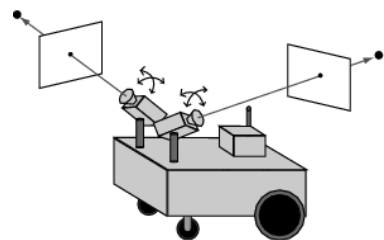


図1 2眼独立注視点制御

Fig. 1 Binocular independent fixation control.

うカメラの首振り運動は含まれないとする。左右のカメラのカメラ座標系を一致させるための回転行列、並進ベクトルを R_{in}, T_{in} とする。また、時刻 t から $t+1$ の間におけるカメラ運動のパラメータを R, T とする。前者が回転を表す行列であり、後者が並進を表すベクトルである。なお、カメラの内部・外部パラメータは校正済みであるとする。

時刻 t において、右カメラの投影中心から右カメラの注視点に向かう単位ベクトルは、時刻 t の右カメラ座標系で v_r^t であるとする。このとき、右カメラの注視点対称からカメラ運動のパラメータに関する制約式

$$\det [R_0 v_r^t \mid R_0 R v_r^{t+1} \mid T] = 0$$

が得られる。ここで、 R_0 はワールド座標系の姿勢を、時刻 t の右カメラ座標系に一致させるための回転行列である。同様に、左カメラの注視点対称からは、

$$\det [R_0 R_{in}^T v_\ell^t \mid R_0 R R_{in}^T v_\ell^{t+1} \mid T - R_0(R - I)R_{in}^T T_{in}] = 0$$

が得られる。ここで、 v_ℓ^t は、時刻 t において、左カメラの投影中心から左カメラの注視点に向かう単位ベクトルを表す（ただし、時刻 t の左カメラ座標系で表示）。

一方、時刻 t において、右カメラで撮影された画像の注視点周辺の領域内にある点 q_r^t のオプティカルフロー u_r^t が得られたとすると、それからカメラ運動パラメータに関する制約式

$$\det [R_0 R (M u_r^t + \tilde{q}_r^t) \mid R_0 \tilde{q}_r^t \mid T] = 0$$

が得られる。ただし、 f_r を右カメラの焦点距離として、

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}^T, \tilde{q}_r^t = ((q_r^t)^T, f_r)^T$$

とする。同様に、左カメラで画像の注視点周辺の領域内にある点 q_ℓ^t のオプティカルフロー u_ℓ^t が得られたとすると、それからカメラ運動パラメータに関する制約式

$$\det [R_0 R R_{in}^T (M u_\ell^t + \tilde{q}_\ell^t) \mid R_0 R_{in}^T \tilde{q}_\ell^t \mid T - R_0(R - I)R_{in}^T T_{in}] = 0$$

が得られる。ここに、 $\tilde{q}_\ell^t = ((q_\ell^t)^T, f_\ell)^T$ である (f_ℓ は左カメラの焦点距離)。

回転、並進ともに3自由度あるので、カメラ運動の自由度は6である。これに対し、注視点対称から得られるカメラ運動の制約式は2個である。また、オプティカルフローに基づく制約式は、1点に対して1個ずつ独立に得られる。したがって、4点以上に対して

オプティカルフローに基づく制約式を得ることで不足分の制約式を補い、それらをまとめて非線形最適化問題としてとらえ、最適化手法を適用することで、カメラ運動のパラメータを推定することができる。

3. 局所的バンドル調整

前章に述べた自己運動推定手法は、時間的に連続する2フレームを用いて逐次的に運動を推定している。したがって、運動の長さが長くなればなるほど誤差が累積し、運動軌跡の推定精度が次第に悪くなる。

推定精度を低下させる要因には、推定の誤りと推定誤差の累積との2種類の誤差があると考えられる。各時刻での運動推定には非線形最適化を行っているので、最適化の過程で局所解に陥る可能性があり、カメラ運動のパラメータを正しく推定できているとは限らない。これに起因する誤差が推定の誤りである。一方、たとえば大域的な最適解が求められたとしても、観測誤差に起因する推定誤差が生じている。しかし、いずれの誤差も引き起こす結果は同じである。すなわち、各画像の注視点とそれ以前に得られた画像系列の注視点とを推定された運動パラメータによって幾何学的に関係づけようとすると、そこに不整合が生じる。ここでは、この不整合を解消することによって推定精度を向上させることを目指す。具体的には、バンドル調整の考え方を取り入れて、この不整合が生じないように運動パラメータの推定値を随時修正することを考える。なお、以下では、2種類の誤差を区別して論じないことにする（以後、便宜上、すべて、推定誤差と記す）。なぜなら、2種類の誤差が相殺し、結果的に幾何学的な整合を保証し高精度な運動推定を実現する場合も考えられるので、それぞれの誤差を必ずしも個別に軽減する必要がないからである。

いま、離散的な時刻で画像系列が得られるとし、時刻 s から時刻 t ($t \geq s+2$) までのカメラ運動のパラメータが連続する2フレーム間で逐次的に求められているとする。そして、時刻 i ($i = s, \dots, t-1$) と時刻 $i+1$ の間で推定された運動パラメータを $R_{i,i+1}, T_{i,i+1}$ とする（図2）。ここに、 $R_{i,i+1}, T_{i,i+1}$ は、それぞれ、時刻 i から時刻 $i+1$ におけるカメラ運動の回転行列、並進ベクトルを表すとする。なお、注視点制御によって、時刻 s から時刻 t の間の各画像間で、

2種類の誤差を区別して議論するという立場も考えられるが、自己運動推定という目的に照らし合わせると、遅延なく高精度な推定が実現されればよいので、最終的には、処理の複雑化による計算コストの増加量とその処理の導入による精度向上の度合いとのトレードオフになると考えられる。

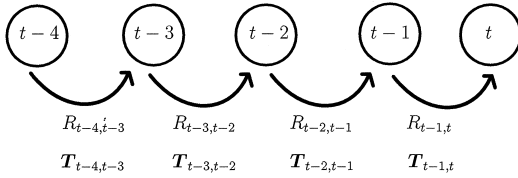


図2 隣接する時刻におけるカメラ運動のパラメータ

Fig. 2 Motion parameters over time-series camera motion.

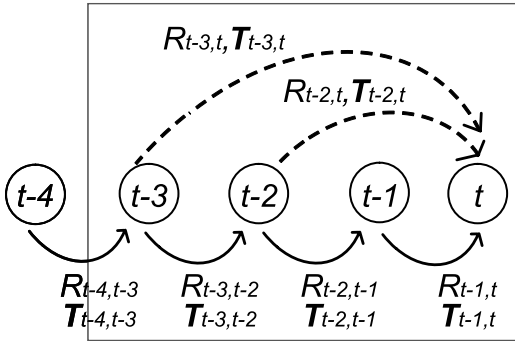


図3 局所的バンドル調整を適用する範囲内のカメラ運動のパラメータの関係 (4 フレームの場合)

Fig. 3 Relationship between motion parameters to which the local bundle adjustment is applied (the case of four frames).

注視点の像 p_i ($i = s, \dots, t$) の対応が得られているとする。

このとき、時刻 t の画像において、時刻 t を含む過去 n 時刻 ($3 \leq n \leq t - s + 1$) に得られた画像系列に注目する。そして、この画像系列に対してバンドル調整を適用することを考える。時間的に連続する 2 フレーム間のカメラ運動のパラメータが推定されているので、これを合成することによって、時刻 j ($t - n + 1 \leq j \leq t - 1$) から時刻 t までのカメラ運動 $R_{j,t}$, $T_{j,t}$ を得ることができる (図 3 参照):

$$R_{j,t} = \prod_{i=j}^{t-1} R_{i,i+1}, \quad T_{j,t} = \sum_{i=j}^{t-1} T_{i,i+1}.$$

そこで、これを用いて、 p_j によって定まるエピポーラ線を時刻 t の画像上に考える。本来、このエピポーラ線は p_t を通るはずであるが、推定誤差やその累積のため、図 4 に示す例のように、 p_t を通らない。これは、各画像の注視点とそれ以前に得られた画像系列の注視点とが幾何学的に不整合である、すなわち、エピポーラ拘束を満たさないことを意味する。

p_j , および、 $R_{j,t}$, $T_{j,t}$ によって得られる時刻 t の画像上のエピポーラ線は、時刻 t の画像上の点の同次座標を \tilde{x} とすると、

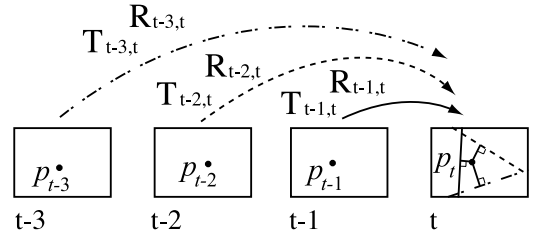


図4 注視点と推定したパラメータから得られるエピポーラ線

Fig. 4 Epipolar line obtained by the current fixation point and estimated parameters.

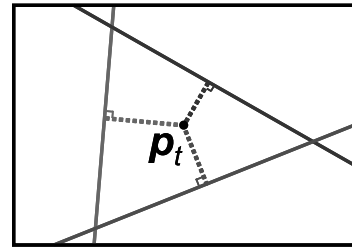


図5 注視点とエピポーラ線のずれの評価

Fig. 5 Distance between the fixation point and its corresponding epipolar line.

$$\tilde{x}^\top E_{j,t} \tilde{p}_j = 0 \quad (1)$$

と表される。ただし、 \tilde{p}_j は p_j の同次座標であり、

$$E_{j,t} = [T_{j,t}]_\times R_{j,t}$$

である。したがって、 \tilde{p}_t とエピポーラ線 (1) とのずれによって、 $R_{j,t}$, $T_{j,t}$ の誤差を評価することができる。そこで、 $(n-1)$ 枚の画像から得られるエピポーラ線群と \tilde{p}_t とのずれの和を評価し、それを最小化するように最新の運動パラメータ $R_{t-1,t}$, $T_{t-1,t}$ を修正する。幾何学的距離でエピポーラ線とのずれを評価 (図 5 参照) すると、最小化するべき評価関数は、

$$\sum_{j=t-n+1}^{t-1} \left(\frac{\tilde{p}_t^\top E_{j,t} \tilde{p}_j}{\sqrt{a_{j,t}^2 + b_{j,t}^2}} \right)^2 \quad (2)$$

となる。ただし、 $(a_{j,t}, b_{j,t}, c_{j,t})^\top = E_{j,t} \tilde{p}_j$ とした。

式 (2) を最小化するように運動パラメータ $R_{t-1,t}$, $T_{t-1,t}$ を修正することによって、過去 n 時刻に得られた画像系列において、注視点に関するエピポーラ拘束が満たされるようになる。ここで修正するパラメータは $R_{t-1,t}$, $T_{t-1,t}$ のみであるので、その数は n の値に依存しない。したがって、局所的なバンドル調整を適用する時間範囲 (フレーム数) を変更しても計算

$[T_{j,t}]_\times$ は、 $T_{j,t}$ から構成される 3×3 の歪対称行列である: 任意の 3 次元ベクトル y に対して、 $[T_{j,t}]_\times y = T_{j,t} \times y$ が成立する。

の手間はほとんど変わらないことになる。

以上のように時々刻々に推定された運動パラメータに対して、局所的にバンドル調整を適用し、そのつど推定結果を修正することによって、入力画像系列の注視点と推定された運動パラメータが幾何学的に整合するようになる。その結果、修正前の運動パラメータに含まれていた誤差が軽減され、逐次的な推定を行っても誤差の累積が抑えられ、推定精度の向上が期待される。なお、式(2)は運動パラメータ $R_{j,t}$, $T_{j,t}$ に関して非線形となっているので、Powell 法²⁴⁾ などの非線形最適化手法を用いることで評価関数の最小化を行うことができる。

4. 実験

提案手法を用いて、カメラの自己運動を推定する実験を行った。初めに、シミュレーションにより、提案手法を用いることでどの程度推定誤差が軽減されるかを定量的に評価した。また、何フレームに対して局所的バンドル調整を適用するのが運動推定に適しているかを評価した。次に、実環境下において、カメラの自己運動軌跡を推定する実験を行った。

4.1 シミュレーションによる定量的評価

ここでは、提案手法によって誤差がどの程度軽減されるかをシミュレーション画像を用いて定量的に評価した。比較のために杉本らの手法^{13),27)} (以下、従来手法とよぶことにする)を実装し、そこで生じる誤差に対する軽減の度合いを評価した。まず、前章において $n=3$ とした場合、すなわち、過去の3フレームに対して局所的バンドル調整を適用した場合において、提案手法と従来手法による運動パラメータの推定誤差を比較した。

カメラの基線長を距離の単位とした。また、カメラの画角は 21.0° とし、焦点距離を 0.025 とした。さらに、右カメラと左カメラの姿勢は一致しているとし、得られる画像のサイズを $512 \text{ pixel} \times 512 \text{ pixel}$ とした。一方、注視点の基線からの奥行きを 10.0 (一定) とした。

実験では、注視点間距離が 10.0 となるように左右のカメラの注視点を発生させた。さらに各画像中で、注視点の周りにオプティカルフローを算出する点を注

視点の周辺領域内に 3 個ずつランダムに発生させた。なお、注視点の周辺領域は、画像上で注視点の像を中心とする $20 \text{ pixel} \times 20 \text{ pixel}$ の範囲に投影される領域とした。

3 フレームの画像を得るために 2 ステップの運動を発生させた。第 1 ステップにおける並進ベクトルは、ノルムが 0.25 になるようにランダムに発生させ、回転行列は、回転角がカメラ座標系の Z 軸周りに 1.0° となるように発生させた。第 2 ステップの並進に関しては、第 1 ステップで発生させた並進ベクトルをカメラ座標系の Z 軸周りに $-1.0^\circ \sim 1.0^\circ$ の間でランダムに回転させた。また、第 2 ステップにおける回転に関しては、第 1 ステップにおけるカメラの回転角に $-1.0^\circ \sim 1.0^\circ$ の間でランダムに発生させた回転角を加えることで発生させた。

カメラ運動の前後において、各点をそれぞれの撮像面に投影し、その投影された点に対してノイズを加えた。ここでは、平均 0.0 pixel、標準偏差 2.0 pixel の正規分布に従って発生させた乱数をノイズとして加えた。そして、各ステップにおいて、カメラの運動の前後で画像上で各点の対応を算出し、カメラ運動の推定を行った。

推定されたカメラの回転 \hat{R} 、および、並進 \hat{T} の誤差は、以下の式で評価した。

$$\frac{\|\hat{\theta}\hat{r} - \theta r\|}{|\theta|}, \quad \frac{\|\hat{T} - T\|}{\|T\|}.$$

ここでは、誤差評価のため回転行列 \hat{R} を、回転軸 \hat{r} ($\|\hat{r}\|=1$) と回転角 $\hat{\theta}$ に分解した。 r ($\|r\|=1$) は回転軸の真値であり、 θ は回転角の真値である。また、 T は並進ベクトルの真値を表す。上記の過程を 200 回繰り返し、200 回の繰り返しにおける推定誤差の平均値と標準偏差を求めた。

図 6 にその実験結果を示す。これより、提案手法による推定誤差の軽減が読みとれる。まず、図 6(b) をみると、回転行列に関しては従来手法と提案手法とで推定誤差にあまり差異がないことが分かる。回転行列に関しては、もともと推定精度が十分高い^{13),27)} のので、それほど改善が得られなかったと考えられる。しかし、図 6(a) をみると、並進ベクトルに関して、提案手法を用いた方が推定誤差の平均値、および、標準偏差がともに小さくなっていることが分かる。この差に意味があるかどうかを判定するために、有意水準 5% でウェルチの検定³⁰⁾ を行ったところ、この差は統計的に有意であるとの結論を得た。

200 回のうちのある試行に対して、エピソード線を実際に引いた結果を図 7 に示す。図 7 から、修正に

ここでの実験の目的は、2 眼独立注視点制御を用いた運動推定手法の有効性を確認することではなく、提案手法による誤差の軽減の度合いを評価することである。したがって、簡単のため左右のカメラの姿勢を一致させており、また、注視点間距離も変化させず 10.0 (一定) とした。なお、注視点間距離を変化させた場合の実験においても、その結果は注視点間距離を 10.0 とした場合と同じような傾向であった。

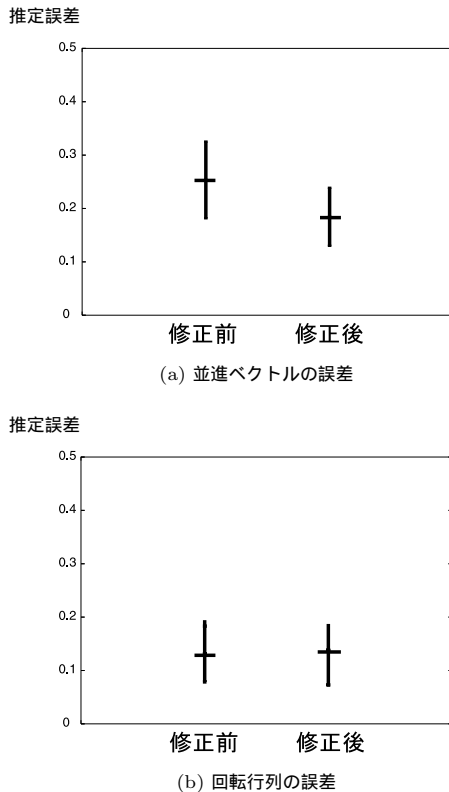


図 6 注視点間距離を 10 基線長とした場合の推定誤差

Fig. 6 Estimation errors in the case where the distance between fixation points is 10 baseline.

よって、過去のフレームから引かれたエピポーラ線を注視点を通るようになっていいることが分かる。このときの注視点と各エピポーラ線とのおよその距離は、修正の前後でそれぞれ、10 pixel, 0.2 pixel であった。200 回の試行における注視点とエピポーラ線との距離はほとんどの場合、この程度であった。

200 回すべての試行において、修正前後で評価関数の値が小さくなっていることを確認しているが、評価関数の値が小さくなるのが必ずしも推定値が真の値に近づくことを意味するとは限らないため（回転と並進での相殺が起こりうるため）、推定精度の改善にはある程度の限界があってしかるべきである。この点をふまえると、提案手法を用いることにより、逐次推定された運動パラメータが入力画像系列と幾何学的に整合し、推定精度は向上していると結論づけることができる。

次に、局所的バンドル調整を適用するフレーム数と推定誤差との関係を見るために、 $3 \leq n \leq 10$ の場合に対して、上と同一の手順で実験を行った。ここでは、10 ステップの運動を発生させた。なお、回転運動ではあまり改善がみられていないので、ここではカメラ

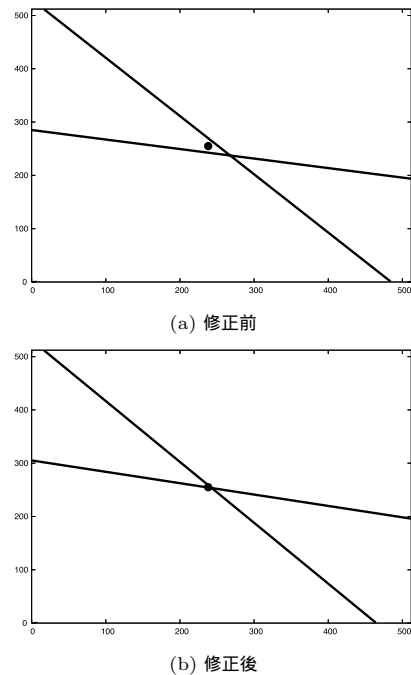


図 7 注視点と修正前後の推定パラメータから得られたエピポーラ線
Fig. 7 Fixation point and its corresponding epipolar line before/after parameter correction.

の並進運動の推定のみに着目した。このときの実験結果のうち、 $3 \leq n \leq 6$ の場合を図 8 に示す。それぞれの場合において、必要なフレーム数が得られるまではバンドル調整による修正は行っていない。たとえば、 $n = 5$ の場合は、4 ステップの運動以降に対してのみバンドル調整を適用している。

図 8 から、 $n = 3$ の場合、最も推定精度が良いことが分かる。一見、局所的バンドル調整を適用するフレーム数を増やせば増やすほど推定誤差の軽減がよりはかれるように思うが、実際は、そうではない。なぜなら、フレーム数が多くなると、バンドル調整を適用するまでの累積誤差が大きくなってしまっていて、バンドル調整を適用する拘束条件の信頼性が低下していると考えられるからである。それゆえ、バンドル調整ではもはやその誤差を十分に軽減できなくなってしまうと考えられる。 $n = 3$ の場合の誤差が最も小さいというのはこのことを裏づけている。累積誤差が大きくならないうちに随時バンドル調整を適用し、誤差を軽減する必要がある。

誤差を最も抑えるには、3 フレームに対して局所的バンドル調整を適用すればよいことが分かったので、 $n = 3$ の場合に対して、累積誤差がどのように変化するかを評価した。このときの実験の手順も同一である。この実験で発生させた運動は、11 ステップである。

推定誤差

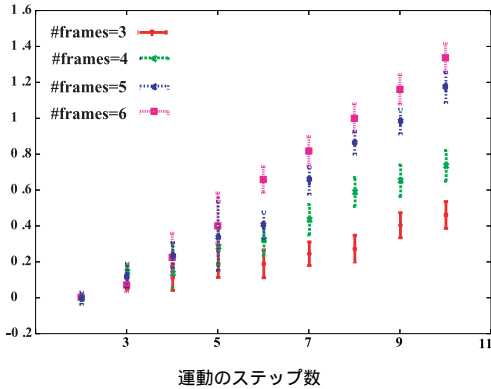


図 8 局所的バンドル調整を適用するフレーム数と並進運動の推定誤差

Fig. 8 The number of frames to which local bundle adjustment is applied and estimation errors.

図 9 に結果を示す．比較のため，従来手法で推定を行った場合についても結果を示した．回転運動に関しては，やはり，提案手法，従来手法ともに，あまり誤差の累積がみられない．しかし，並進運動に関しては，図 9 (a) から容易に読みとれるように，従来手法では運動のステップ数が増えるのにほぼ比例して誤差が累積しているのに対し，提案手法では誤差の累積の割合は大幅に改善されている．これより，提案手法によって，長い運動に対しても安定に運動推定を実現することができると思われる．

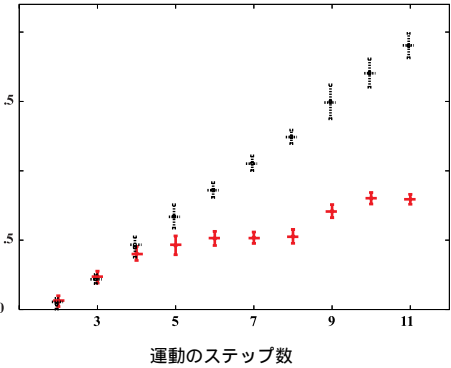
4.2 実画像を用いた運動軌跡推定

前節でのシミュレーション実験の結果をふまえて，実環境下において，提案手法を用いてカメラの自己運動軌跡を推定する実験を行った．また，提案手法の有効性を確認するため，従来手法との比較も行った．

本実験では，アクティブカメラとして SONY 製 EVI-G20 を用いた．そして，基線長が約 20 cm となるように 2 台のカメラを三脚に配置し，注視点間距離が 2~3 m 程度になるように，2 台のカメラの向きを設定した (図 10)．このとき，カメラ座標系の Z 軸は，カメラの光軸と平行であるとし，ワールド座標系は，初期位置における右カメラのカメラ座標系に一致しているとした．そして，右カメラの投影中心の運動軌跡を推定した．

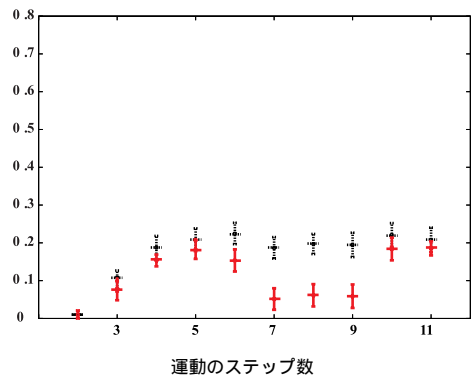
図 11 に示す軌跡 (約 5 m) に沿って実験室内でカメラを運動させ，提案手法，および，従来手法を用いて，その運動を推定した．実験では，図 12 に示したように，カメラ運動の軌跡を 130 ステップに刻んだ (2 つの直線部分がそれぞれ 20 ステップ，2 つの円弧の部分がそれぞれ 45 ステップである)．そして，各

推定誤差



(a) 並進ベクトル

推定誤差



(b) 回転行列

図 9 3 フレームを用いた局所的バンドル調整を適用した場合の累積誤差 (赤は提案手法，黒は従来手法を表す)

Fig. 9 Accumulation errors when the local bundle adjustment with 3 frames is applied (read means with correction and black means without correction).

ステップごとに提案手法，および，従来手法を適用した．なお，それぞれの手法では，計 6 点を左右のカメラの画像から均等に選び，そのオプティカルフローを用いた．

注視点制御における注視点の対応づけは手動で行った．また，運動の過程において，注視点画像から外れてしまった場合には注視点の更新が必要となるが，この更新は手動で行った (更新回数は左右のカメラともに 7 回であった)．一方，各カメラで得られた画像ごとに，注視点を中心とする 100 pixel × 100 pixel の

注視点の更新によって，幾何学的整合性を評価する特徴点が変わる．また，注視点の更新の影響による誤差も発生する．このため，厳密には，注視点の更新の前後で幾何学的整合性が必ずしも保証されるわけではない．この問題の影響をできるだけ排除するために，本実験では，注視点の更新を手動で行った．なお，この問題の解決は今後の課題として残されている．



図 10 実験時のカメラ配置
Fig. 10 Two camera setup.

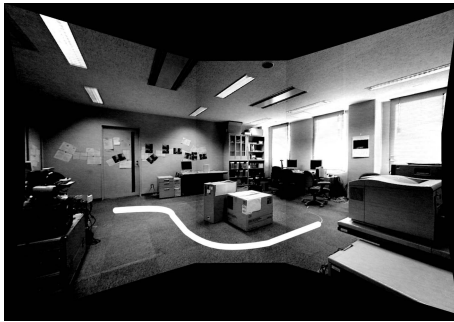


図 11 実験環境
Fig. 11 Experimental environment.

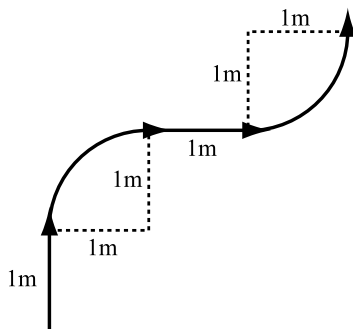


図 12 カメラ運動の軌跡
Fig. 12 Trajectory of the camera motion.

領域においてオプティカルフローを算出し^{4),17)}, さらに、注視点を中心とする $30 \text{ pixel} \times 30 \text{ pixel}$ の領域からランダムに 3 点ずつ選んだ。そして、運動推定には、この 6 点のオプティカルフローを利用した。また、提案手法では、時間的に連続する 3 フレームにおける左右のカメラそれぞれで、注視点対応を用いて局所的バンドル調整を行った。

提案手法、および、従来手法を利用してカメラ運動

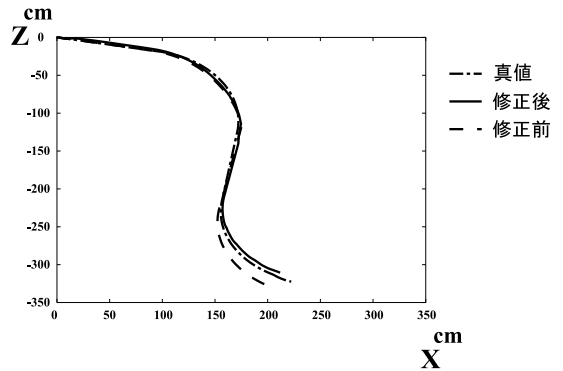


図 13 カメラ位置の推定結果
Fig. 13 Estimated camera positions.

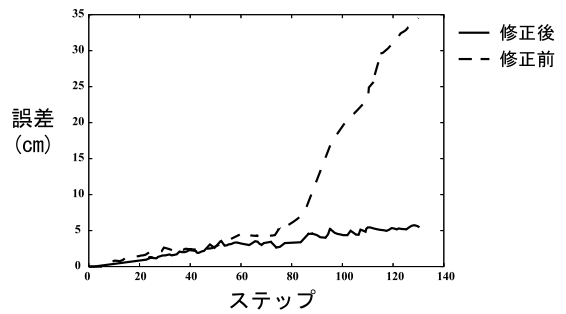


図 14 カメラ位置の推定誤差
Fig. 14 Estimation errors of the camera position.

表 1 130 ステップにおける推定誤差の平均と標準偏差
Table 1 The average and the standard deviation of estimation errors over 130 steps.

	修正前	修正後
平均 [cm]	0.268	0.0394
標準偏差 [cm]	0.102	0.00739

を推定した各ステップでの結果をつなぎ合わせることで、運動軌跡の推定結果を得た。得られたカメラの推定位置とその真値をワールド座標系の XZ 平面にそれぞれ投影し、重ね合わせた結果を図 13 に示す。また、提案手法および従来手法それぞれについて、各ステップにおけるカメラ位置の推定誤差を図 14 に示す。図 13、図 14 において、実線が提案手法を用いた結果であり、点線が従来手法を用いた結果である。また、1 点鎖線が真値を表している。表 1 に、130 ステップの運動における位置の推定誤差の平均と標準偏差を示す。

図 13 をみると、従来手法に比べ、提案手法を用いた方が真値に近い結果が得られていることが分かる。図 14 をみると、特に、80 ステップ目あたり（軌跡の 2 つ目の円弧部分）において、従来手法は推定に失敗して真の軌跡から大きなずれが生じはじめているが、

提案手法では、修正によって誤差が著しく減少し、軌跡推定に成功していることが分かる。従来手法では、直前のステップにおいて推定された運動パラメータの値を初期値として非線形最適化計算を実行している。このため、運動が変化するタイミングでは、推定に失敗することがある。実際、80 ステップ目あたりは、直線運動から回転運動に変化するタイミングである。提案手法によって推定の失敗が回避されたことは、バンドル調整は、失敗した推定結果を回復する効果があることを意味し、運動が変化するタイミングでも提案手法は精度良い運動推定を実現していると考えられる。最終的な推定位置の誤差は、従来手法では約 35 cm、提案手法では約 4 cm と、大幅な改善がみられる。表 1 から、提案手法の方が 130 ステップの位置推定における誤差の平均、および、標準偏差がともに小さくなっていて、高精度で安定に推定を行っていることが分かる。

5. おわりに

本稿では、逐次型の自己運動推定手法として 2 眼独立注視点制御を用いた運動推定手法^{13),25)–27)} を例にとり、バンドル調整を局所時間内に得られた運動パラメータの推定結果に適用し、そのつど推定結果を修正する手法を提案した。これによって、入力画像系列の注視点と運動パラメータが幾何学的な整合性を保つようになり、逐次的に推定された運動パラメータに含まれていた誤差が軽減された。実際、実験によって、(1) 局所的バンドル調整によって統計的に有意に誤差が軽減されること、(2) 局所的バンドル調整を 3 フレームに対して適用すると誤差を最も効果的に軽減することができること、(3) 提案手法によって長い運動に対しても安定な運動推定を実現することができることを示し、逐次型運動推定の精度向上を確認した。

本稿では、左右のカメラの注視点対称に対して局所的バンドル調整を適用したが、画像中のより多くの特徴点を用いて過去の画像系列との幾何学的な整合性を総合的に評価することで、さらなる推定精度の向上が期待される。どのような特徴点をどの程度用いるのがよいかなどに関しては、今後検討する予定である。また、局所的バンドル調整を選択的に適用するなどの工夫によって推定精度を維持しつつ処理の高速化をはかることも今後の課題として残されている。

バンドル調整は幾何学的な拘束を用いて推定誤差を補正するアプローチであるが、推定誤差の分布を導入し確率の枠組みに基づいて誤差を補正するアプローチと比べ、どのような場合にどちらのアプローチが適し

ているかを明らかにしたり、両アプローチの長所を活かした統合手法を議論したりすることも重要であると考えられる。

謝辞 本研究の一部は、科学研究費補助金 13224051、および、18049046 の補助を受けて行った。

参考文献

- 1) Aoki, H., Schiele, B. and Pentland, A.: Real-time Personal Positioning System for Wearable Computers, Vision and Modeling Technical Report, TR-520, Media Lab. MIT (2000).
- 2) Azarbayejani, A. and Pentland, A.: Recursive Estimation of Motion, Structure and Focal Length, *IEEE Trans. PAMI*, Vol.17, No.6, pp.545–561 (1994).
- 3) Barshan, B. and Durrant-Whyte, H.F.: Internal Navigation Systems for Mobile Robots, *IEEE Trans. RA*, Vol.11, No.3, pp.328–342 (1995).
- 4) Bouguet, J.-Y.: *Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm*, Intel Corporation, Microprocessor Research Labs, OpenCV Documents (1999).
- 5) Davison, A.J., Mayol, W.W. and Murray, D.W.: Real-Time Localization and Mapping with Wearable Active Vision, *Proc. IEEE/ACM Int. Symposium on Mixed and Augmented Reality*, pp.18–27 (2003).
- 6) Davison, A.J. and Murray, D.W.: Mobile Robot Localization using Active Vision, *Proc. ECCV*, Vol.2, pp.809–825 (1998).
- 7) DeSouza, G.N. and Kak, A.C.: Vision for Mobile Robot Navigation: A Survey, *IEEE Trans. PAMI*, Vol.24, No.2, pp.237–267 (2002).
- 8) Dissanayake, M.W.M.G., Newman, P., Clark, S., Durrant-Whyte, H.F. and Csorba, M.: A Solution to the Simultaneous Localization and Map Building (SLAM) Problem, *IEEE Trans. RA*, Vol.17, No.3, pp.229–241 (2001).
- 9) Gonçalves, N. and Araújo, H.: Estimation of 3D Motion from Stereo Images, *Proc. 16th International Conference on Pattern Recognition (ICPR)*, Vol.I, pp.335–338 (2002).
- 10) Guivant, J.E. and Neboit, E.: Optimization of the Simultaneous Localization and Map-Building Algorithm for Real-Time Implementation, *IEEE Trans. RA*, Vol.17, No.3, pp.242–257 (2001).
- 11) 原口一馬, 島田伸敬, 白井良明: 隠蔽を考慮した確率モデルによるロボットの移動, 第 18 回人工知能学会全国大会 (2004).
- 12) Hartley, R. and Zisserman, A.: *Multiple View*

- Geometry in Computer Vision*, Cambridge Univ. Press (2000).
- 13) 池田友彦, 杉本晃宏: アクティブカメラの注視点制御に基づく自己運動の推定, 電子情報通信学会論文誌, Vol.J88-D-II, No.8, pp.1480-1489 (2005).
 - 14) Kriegman, D.J., Triendl, E. and Binford, T.O.: Stereo Vision and Navigation in Buildings for Mobile Robot, *IEEE Trans. RA*, Vol.5, No.6, pp.792-803 (1989).
 - 15) 倉爪 亮, 長谷川勉: 全方位カメラとデッドレコニング機能を有するサッカーロボットのロバストな自己位置同定手法: メディアンフィルタによる観測誤差情報の除去と非線型最尤推定法の適用, 第20回ロボット学会学術講演会予稿集, 3A22 (2002).
 - 16) Kuorogi, M., Kurata, T. and Sakaue, K.: A Panorama-Based Method of Personal Positioning and Orientation and Its Real-Time Applications for Wearable Computers, *Proc. Int. Symposium on Wearable Computers*, pp.107-114 (2001).
 - 17) Lucas, D.D. and Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision, *Proc. Imaging Understanding Workshop*, pp.121-130 (1981).
 - 18) 松山隆司, 杉本晃宏, 佐藤洋一, 川嶋宏彰: 人間と共生する情報システムの実現を目指して, 人工知能学会誌, Vol.19, No.2, pp.257-266 (2004).
 - 19) Molton, N. and Brady, M.: Practical Structure and Motion from Stereo When motion is Unconstrained, *Int. J. of Computer Vision*, Vol.39, No.1, pp.5-23 (2003).
 - 20) Nakagawa, T., Okatani, T. and Deguchi, K.: Active Construction of 3D Map in Robot by Combining Motion and Perceived Images, *Proc. ACCV*, Vol.1, pp.563-568 (2004).
 - 21) Nagatani, K., Ishida, H., Yamanaka, S. and Tanaka, Y.: Three Dimensional Localization and Mapping for Mobile Robot in Disaster Environments, *Int. Conf. on Intelligent Robots and System*, pp.3112-3117 (2003).
 - 22) Nishizawa, T., Ohya, A. and Yata, S.: An Implementation of On-board Position Estimation for a Mobile Robot, *Proc. Int. Conf. on Robotics and Automation*, pp.395-400 (1995).
 - 23) 登内, 坪内, 有本: 移動ロボットにおける空間有限性を考慮した位置推定—内界情報センサと作業領域に関する知識のベイズ融合法, 日本ロボット学会誌, Vol.12, No.6, pp.695-699 (1994).
 - 24) Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T.: Numerical Recipes in C [日本語版], 技術評論社 (1993).
 - 25) 杉本晃宏, 長友 渉, 松山隆司: 装着型能動視覚センサを用いた人物の位置および運動の推定, 画像の認識・理解シンポジウム (MIRU2002), Vol.1, pp.215-222 (2002).
 - 26) Sugimoto, A., Nagatomo, W. and Matsuyama, T.: Estimating Ego Motion by Fixation Control of Mounted Active Cameras, *Proc. ACCV*, pp.67-72 (2004).
 - 27) Sugimoto, A. and Ikeda, T.: Diverging Viewing-Lines in Binocular Vision: A Method for Estimating Ego Motion by Mounted Active Cameras, *Proc. 5th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras*, pp.67-78 (2004).
 - 28) Sumi, Y., Ishiyama, Y. and Tomita, F.: 3D Localization of Moving Free-Form Objects in Cluttered Environments, *Proc. ACCV*, Vol.1, pp.43-48 (2004).
 - 29) Thrun, S.: Robotic Mapping: A Survey, *Exploring Artificial Intelligence in the New Millennium*, Morgan Kaufmann (2002).
 - 30) 東京大学教養学部統計学教室 (編): 統計学入門, 東京大学出版会 (1991).
 - 31) Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A.: Bundle Adjustment — A Modern Synthesis, Vision Algorithms: Theory and Practice, Triggs, B., Zisserman, A. and Szeliski, R. (Eds.), Lecture Notes in CS, Vol.1883, pp.298-372, Springer (2000).
 - 32) 植芝俊夫, 岡谷貴之, 佐藤智和: カメラキャリブレーション, 情報処理学会 CVIM 研究会, CVIM148, pp.1-18 (2005).
 - 33) 渡辺, 油田: 車輪型移動ロボットのデッドレコニングシステムにおける位置データ誤差, 第6回日本ロボット学会学術講演会, pp.347-350 (1988).
 - 34) Weng, J., Cui, Y. and Ahuja, N.: Transitory Image Sequences, Asymptotic Properties, and Estimation of Motion and Structure, *IEEE Trans. PAMI*, Vol.19, No.5, pp.451-463 (1997).
 - 35) Werman, M., Banerjee, S., Dutta Roy, S. and Qiu, M.: Robot Localization Using Uncalibrated Camera Invariants, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol.II, pp.353-359 (1999).
 - 36) Zhang, Z. and Shan, Y.: Incremental Motion Estimation through Local Bundle Adjustment, Technical Report MSR-TR-01-54, Microsoft Research (2001).
 - 37) Zhang, Z. and Shan, Y.: Incremental Motion Estimation through Modified Bundle Adjustment, *Proc. International Conference on Image Processing (ICIP'03)*, Vol.II, pp.343-346 (2003).

(平成 17 年 9 月 18 日受付)

(平成 18 年 3 月 20 日採録)

(担当編集委員 谷口 倫一郎)



池田 友彦

2003 年千葉大学工学部情報画像工学科卒業．2005 年同大学院自然科学研究科知能情報工学専攻修士課程修了．同年東レ株式会社入社．搭載型カメラを用いた自己運動推定の

研究に従事．



杉本 晃宏 (正会員)

1987 年東京大学工学部計数工学科卒業．1989 年同大学院工学系研究科修士課程修了 (数理工学専攻)．同年日立製作所基礎研究所に入社．1991 ~ 1995 年 ATR に出向．1999

年京都大学講師．2002 年国立情報学研究所助教授，2006 年同教授，現在に至る．総合研究大学院大学複合科学研究科教授併任．博士 (工学)．視覚情報処理や離散システム・アルゴリズム等に興味を持ち，数理的手法に基づいたコンピュータビジョンの研究に従事．2001 年情報処理学会論文賞．IEEE, 日本応用数理学会，電子情報通信学会等の会員．