

NAIST-IS-MT351203

修士論文

ステレオ動画像からの運動予測に基づく
物体の追跡

田村 洋

1995年 2月 20日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科において
修士(工学) 授与の要件として提出された修士論文である。

提出者： 田村 洋

指導教官： 横矢 直和 教授
千原 國宏 教授
荻原 剛志 助教授

ステレオ動画像からの運動予測に基づく 物体の追跡*

田村 洋

内容梗概

コンピュータビジョンにおけるシーンからの3次元情報の抽出手法の中で、両眼ステレオ法は観測対象へ影響を与えない、画像の入力が容易である、などの利点がある。しかし、ステレオ画像から3次元情報を抽出するための処理は計算量が多く、動画像をリアルタイムに処理することは困難である。

本論文では、ステレオ動画像から抽出した物体上の特徴点を効率良く追跡する手法を提案する。本手法では、ステレオ動画像から得られる物体の3次元空間における軌跡に基づいて次の移動位置を予測し、左右の2次元画像平面に微小探索空間を設定する。この予測に使用する特徴点の軌跡は、各フレームから獲得した特徴点の時系列位置からの多項式近似と、情報量基準 AIC 最小化原理に基づいた多項式の次数選択を逐次行なうことにより実現する。また、エピポーラ拘束に基づいたステレオ画像対応点探索とフレーム間の対応点探索を同時並行的に行ない、特徴点を連続的に追跡する。

提案手法を並列計算機上での分散処理として実現し、実画像を用いて人物の連続的な追跡を行なう実験を行ない、本手法の有効性を確認した。

キーワード

両眼ステレオ法, 動画像, 物体追跡, 多項式近似, 情報基準量, 相関係数, 運動予測

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT351203, 1995年2月20日.

Motion Tracking from Stereo Image Sequences Based on Motion Prediction*

Yoh Tamura

Abstract

Binocular stereo is a typical technique for obtaining 3D information of a scene in computer vision. This method has several advantages; for example, an image acquisition for the method is easy and it does not affect the scene. However, the stereo method is computationally expensive, especially in determining correspondences between stereo images. Therefore, real-time processing of dynamic stereo image sequences is very difficult.

This paper describes an efficient method for tracking feature points on an object in stereo image sequences. In this method, the locations of feature points in each frame are first predicted from the trajectories of the points obtained in the previous frames. These are predicted by using the polynomial approximation of their 3D trajectories. The degree of the polynomial is automatically selected based on the AIC minimization principle. In addition, the coordinates of feature points in the current frame are determined by cooperatively executing the stereo matching based on the epipolar constraint and the interframe matching.

The proposed method is implemented on a multi-processor system and has been successfully applied to stereo image sequences obtained in a real-world environment for human motion tracking.

Keywords:

binocular stereo, image sequences, motion tracking, information criterion, polynomial approximation, correlation coefficient, motion prediction

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT351203, February 20, 1995.

目次

1. 序論	1
2. 両眼ステレオ法による 3 次元情報の獲得	3
2.1 非接触 3 次元計測法	3
2.2 両眼ステレオ法の原理	4
2.3 ステレオ画像の対応づけ	5
2.4 カラー画像への対応	7
3. ステレオ動画像からの物体の追跡	9
3.1 動画像解析の基本原則	9
3.2 ステレオ動画像からの対応点探索	10
3.3 3 次元運動予測	12
3.3.1 3 次元運動予測による計算量の低減	12
3.3.2 軌跡の多項式近似による 3 次元運動の予測	14
3.3.3 情報基準量に基づく多項式の次数選択	14
3.4 物体の追跡処理手順	17
4. ステレオ動画像による実験	20
4.1 ステレオ動画像の入力	20
4.1.1 ステレオ静止画像入力装置	20
4.1.2 ステレオ動画像入力装置	20
4.2 実験プログラムの構成と実行環境	23
4.3 連続ステレオ静止画像による実験と評価	28
4.3.1 実験画像	28
4.3.2 ステレオ画像間とフレーム間の対応点探索による物体の追跡結果	29
4.3.3 軌跡の多項式近似における次数と AIC の関係	34
4.3.4 軌跡の多項式近似精度と探索窓の大きさの関係	38
4.3.5 対応点探索処理の計算時間	41
4.4 人物の実動画像による実験と評価	43
4.4.1 実験画像	43
4.4.2 特徴点追跡結果	43
4.4.3 軌跡の多項式近似精度と探索窓の大きさの関係	47
5. 結論	51

謝辞	52
参考文献	53

目 次

2.1	非接触 3 次元計測法の種類	3
2.2	両眼ステレオ法の原理	4
2.3	テンプレート・マッチングの図解	5
2.4	エピポーラ拘束による対応点の位置	7
3.1	動画像の概念図	10
3.2	現時刻位置を中心とした次時刻の対応点探索	11
3.3	ステレオ動画像対応点探索における相関係数の算出範囲	11
3.4	次時刻の 3 次元運動予測による探索窓の設定	13
3.5	観測値からの曲線近似	15
3.6	特徴点追跡処理の流れ	17
4.1	ステレオ静止画像入力装置の構成	21
4.2	ステレオ動画像入力装置の構成	22
4.3	インタレース画像の前処理方法	24
4.4	インタレース画像の前処理結果	25
4.5	実験プログラムの構成	26
4.6	プロセス間の制御構造	27
4.7	連続ステレオ静止画像の例	30
4.8	追跡物体上の特徴点位置	31
4.9	提案手法による特徴点の追跡結果	32
4.10	初期画像を基準画像に固定した場合の特徴点追跡結果	33
4.11	X 方向の多項式における次数と AIC 値の関係	35
4.12	Y 方向の多項式における次数と AIC 値の関係	36
4.13	Z 方向の多項式における次数と AIC 値の関係	37
4.14	探索窓の大きさと多項式近似に使用する時系列座標数による 予測位置と対応点のずれの割合	40
4.15	探索窓の大きさと軌跡の多項式近似用時系列座標数による 対応点探索計算時間	42
4.16	人物の実ステレオ動画像の例	44
4.17	追跡対象人物の特徴点位置	45
4.18	人物の追跡結果	46
4.19	探索窓の大きさと多項式近似に使用する時系列座標数による 予測位置と対応点のずれの割合	48
4.20	物体の回転の影響による誤対応	50

表 目 次

4.1	各軸の移動量	20
4.2	撮影条件	28
4.3	探索窓の大きさと多項式近似に使用する時系列座標数による追跡の成否	39
4.4	探索窓の大きさと多項式近似に使用する時系列座標数による誤対応の割合	39
4.5	撮影条件	43
4.6	探索窓の大きさと多項式近似に使用する時系列座標数による追跡の成否	47
4.7	探索窓の大きさと多項式近似に使用する時系列座標数による誤対応の割合	47

1. 序論

計算機の性能向上に伴い、画像からの3次元情報の抽出や、動画像を用いた移動物体の抽出・追跡などの画像処理の要求が高まっている。また、3次元情報を抽出する数多くの方法の中で、両眼ステレオ法は観測対象へ影響を与えない、画像の入力が容易であるなどの特徴を持ち、動画像への拡張も容易なことから、数多くの研究がなされている[1, 2, 3]。

画像から物体の運動を解析する方法は、特徴点の対応づけにより運動を算出する方法[4, 5]、微小運動からオプティカル・フローを求め運動を算出する方法[6, 7, 8]などが考えられている。動画像から物体の運動を推定する方法として、Horn は[4]、剛体上の複数の点(3点以上)が運動の前後で正しく対応付けられている場合に、4元数を用いて運動を推定できることを示した。また、金谷は[9]、3次元運動を計算するアルゴリズムを整理し批評した。

両眼ステレオ法における最も重要な問題はステレオ画像間における対応点の決定方法であり数多くの研究がなされている。Jenkin と Tsotsos は[10]、ステレオ画像と動画像の対応付けを同時に行ない、数多くの可能な対応付けの中から物理的に妥当な3次元移動ベクトルを与える対応付けを選択した。山本とチポラは[11]、左右のステレオ動画像からそれぞれ抽出された運動軌跡のステレオ対応付けが、エピポーラ拘束のみを用いて多重対応の数を減少できることを示した。阿部と横矢は[12]、ステレオ画像間の対応付けに動的輪郭モデルを導入し、雑音に対しロバストな対応点探索方法を示した。また、大崎らは[13]、モデル当てはめと運動推定を交互に行なうことによりロバストな追跡を行なった。

ステレオ画像から3次元情報を抽出するための処理は計算量が多く、動画像をリアルタイムに処理することは困難な問題である。なかでもステレオ画像間における対応点探索問題は、非常に多くの処理時間が掛かることから、ステレオ動画像をリアルタイムに処理するボトルネックになっている。動画像が物体の動きに比べて時間的に密に得られるとすれば、動画像間の対応点探索は近傍処理のみで良くなり、処理時間を減少させることが可能である。また、探索範囲を狭くできることから誤対応の可能性も減少し、より簡単なアルゴリズムで対応点を見つけることが

可能になる。しかし、一般に利用できる動画像は 1/30 秒程度の間隔で入力されるため、人間や車両などの動物体を対象とする場合には、十分に短い間隔であるとは言えない。このため探索空間を狭くする手法の導入が必要になる。探索空間を狭くするための一手法として、解析対象の対応点がどのように動いているか推定し次画像中に現われる位置を予測することが考えられる[14]。

本論文では、移動物体の実時間追跡を目的として、ステレオ動画像から抽出した物体上の特徴点を効率良く追跡する手法を提案する。本手法では、ステレオ動画像から得られる 3 次元空間の軌跡を多項式に近似することにより次画像上の移動位置を予測し、左右の 2 次元画像平面に投影した狭い探索空間内において対応点を決定してマッチング処理を減少させ、処理の高速化を図る。この予測に使用する特徴点の軌跡は、各フレームから獲得した特徴点の時系列位置からの多項式近似と、情報量基準 AIC 最小化原理に基づいた多項式の次数選択を逐次行なうことより実現する。また、フレーム間の対応点探索とエピポーラ拘束に基づいたステレオ画像間対応点探索に相関係数を用いたマッチング処理を適用し、これらを同時並行的に処理することによりステレオ動画像の対応点探索を行なう。

提案手法を並列計算機上での分散処理として実装し、連続静止画像と人物の実動画像を用いた移動物体上の特徴点の連続的な追跡を行なう実験を行ない、高速な追跡が可能であることを示す。

以下、第 2 章では、提案手法の基礎となる両眼ステレオ法による 3 次元情報の獲得方法について、第 3 章では、提案手法の説明と実装する際の処理の流れについて、第 4 章では、実験環境と実画像を用いた物体上の特徴点の追跡実験について、第 5 章では、本論文のまとめとして、結論と今後の課題について述べる。

2. 両眼ステレオ法による 3 次元情報の獲得

この章では、本研究の基盤となる両眼ステレオ法を用いた 3 次元情報獲得の基本原理について述べる。

2.1 非接触 3 次元計測法

シーンの 3 次元情報を獲得する方法は図 2.1 に示すような能動的手法と受動的手法がある[15]。能動的計測は、対象とするシーンに何らかのエネルギーを照射し、その反射を測定することにより対象までの距離を測定する方法である。受動的計測は、対象に対して何ら影響を与えることなく、通常画像から観測系の工夫と対象に関する制約条件に基づいて、画像生成過程の逆問題を解くことにより 3 次元情報を得る方法である。

本研究に用いた両眼ステレオ法は、受動的手法に分類される。両眼ステレオ法は、画像入力が容易、自然光での測定が可能などの長所があるが、従来方法では計算量が多い、測定精度が低いなどの欠点がある。

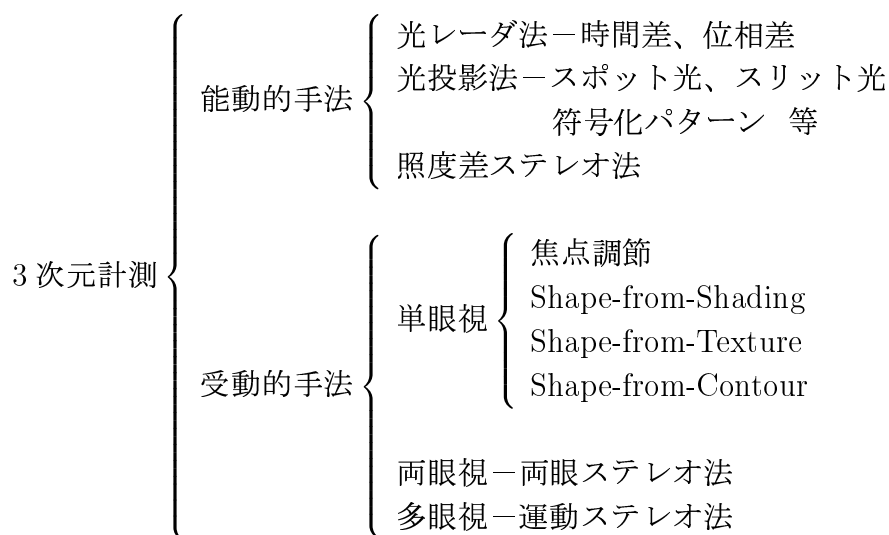


図 2.1 非接触 3 次元計測法の種類

2.2 両眼ステレオ法の原理

両眼ステレオ法は、カメラなどの視覚センサーを左右に2台並べ両眼立体視により3次元情報を獲得する方法であり、もっとも一般的な距離測定法として用いられてきた三角測量法の原理に基づいている。

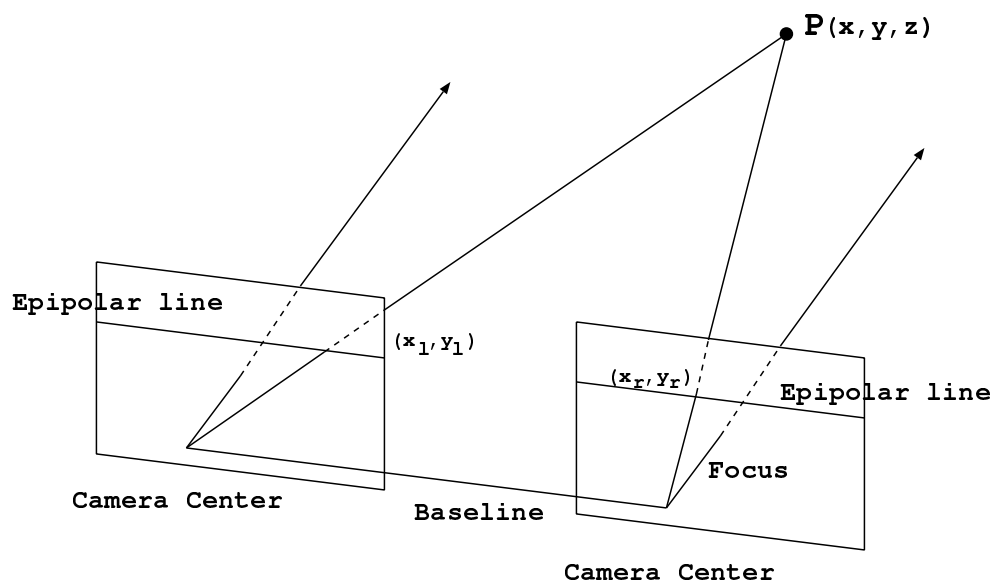


図 2.2 両眼ステレオ法の原理

図 2.2 に示すような、左右のカメラの光軸が平行な標準ステレオ・カメラ・モデルを考える。シーンはそれぞれの画像平面に焦点距離 (focus) で中心投影される。左右投影中心を結ぶ線分を基線 (baseline)、この基線を含む任意の平面をエピポーラ面、この平面と各投影面との交線をエピポーラ線と呼ぶ。

ここで、空間中の点 P の座標を2つのレンズ中心の中点を原点として定め、 X 軸を基線と平行な方向、 Y 軸を基線と光軸に垂直な方向、 Z 軸を光軸方向とした座標で表す。点 P の画像平面上の像がそれぞれ $(x_l, y_l), (x_r, y_r)$ で検出されたとすると、点 P の座標値 (x, y, z) は以下の式で与えられる [16]。

$$x = baseline \cdot \frac{(x_l + x_r)/2}{x_l - x_r} \quad (2.1)$$

$$y = baseline \cdot \frac{(y_l + y_r)/2}{x_l - x_r} \quad (2.2)$$

$$z = baseline \cdot \frac{focus}{x_l - x_r} \quad (2.3)$$

2.3 ステレオ画像の対応づけ

両眼ステレオ法により 3次元位置を算出するためには、基線長や焦点距離を既知とすれば、右の画像の中に見えている点が左の画像のどこに見えているか探し出す、対応点決定の方法が重要である。

任意のデータから、与えられた特定のパターンと一致する部分を見つけ出すことを一般的にパターン・マッチングと言ひ、対象パターンと参照パターンがともに画像の形で表現されている場合、テンプレート・マッチングと呼ばれる[17]。

テンプレート・マッチングは、図 2.3 に示すように、既知の対象を表すテンプレート $I_t(x, y)$ を動かしながら画像 $I_f(x, y)$ の各点での相互相関を求め、それが最大となる点をマッチング画像とする操作である。

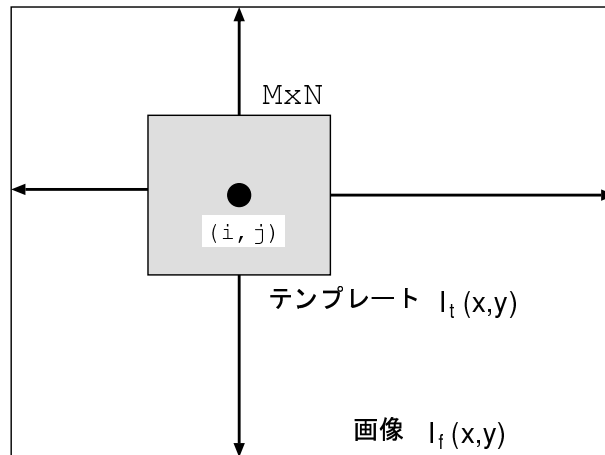


図 2.3 テンプレート・マッチングの図解

マッチング処理を行なう小領域の大きさを $M \times N$ の矩形領域とし、その内部の各点の明るさをそれぞれ $I_f(i, j)$, $I_t(i, j)$ とすると、これらの小領域の相互相関は次式で与えられる。 $C(f, t; i, j)$ は統計における相関係数に対応した量である。

$$C(f, t; i, j) = \frac{\sum_{j=1}^N \sum_{i=1}^M ((I_f(i, j) - \mu_f)(I_t(i, j) - \mu_t))}{\sqrt{\sigma_f^2 \sigma_t^2}} \quad (2.4)$$

ただし、

$$\begin{aligned} \mu_f &= \frac{1}{MN} \sum_{j=1}^N \sum_{i=1}^M I_f(i, j) \\ \mu_t &= \frac{1}{MN} \sum_{j=1}^N \sum_{i=1}^M I_t(i, j) \\ \sigma_f^2 &= \sum_{j=1}^N \sum_{i=1}^M \{I_f(i, j) - \mu_f\}^2 \\ \sigma_t^2 &= \sum_{j=1}^N \sum_{i=1}^M \{I_t(i, j) - \mu_t\}^2 \end{aligned}$$

標準ステレオ・カメラの配置から、左画像上で見えるシーン中の点の位置は、右画像上で見える位置よりも必ず右方向にあり、対応するエピポーラ線上にある (図 2.4 参照)。これがエピポーラ拘束である。つまり、一方の画像上の 1 点に対応した他方の画像上の点は、エピポーラ線上の小領域についてテンプレート・マッチングにより探すことができる。つまり、エピポーラ線に沿って式 (2.4) を計算し、この値が最大となる部分に対応点とする。

ステレオ画像からの対応点探索は、相互相関を用いるこの方法が最も確実と思われることから、本論文で使用するが、雑音の影響を受け易い、計算コストが大きいなどの欠点がある。計算コストを減少するために次のような方法が提案されている [17, 18]。

- 粗精探索法

ピラミッド画像を用いて画像内での探索回数を減らす方法である。粗い画像のすべての点についてマッチング尺度を計算し、最も一致

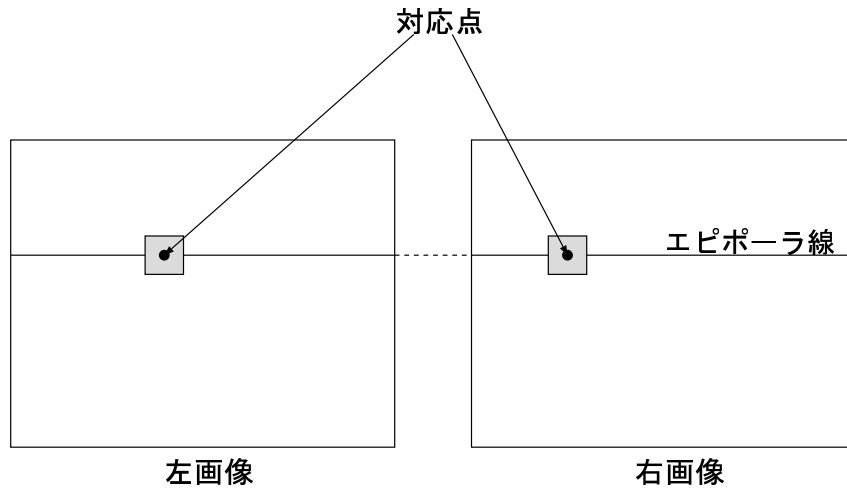


図 2.4 エピポーラ拘束による対応点の位置

する点を見つける。次に、精画像では、粗画像で見つかった点に対応した点の近傍についてだけマッチングの計算を行なう。これにより、全体としての計算回数を減らすことができる。

- 残差逐次検定法

SSDA (Sequential Similarity Detection Algorithm) と呼ばれ、小領域間の対応する画素値の差を累積し、残差が最小になる領域を一致したとする。非対応領域では、残差が閾値を越えた時点で処理を打ち切りことができるため、計算時間を大幅に改善できる特徴がある。

2.4 カラー画像への対応

カラー画像を利用することによりマッチング精度が向上することが報告されている[19]ことから、本論文では相関係数を基本としたマッチング処理をカラー画像に対して行なう。

RGB 各色の階調をそれぞれグレースケール画像における輝度値に対応するものとする。 (i, j) における色 Q ($Q = R, G, B$) の階調を $Q(i, j)$

とすると、式 (2.4) は次のようになる。

$$C_{RGB}(f, t; i, j) = \frac{\sum_{j=1}^N \sum_{i=1}^M \sum_{Q=R,G,B} ((Q_f(i, j) - \mu_{Q_f})(Q_t(i, j) - \mu_{Q_t}))}{\sqrt{\sigma_{Q_f}^2 \sigma_{Q_t}^2}} \quad (2.5)$$

ただし、

$$\begin{aligned} \mu_{Q_f} &= \frac{1}{3MN} \sum_{j=1}^N \sum_{i=1}^M \sum_{Q=R,G,B} Q_f(i, j) \\ \mu_{Q_t} &= \frac{1}{3MN} \sum_{j=1}^N \sum_{i=1}^M \sum_{Q=R,G,B} Q_t(i, j) \\ \sigma_{Q_f}^2 &= \sum_{j=1}^N \sum_{i=1}^M \sum_{Q=R,G,B} \{Q_f(i, j) - \mu_{Q_f}\}^2 \\ \sigma_{Q_t}^2 &= \sum_{j=1}^N \sum_{i=1}^M \sum_{Q=R,G,B} \{Q_t(i, j) - \mu_{Q_t}\}^2 \end{aligned}$$

3. ステレオ動画像からの物体の追跡

本章では、図 2.2 に示した視線の平行な 2 台の等価なカメラからなるステレオ撮像系を用いた動画像から、効率良く物体を追跡する方法を提案する。

以下の順番で提案手法を説明する。

(1) 動画像解析の基本原理

(2) ステレオ動画像からの対応点探索

フレーム間対応点探索とステレオ画像間対応点探索の同時実行によるステレオ動画像からの対応点探索方法を提案する。

(3) 軌跡の多項式近似による対応点探索範囲の縮小

対応点探索範囲が広いと多くの演算時間が必要になる。そこで、軌跡の多項式近似を応用した対応点の予測を行ない、対応点探索範囲を狭くする方法を提案する。

(4) 多項式の次数選択

対応点の予測を行なうためには、軌跡の特徴が最も反映された多項式の次数を決める必要がある。そこで、情報基準量 AIC 最小化原理に基づく多項式の次数選択を導入する。

(5) 提案手法を実装するための処理の流れ

3.1 動画像解析の基本原理

1 台のカメラで撮影された時系列画像は、図 3.1 に示すような連続した静止画像として表現できる。

物体の動きは、一般に非定常であるが、空間的に狭い領域で対象とする画像中の物体の形状および速度が一定であるとすれば、短時間であれば定常と見なすことができる。

したがって、時刻 t における画像上の点 P が Δt 時間経過後 $P + \Delta P$ に移動する場合、時刻 t の画像上の点 P と、時刻 $t + \Delta t$ の画像における

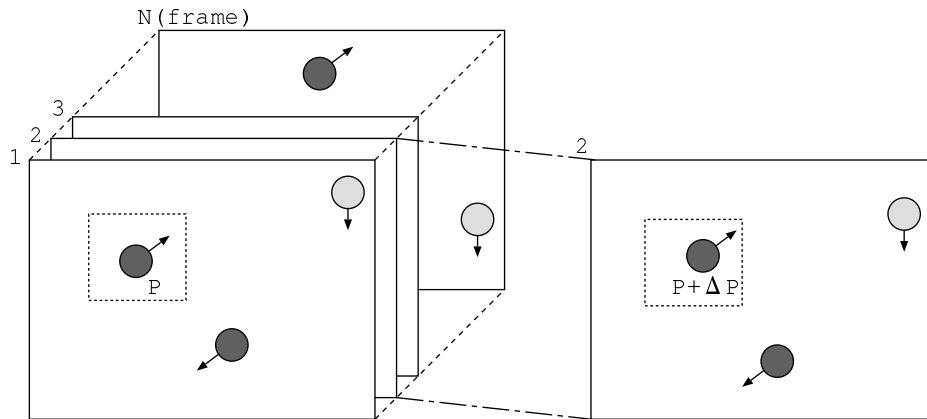


図 3.1 動画像の概念図

点 P を中心とした小領域間の相関を求めることにより $P + \Delta P$ の位置を探索できる。

3.2 ステレオ動画像からの対応点探索

図 3.2 に示すように、時刻 t に P_t にあった点が、時刻 $t+1$ に P_{t+1} に移動したとする。この点の移動は、左右の画像上でそれぞれ P_l から P_{l+1} 、 P_r から P_{r+1} への移動として観察することができる。

したがって、時刻 t における画像上の位置を中心に、時刻 $t+1$ の画像上で探索窓 W_l 、 W_r を設定し探索を行なうことにより画像上の位置を特定できる。

しかし、左右の画像で独立に対応点の探索を実行すると、探索で得られた対応点はエピポーラ拘束に従っていない可能性がある。このため、ステレオ動画像で特徴点の追跡を行なうためには、時系列画像の同一性を保証するフレーム間での類似度と、ステレオ画像の拘束条件を満たす左右画像間の類似度から対応点を決める必要がある。すなわち、フレーム間対応点探索とステレオ画像間対応点探索を同時に行わなければならない。

そこで、以下の 3 種類の相関係数を求め、式 (3.1) に示す相関係数の和 $C(x_l, x_r, y)$ が最大になる左右の点 (x_l, y) 、 (x_r, y) を次フレームの対応点とすることで、ステレオ画像間対応点とフレーム間対応点を決定する (図 3.3 参照)。

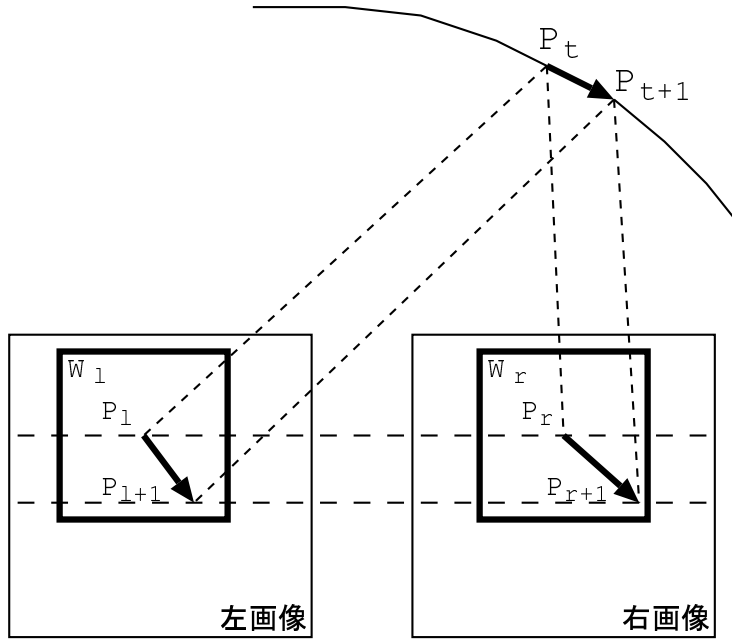


図 3.2 現時刻位置を中心とした次時刻の対応点探索

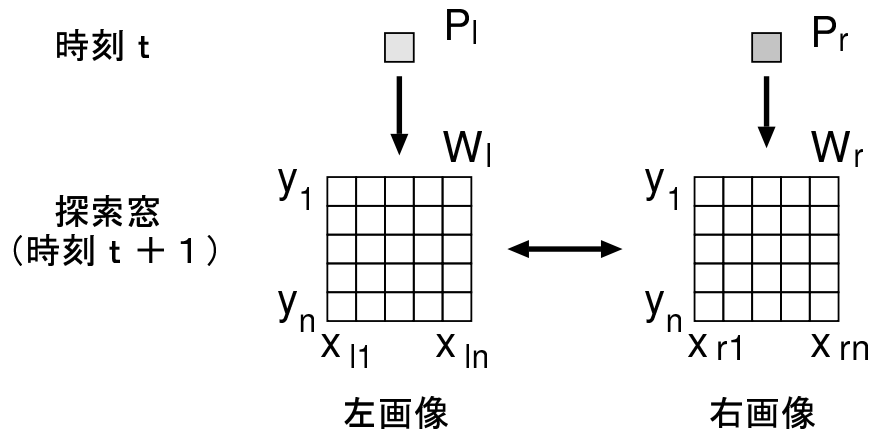


図 3.3 ステレオ動画像対応点探索における相関係数の算出範囲

- 左画像内の特徴点 P_l と左探索窓 W_l 内の各点 (x, y) とのフレーム間相関係数 : $C_l(x, y)$
- 右画像内の特徴点 P_r と右探索窓 W_r 内の各点 (x, y) とのフレーム間相関係数 : $C_r(x, y)$
- エピポーラ拘束を満たす探索窓 W_l 内の点 (x_l, y) と W_r 内の点 (x_r, y) とのステレオ画像間相関係数 : $C_{rl}(x_l, x_r, y)$

$$C(x_l, x_r, y) = C_l(x_l, y) + C_r(x_r, y) + C_{rl}(x_l, x_r, y) \quad (3.1)$$

画像上で見える動物体の大きさ・方向は刻一刻と変化する。したがって、動画像を使用した連続追跡では、式 (3.1) が最大になる左右画像上の点を次の探索における基準画像にする必要がある。

3.3 3次元運動予測

3.3.1 3次元運動予測による計算量の低減

現在の画像上で対象物体の位置が判っている場合、次フレームでは現在位置の近傍に比較的大きな探索窓を設定し探索を行えば対応点を見つけることが可能である。しかし、探索窓 W の大きさを $n \times n$ 画素とする場合、相関係数の演算回数は、

- $C_l(x, y)$ 演算回数
探索窓の画素数 : n^2 回
- $C_r(x, y)$ 演算回数
探索窓の画素数 : n^2 回
- $C_{rl}(x_l, x_r, y)$ 演算回数
(エピポーラ線上の画素数)² × (エピポーラ線数) : n^3 回

であるから合計 $2n^2 + n^3$ 回となり、探索窓が広がると演算回数が急激に増加し、処理性能が大幅に低下することを示している。このことは、動

画像のリアルタイム処理にとって大きな問題であり、探索窓を小さくするアルゴリズムの導入が必要である。

探索窓を小さくする方法として対象物体の運動を予測し、予測した移動位置近傍の狭い範囲だけ探索を行なうことにより、相関係数の演算回数を低減する方法が考えられる。つまり、対象物体の3次元運動を推定し、次画像フレーム上での特徴点の3次元予測位置を求め、左右の2次元画像上に投影した位置の近傍に小さな探索窓を設定し、式(3.1)が最大になる点を探査する(図3.4参照)。

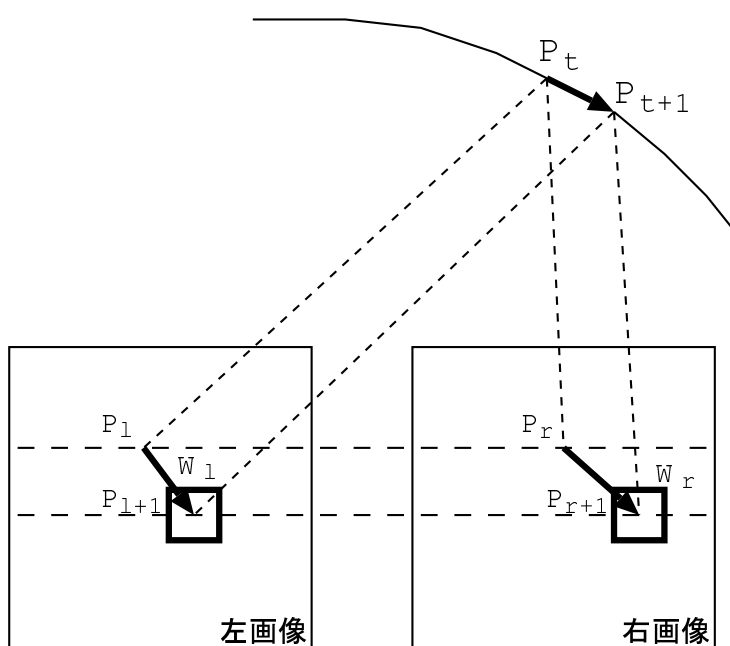


図 3.4 次時刻の3次元運動予測による探索窓の設定

3次元位置から2次元画像上への投影式は、式(2.1)～(2.3)より、以下の式で定義される。

$$x_r = \frac{(x - baseline/2) \cdot focus}{z} \quad (3.2)$$

$$x_l = \frac{(x + baseline/2) \cdot focus}{z} \quad (3.3)$$

$$y_r = y_l = \frac{\text{focus} \cdot y}{z} \quad (3.4)$$

3.3.2 軌跡の多項式近似による 3 次元運動の予測

すでに述べたように、物体の動きは短時間であれば直線で近似することができる。しかし、実際の動画像は 1/30 秒間隔で撮影され、この間隔は人間や車両などの動物体の動きを対象とする場合には、十分に短い間隔であるとは言えない。つまり、2 枚の画像フレームから求まる移動量から次の移動位置を単純に求めることはできない。そこで、本手法では、フレーム間隔が一定であることを利用して軌跡の多項式近似を行ない次フレーム上の位置を推定する。

時刻 t における特徴点の 3 次元空間内での位置を $P(x_t, y_t, z_t)$ とし、ステレオ動画像の各フレームから特徴点の 3 次元位置を時系列的に求めたとすれば、これらの点の軌跡を時間の関数として次の k 次多項式で近似することは最小 2 乗法を用いて容易に行なえる。

$$x_t = a_{x0} + a_{x1}t + a_{x2}t^2 + \cdots + a_{xk}t^k \quad (3.5)$$

$$y_t = a_{y0} + a_{y1}t + a_{y2}t^2 + \cdots + a_{yk}t^k \quad (3.6)$$

$$z_t = a_{z0} + a_{z1}t + a_{z2}t^2 + \cdots + a_{zk}t^k \quad (3.7)$$

したがって、次フレームのステレオ画像上に現れる特徴点の 3 次元座標は、最新の t フレームまでの座標で式 (3.5) ~ (3.7) の係数を求めたとき、フレーム番号 $t+1$ を与えることにより予測することができる。

3.3.3 情報基準量に基づく多項式の次数選択

軌跡の多項式近似を行なう上で問題となるのは、式の次数をどのように設定するかである。図 3.5 に示すような観測値が得られた場合、図に示すような 1 次 ~ 4 次式以上の曲線で近似できたとしても、どの曲線がもっとも良い推定値を示すか判断できない。

最小 2 乗法の立場から見ると、残差 2 乗平均が小さくなるほど曲線のモデルは良く合っていることになる。しかし、次数を大きくすると、デー

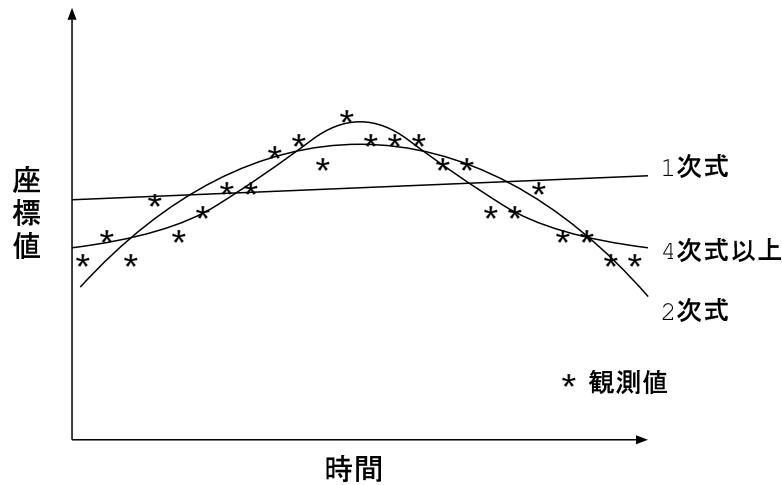


図 3.5 観測値からの曲線近似

タへの見かけ上の当てはまりは良くなるが、データに含まれる雑音に過剰に反応するため、安定したモデルを得ることができない。逆に、次数が小さいと対象の特性を十分に反映できない状態になる。したがって、モデルを選択するための客観的な判定基準が必要である。この判定基準として、情報量基準 AIC (Akaike's Information Criterion) [20] を導入する。

AIC は、現実に観測されるデータを用いて統計的モデルの悪さを評価するために導入された統計量であり次式で定義される。

$$AIC = (-2)(\text{最大対数尤度}) + 2(\text{パラメータ数}) \quad (3.8)$$

ただし、パラメータ数はモデル内にあって自由に変化させられるものの数である。

比較するモデル間の差が大きいときは、右辺第 1 項に著しい差が現われる。また、第 1 項にあまり差が無い場合は、第 2 項が作用してパラメータの少ない方が良いモデルと評価される。AIC の値の絶対的な大きさはデータに依存するが、2 つ以上のモデルの AIC の値の差がモデルの相対的な良さの指標となり、AIC が小さい方が良いモデルということになる。したがって、この値が最小となる多項式のモデルを最良と判断し、その多項式の次数選択を行ない運動予測に適用する。

本論文では、観測された特徴点の座標から式 (3.5) ~ (3.7) を最小 2 乗法を用いて近似する。最小 2 乗法におけるモデルパラメータの最尤推

定値は χ^2 分布を最小にすることによって得られる[21]。例えば、

$$y(x) = a_0 + a_1x + a_2x^2 + \dots + a_kx^k \quad (3.9)$$

の k 次の多項式を考えた場合、モデルの一般形は

$$y(x) = \sum_{l=0}^k a_l X_l(x) \quad (3.10)$$

で表される。このとき、 χ^2 分布は次のようになる。

$$\chi^2 = \sum_{i=1}^N \left[\frac{y_i - \sum_{l=0}^k a_l X_l(x_i)}{\sigma_i} \right]^2 \quad (3.11)$$

σ_i は i 番目のデータの測定誤差である。

式 (3.11) が最小になるのは、 χ^2 を $k+1$ 個のパラメータ a_l で微分した値が全て 0 になるときであり、次の方程式が与えられる。

$$0 = \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[y_i - \sum_{j=0}^k a_j X_j(x_i) \right] X_l(x_i) \quad l = 0, \dots, k \quad (3.12)$$

この連立方程式を解くことにより解が得られることから、誤差が最小になるときに尤度は最大になると言える。

また、最小 2 乗法で求めた k 次の多項式の残差の 2 乗平均 S_k^2 は

$$S_k^2 = \frac{1}{N} \sum_{i=1}^N \left[y_i - \sum_{j=0}^k a_j X_j(x_i) \right]^2 \quad (3.13)$$

であるから、求めた多項式の最大対数尤度は次のようになる。

$$\xi = N \log_e \frac{1}{\sqrt{S_k^2}} \quad (3.14)$$

この式を、式 (3.8) に代入すると k 次式への近似における AIC の計算式は

$$\begin{aligned} AIC &= -2 \times (-N \log_e \sqrt{S_k^2}) + 2(k+1) \\ &= N \log_e S_k^2 + 2(k+1) \end{aligned} \quad (3.15)$$

のようになる。したがって、式 (3.15) が最小になる k 次の多項式を最適解とみなし、次数選択を行えば良い。

3.4 物体の追跡処理手順

提案手法を用いて、対象物体から抽出した特徴点を連続的に追跡する手順について説明する。図 3.6 に特徴点の追跡処理の流れを示す。

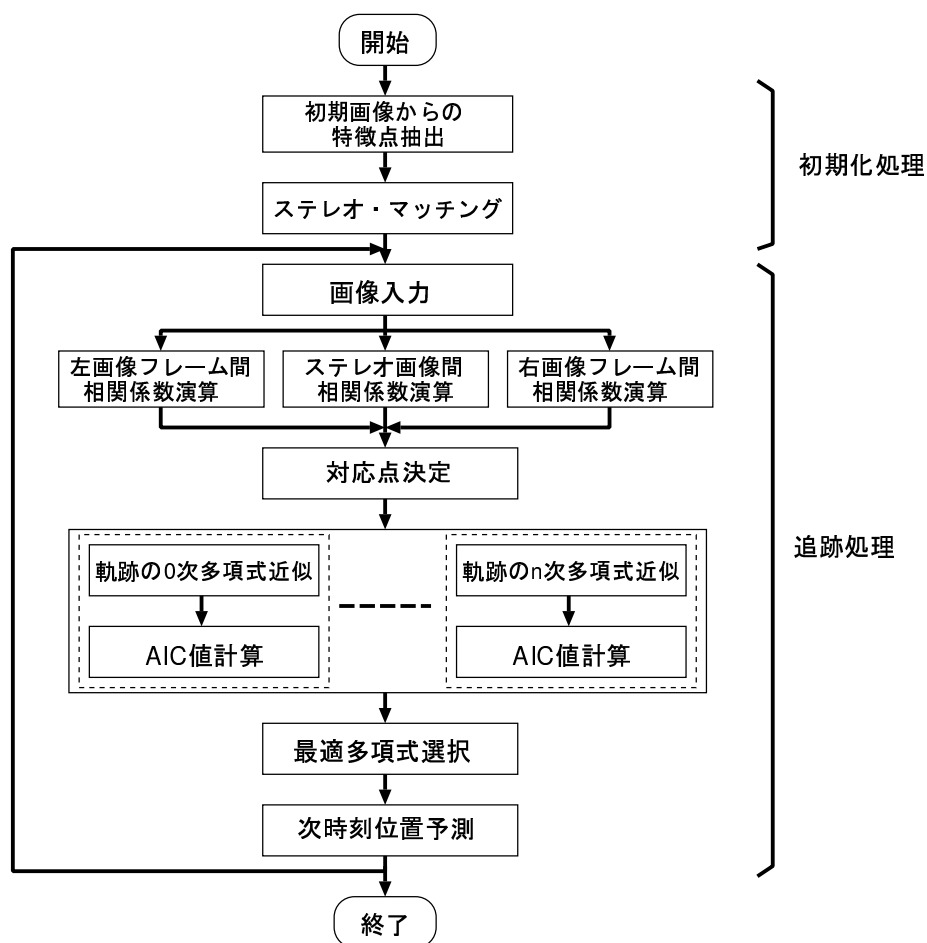


図 3.6 特徴点追跡処理の流れ

(1) 初期化処理

初期化処理では、左右の画像から追跡対象の特徴点抽出を行なう。

(a) 特徴点の抽出

両眼ステレオ法では、左右一对の画像の一方から、対応づけが容易そうな特徴点をあらかじめ抽出し、それらの点についての

み他方の画像中で対応点探索を行なうことが多い。この目的に用いる演算子を、一般に *interest* 演算子と呼んでいる。その代表的なものが方向分散の比の最小値が極大になる点を選択する、Moravec のインタレスト・オペレータ[22]であり、次の式で定義される。

$$\min(\text{dirvar1}, \text{dirvar2}, \text{dirvar3}, \text{dirvar4}) \quad (3.16)$$

ただし、

$$\begin{aligned} \text{dirvar1} &= \text{mean}[(\text{int}_{i,j} - \text{int}_{i+1,j})^2] \\ \text{dirvar2} &= \text{mean}[(\text{int}_{i,j} - \text{int}_{i,j+1})^2] \\ \text{dirvar3} &= \text{mean}[(\text{int}_{i,j} - \text{int}_{i+1,j+1})^2] \\ \text{dirvar4} &= \text{mean}[(\text{int}_{i+1,j} - \text{int}_{i+1,j+1})^2] \end{aligned}$$

ここで、 $\text{int}_{i,j}$ は点 (i, j) における画素の明るさ、 mean は平均値、 dirvar は方向分散値、 \min は最小値を表す。

(b) ステレオ・マッチング

左右画像上の追跡対象の特徴点は、左右一方の初期画像から Moravec のインタレスト・オペレータにより特徴点を抽出した後、テンプレート・マッチングにより左右画像上の対応点を決定する。

(2) 追跡処理

初期化処理で抽出した左右画像上の特徴点を連続して追跡する手順は以下の通りである。

(a) 相関係数演算

第 3.2 節で説明した左画像フレーム間相関係数、右画像フレーム間相関係数、ステレオ画像間相関係数の 3 種類の相関係数の組合せを全て計算する。

(b) 対応点の決定

求めた相関係数から、式 (3.1) が最大になる左右画像上の 2 次元座標を求め、3 次元位置を式 (2.1) ~ (2.3) より求める。

(c) 軌跡の多項式近似

獲得した 3 次元時系列座標から、0 ~ n 次の多項式 (式 (3.5) ~ (3.7)) を最小 2 乗法で近似する。

ただし、以下の問題点がある。

- 最小 2 乗法で多項式を近似する場合、多項式の次数より十分大きな標本数が必要である。したがって、追跡開始直後の時系列座標数が少ない場合、軌跡の多項式近似を行わずに、2 フレーム間の移動量による直線近似による予測位置、あるいは、現在位置などを中心とした広い範囲での探索を行なう必要がある。
- 多項式近似に必要な最大次数を判断する基準が必要である。

(d) AIC 値計算

全ての多項式について、AIC の値 (式 (3.15)) を求める。

(e) 最適多項式選択

全ての多項式の中から、AIC の値が最小になる X,Y,Z 各方向の式を選択する。

(f) 次時刻位置予測

選択した式から次の予測座標を求め、式 (3.2) ~ (3.4) により 2 次元画像上の位置を算出し、探索窓を設定する。

ただし、画像の解像度や雑音の影響による軌跡の多項式近似の精度が不明なため、最適な探索窓の大きさを求める必要がある。

以上の操作を繰り返すことにより、特徴点の連続的な追跡を行なうことができる。

4. ステレオ動画像による実験

本章では、実験を行なうために構築した2種類のステレオ動画像入力装置と実験プログラムの構成について説明したあと、種々の画像を使用した実験結果について述べる。

4.1 ステレオ動画像の入力

4.1.1 ステレオ静止画像入力装置

ステレオ静止画像入力装置は、本研究で使用するステレオ画像のうち、アルゴリズムの性能評価に利用できるように対象物体の動きを制御しながら画像を入力するシステムとして構築した。図 4.1 にステレオ静止画像入力装置の構成を示す。各軸はパルスモータにより表 4.1 に示すような精密移動が可能である。

表 4.1 各軸の移動量

	最大移動量	最小移動量
カメラ操作台 X	±50mm	0.005mm/パルス
カメラ操作台 ピッチ	±12 度	0.001 度/パルス
カメラ操作台 ヨー	360 度	0.001 度/パルス
3 軸物体設置台 X	±250mm	0.005mm/パルス
3 軸物体設置台 Y	±250mm	0.005mm/パルス
3 軸物体設置台 回転	360 度	0.0025 度/パルス

本装置では、対象物体の軌道をプログラムにより制御し、3 軸物体設置台をステップ移動させながら連続して撮影を行なうことにより、疑似ステレオ動画像を得ることができる。

撮影した画像は、24 ビット・カラーの 640 × 486 画素でファイルに格納できる。

4.1.2 ステレオ動画像入力装置

ステレオ動画像入力装置は、本研究で使用する自然画像をステレオ画像として入力するシステムとして構築した。図 4.2 にステレオ動画像入力装置の構成を示す。

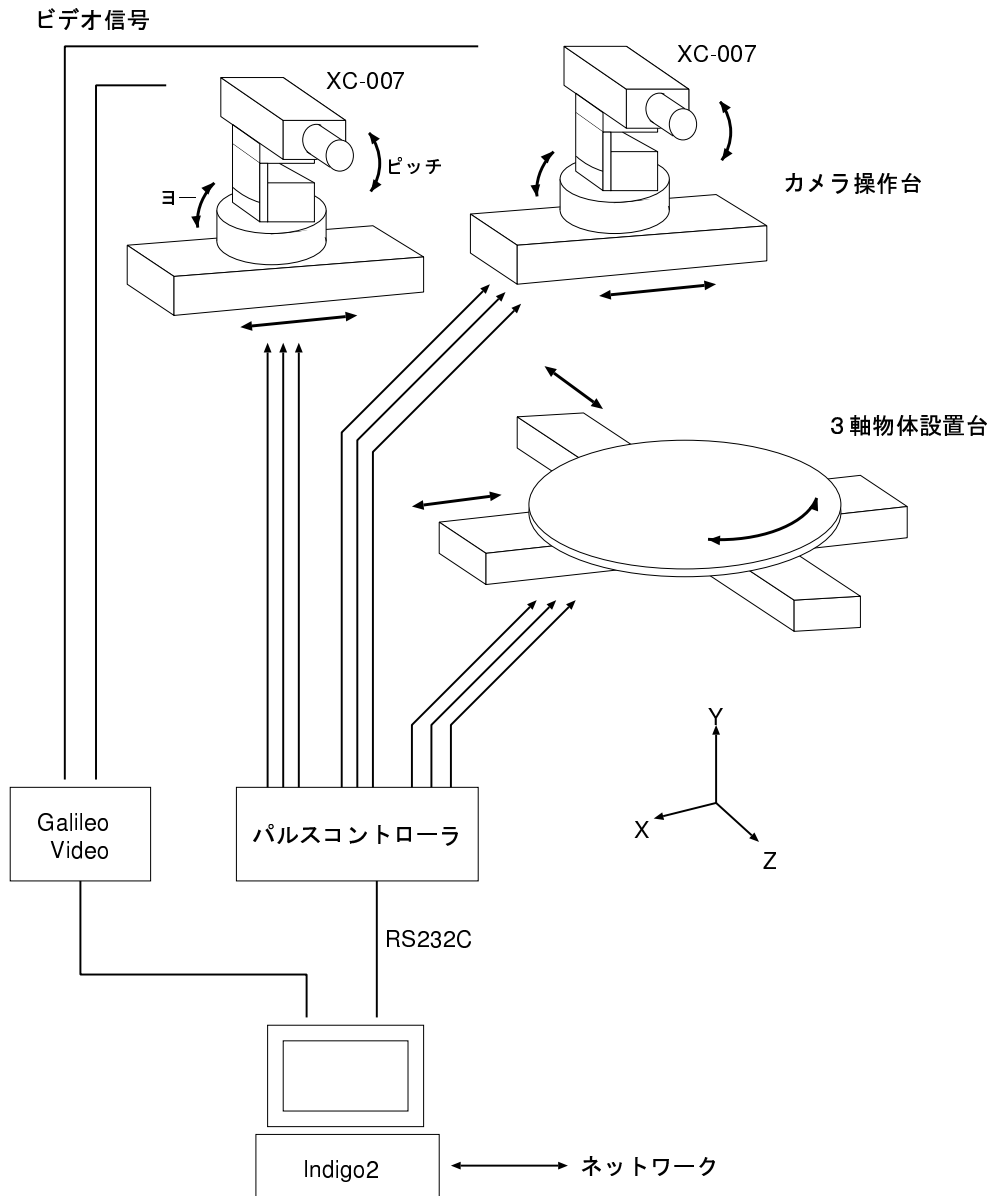


図 4.1 ステレオ静止画像入力装置の構成

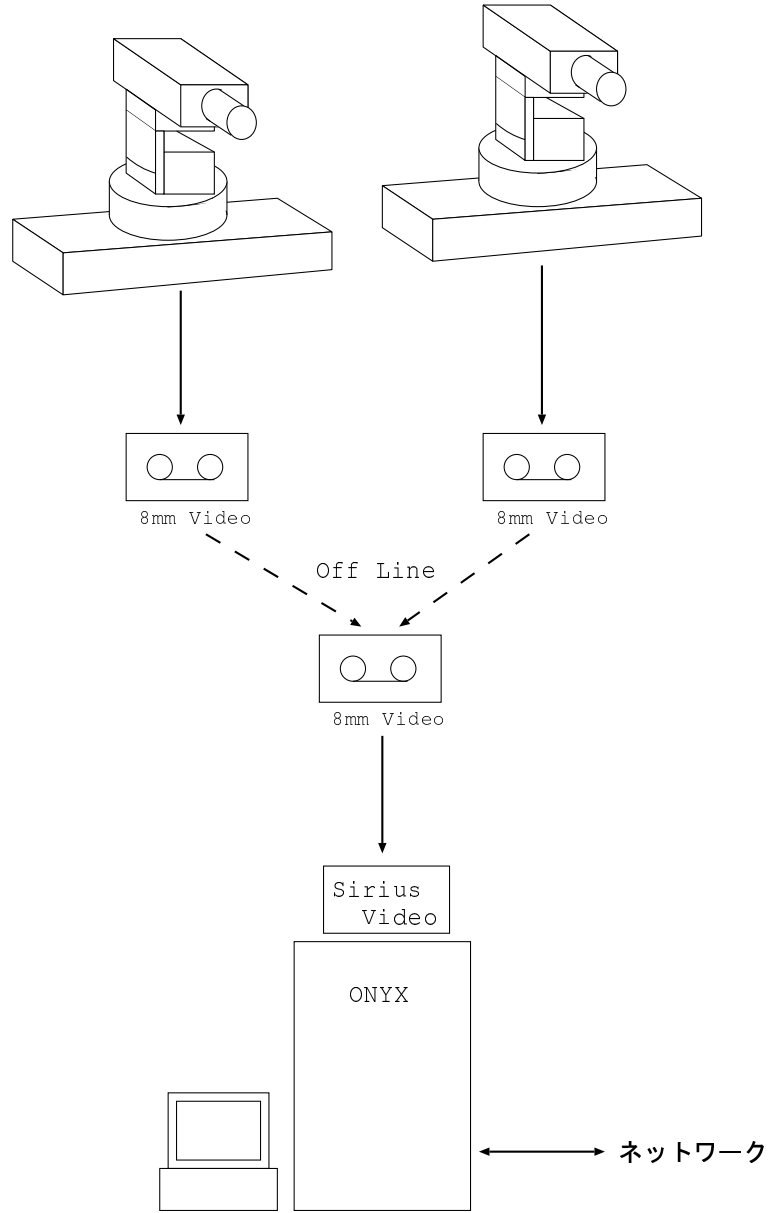


図 4.2 ステレオ動画像入力装置の構成

本装置で使用したカメラ系は、ステレオ静止画像入力装置と同じであるが、動画像を入力するため、8mm ビデオ装置に録画したあと、Sirius Video 装置から 30 フレーム/秒または、60 フィールド/秒の速度で動画像をメモリに格納し、画像入力完了後ディスクに連続して画像を記録する。Sirius Video 装置からは画像サイズ 720 × 486 画素の 24 ビット・カラーでファイルに格納することができる。ただし、走査線上にデータが無い部分も含む。

なお、連続して記録できる時間は計算機のメモリ容量に左右されるが、本装置で使用した画像処理用計算機 SGI 社製 ONYX (主記憶 512MB) では約 4 秒間の記録が可能である。

本装置の構成では、8mm ビデオ装置に録画した画像を使用しているため、Sirius Video 装置での水平同期ずれが見受けられた。このため、ディスクに格納した画像では走査線により 1 画素程度のずれが起り、雑音が含まれたものになった。

また、Sirius Video 装置のアナログ・ビデオ・インタフェースは、インタレース信号である。インタレース画像は 1 フレーム内に時系列の異なる 2 枚の画像が含まれるため、画像解析に利用できるよう図 4.3 に示すフィールド単位の画像分割と、走査線方向に 1/2 に縮小する前処理を行なった。処理結果の例を図 4.4 に示す。

4.2 実験プログラムの構成と実行環境

実験では、対象物体から抽出した特徴点の追跡を行なった。ここでは、実験プログラムの構成や並列処理による高速化などについて述べる。

実験プログラムは、図 4.5 に示す実行制御プロセスと、共有メモリ結合した並列処理プロセスで構成した。

実行制御プロセスは、プログラム全体の実行を管理する部分で、初期化処理、画像ファイルの読み出し、並列処理プロセスの実行管理、画像表示などを行なう。

並列処理プロセスは、各特徴点の追跡に必要な、相関係数の演算、対応点の決定、軌跡の多項式近似・予測などを行なう部分である。各プロセスは、特徴点の抽出処理で検出した追跡対象の 1 特徴点に対し、追跡処理に必要な相関係数演算の 3 プロセス (フレーム間演算 2、ステレオ画像間演算 1) が実行制御プロセスで生成され、複数処理を分散処理することで高速化を図った。

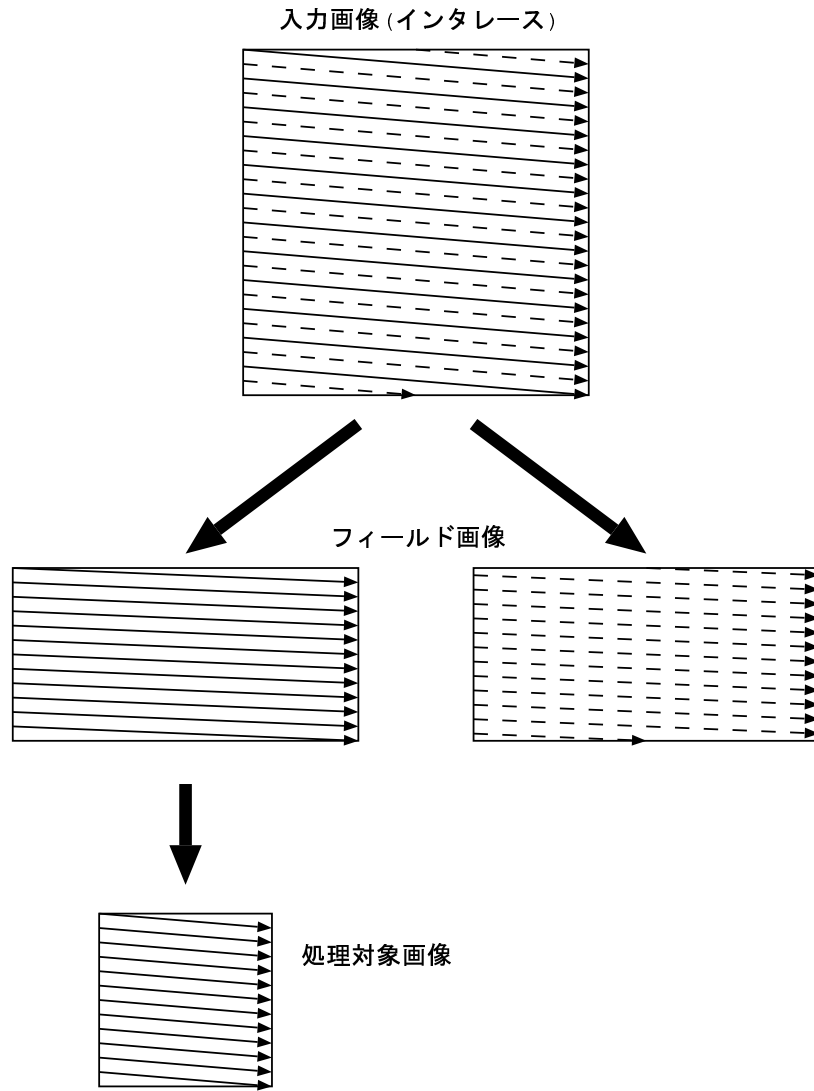


図 4.3 インタレース画像の前処理方法



図 4.4 インタレース画像の前処理結果

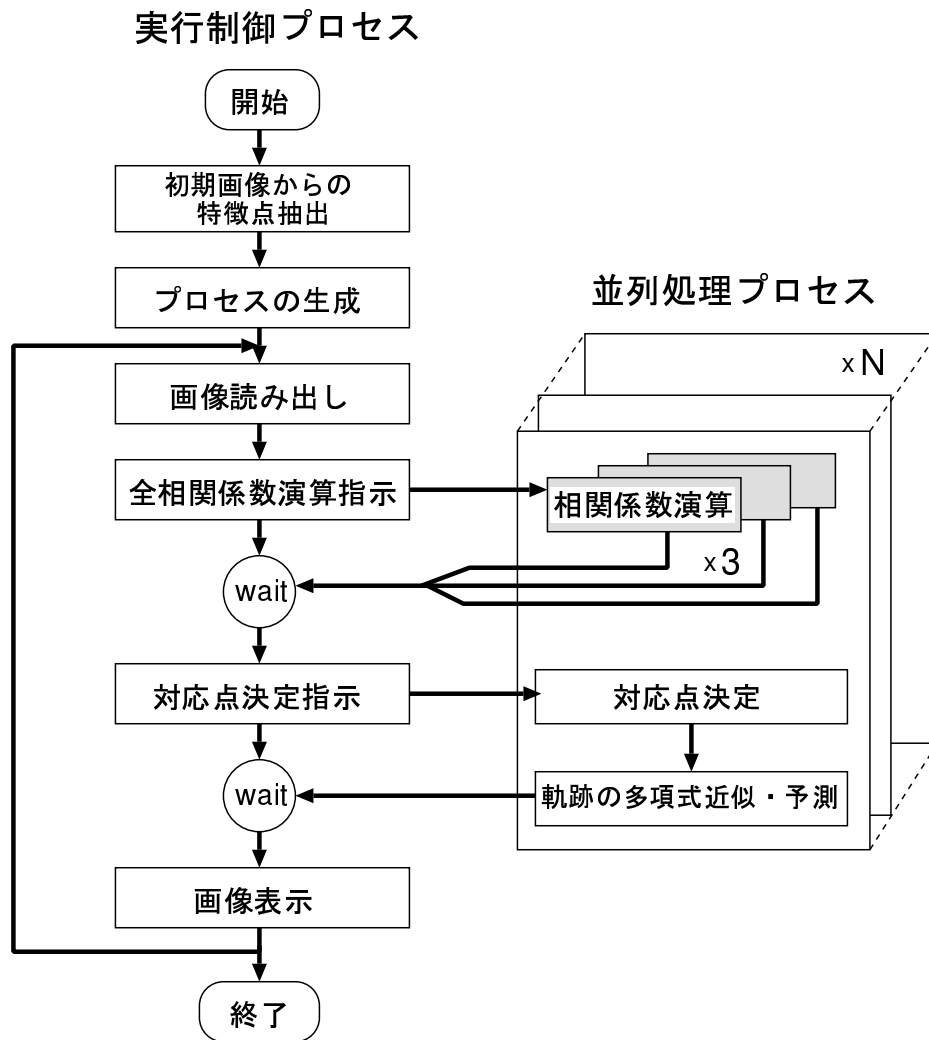


図 4.5 実験プログラムの構成

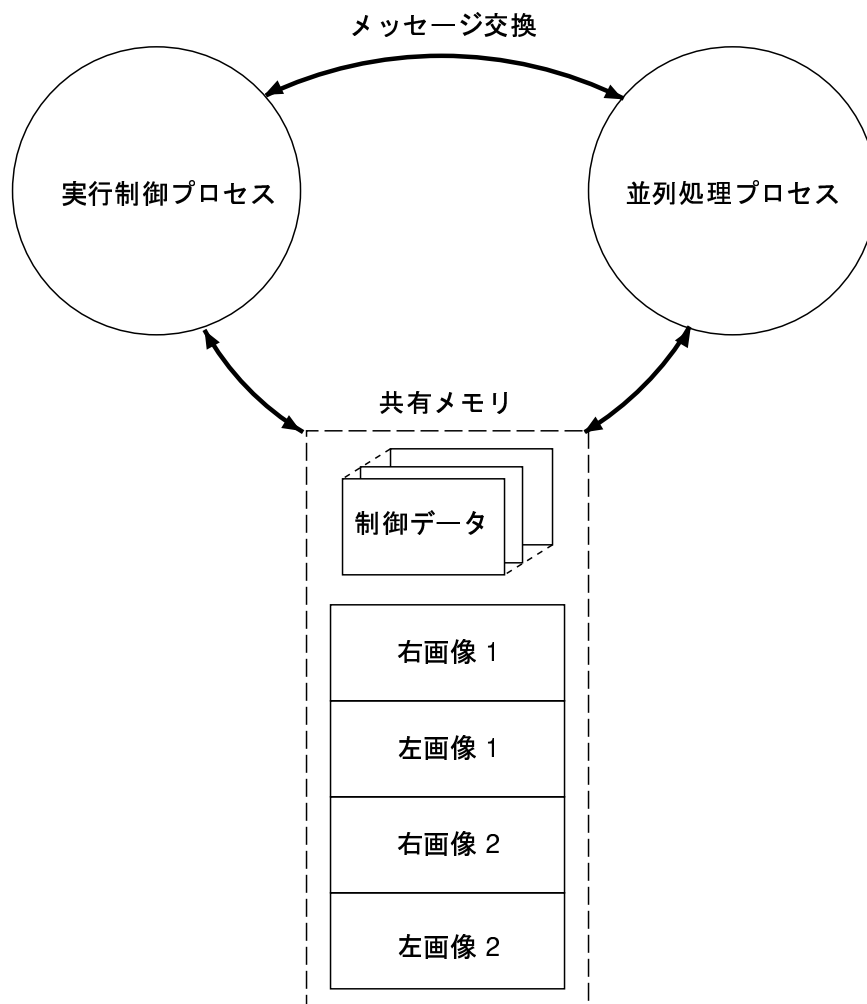


図 4.6 プロセス間の制御構造

また、実行制御には図 4.6 に示すような共有メモリ結合とメッセージ交換を利用した。

共有メモリは、プログラム全体で共通に使用する部分と、個々の追跡処理に必要なデータを格納する部分で構成した。画像記憶用メモリはプログラム全体で共通に使用する部分で、処理に必要な 2 組のステレオ画像を格納する 4 ブロックで構成した。制御データは、特徴点の位置、探索窓の大きさ、相関係数の演算結果など追跡処理に必要な情報を記憶する部分で、追跡する特徴点 1 つに対し 1 ブロック使用する構成とした。

メッセージ交換は、実行制御プロセスと並列処理を行なう各プロセス間で行ない、処理の開始、終了、処理内容などの通信を行なう。

複数の処理を同時に行ない処理の高速化を図るため、MIMD 型並列計算機 SGI 製 ONYX (88 SPECint92, 97 SPECfp92 × 8CPU) 上で実行した。なお、起動するプロセス数は、(特徴点数 × 3) + 1 のため、複数の特徴点追跡ではプロセッサ数以上のプロセスを実行することになる。この場合のプロセスの CPU 割り当ては、オペレーティングシステム (IRIX Release 5.2) が行なった。

4.3 連続ステレオ静止画像による実験と評価

本節では、連続ステレオ静止画像を使用した特徴点の追跡実験により、提案手法の有効性、探索窓の大きさ、多項式近似に使用するパラメータの範囲などを明らかにする。

4.3.1 実験画像

実験には、3 軸物体設置台を楕円軌道で移動させながら表 4.2 に示す条件で撮影した 180 組のステレオ画像を使用した。ステレオ画像の例を

表 4.2 撮影条件

設置台の動き	楕円軌道 (長径 200mm, 短径 100mm)
撮影間隔	2 度/フレーム
撮影距離	1.6m
カメラ間距離	80mm
画像サイズ	640 × 486 画素

図 4.7 に、Moravec のインタレスト・オペレータにより抽出した追跡対象の特徴点を図 4.8 に示す。

4.3.2 ステレオ画像間とフレーム間の対応点探索による物体の追跡結果

本項では、エピポーラ拘束に基づいたステレオ画像対応点探索とフレーム間対応点探索により、正しい対応点を決定可能であるか検討する。

図 4.9 に 2 つのレンズ中心の中点を原点として定め、X 軸を基線と平行な方向、Y 軸を基線と光軸に垂直な方向、Z 軸を光軸方向とした 3 次元座標で図 4.8 の特徴点 3 と 5 を追跡した結果を示す。(本実験で用いた、軌跡の多項式近似における式の次数、探索窓の大きさについては後述する。)

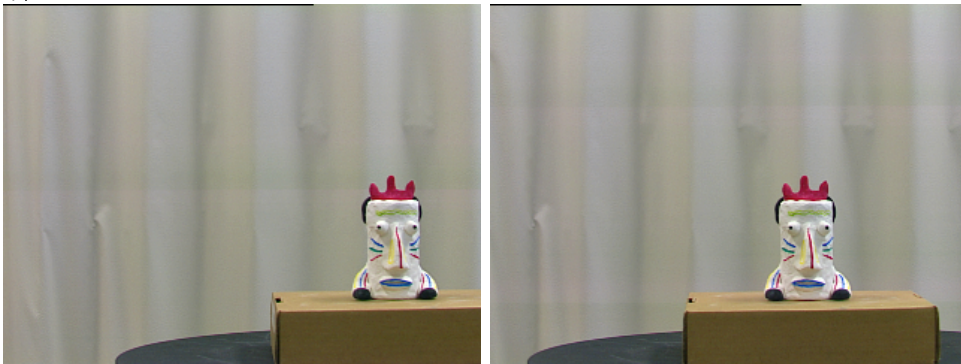
図 4.9 から明らかなように、追跡結果は楕円を描いておらず徐々に軌道から外れて行く。これは、対象物体が動いており画像中の特徴点の大きさ・方向が時間とともに変化することから、対応点探索のための基準画像をフレーム毎に最新の対応点画像に更新して使用したため、量子化誤差により基準点が徐々にずれて行くことに起因している。この問題を解決するためには、

- 一定時間間隔でインタレスト・オペレータを再適用し対応点を補正
- 特徴量をエネルギーとする関数を導入しずれを補正
- 初期画像を基準画像として固定

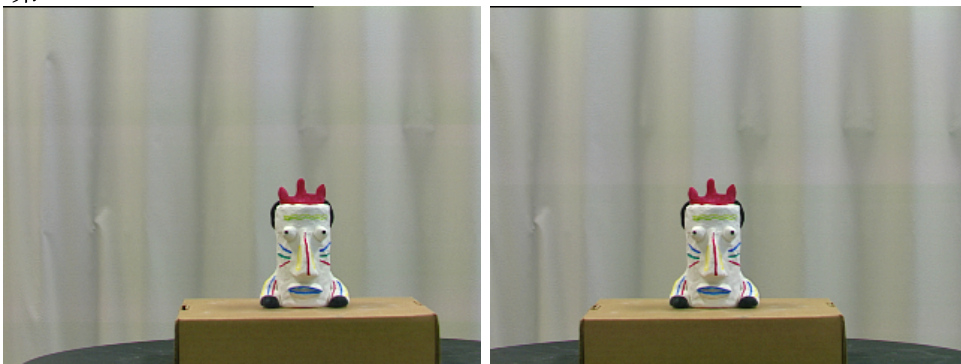
などの方法が考えられる。

本実験では、使用した画像中で物体の大きさがあまり変化せず回転を伴わないため、初期画像を基準画像として固定した。図 4.10 に初期画像を基準画像として固定した場合の特徴点 3 と 5 の追跡結果を示す。同図から明らかなように、良好な軌跡を得ることができた。このことは、エピポーラ拘束に基づいたステレオ画像対応点探索と初期画像を基準画像としたフレーム間の対応点探索により、正しい対応点を決定できることを示している。

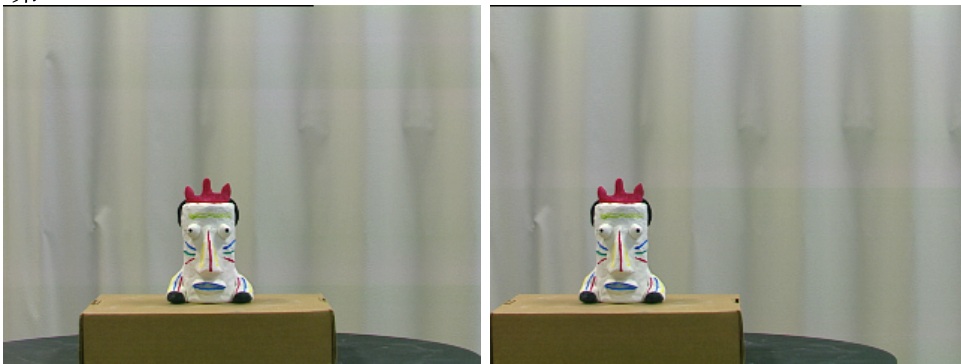
第 0 フレーム



第 45 フレーム



第 90 フレーム



第 135 フレーム

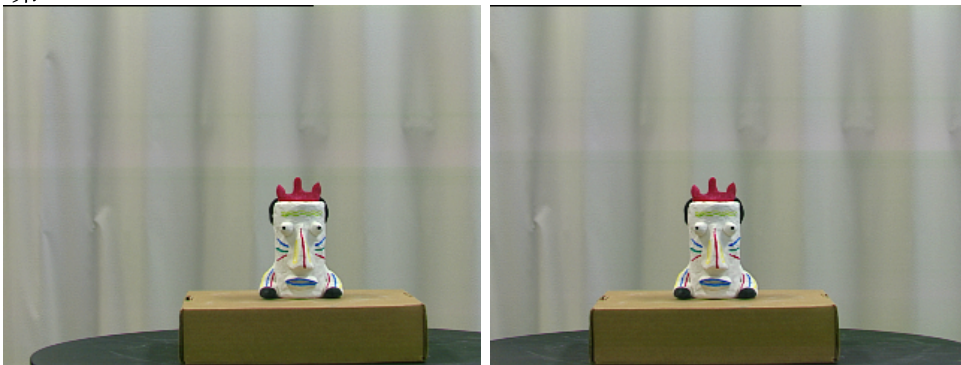


図 4.7 連続ステレオ静止画像の例

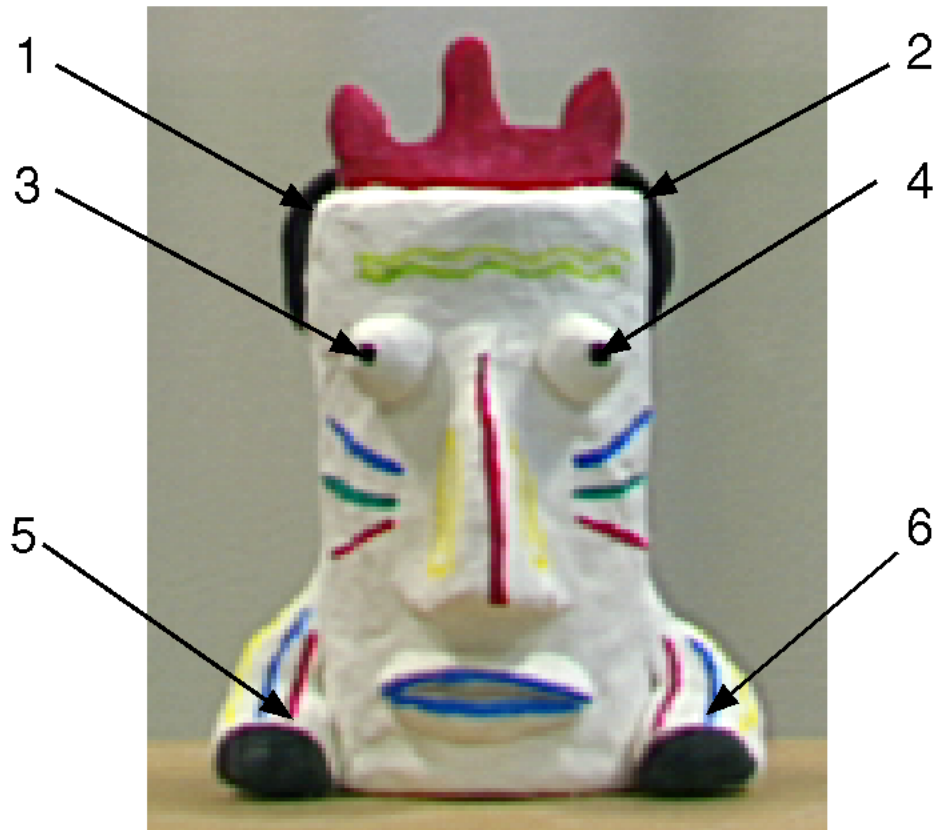


図 4.8 追跡物体上の特徴点位置

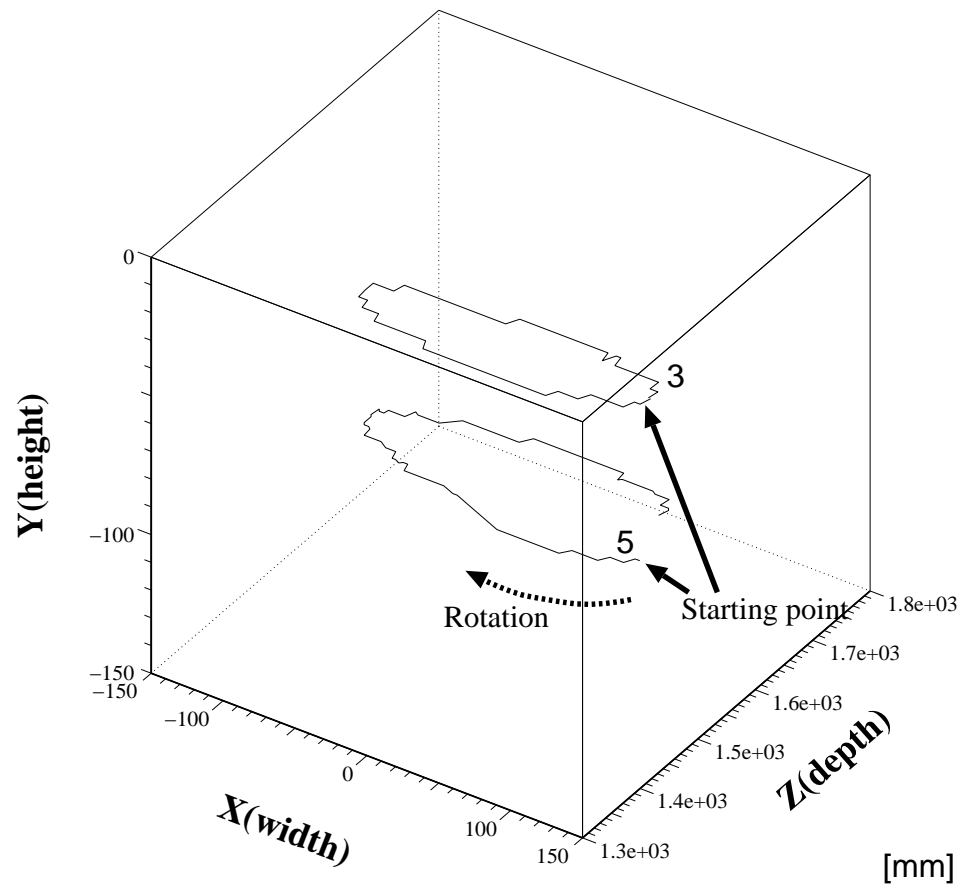


図 4.9 提案手法による特徴点の追跡結果

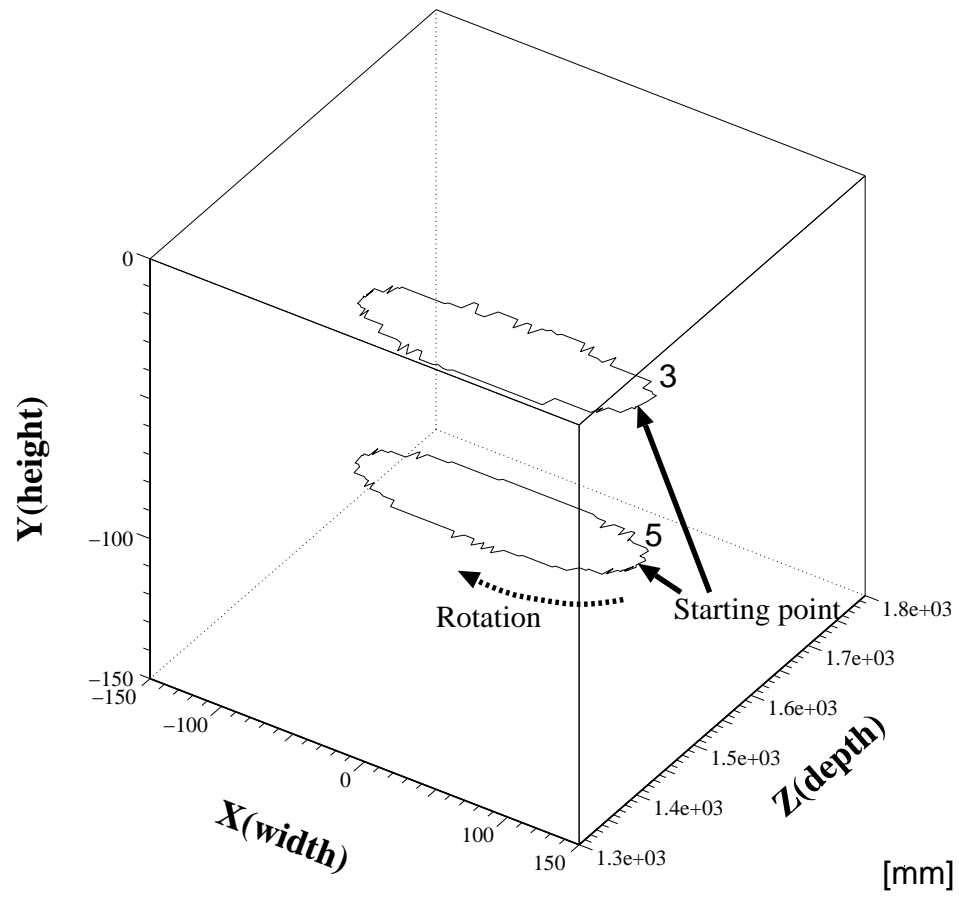


図 4.10 初期画像を基準画像に固定した場合の特徴点追跡結果

4.3.3 軌跡の多項式近似における次数と AIC の関係

軌跡の多項式近似における次数選択は、何次式までの多項式で行なえば良いか求める必要がある。そこで、多項式近似に用いる時系列座標数 40 における 0 ～ 15 次の多項式による AIC の値を求め、次数の検討を行った。

図 4.11 ～ 4.13 に各方向の多項式における AIC の値の変化を示す。図には多項式に含まれる誤差が、平均 0.5 画素の量子化誤差のみと仮定したとき AIC の値がどのように変化するか併記した。

- X 方向の多項式における AIC の値

AIC の値は、3 次式を境にほとんど変化しなくなる傾向がある (図 4.11 参照)。

- Y 方向の多項式における AIC の値

3 軸物体設置台は縦方向に移動しないが、カメラと 3 軸物体設置台が平行に設置されていないため、両者の傾きによる変移と量子化誤差に従うことになる。このため、0 次を除き量子化誤差のみによる AIC 曲線を下まわっている (図 4.12 参照)。

- Z 方向の多項式における AIC の値

画像から得られる Z 方向の座標値は他の方向に比べ分解能が低いため、AIC の値は、全体的に大きな値になっている (図 4.13 参照)。

しかし、0 ～ 1 次への傾斜は X 方向の AIC の変化と同じ傾向を示していることから、カメラ間距離を広くし分解能を上げると、高次でも X 方向の AIC の値に近付くと考えられる。

また、図には示していないが、連続した多くのフレームで座標値が変化しない場合、軌跡を正確に近似でき AIC の値が非常に小さくなることがあった。

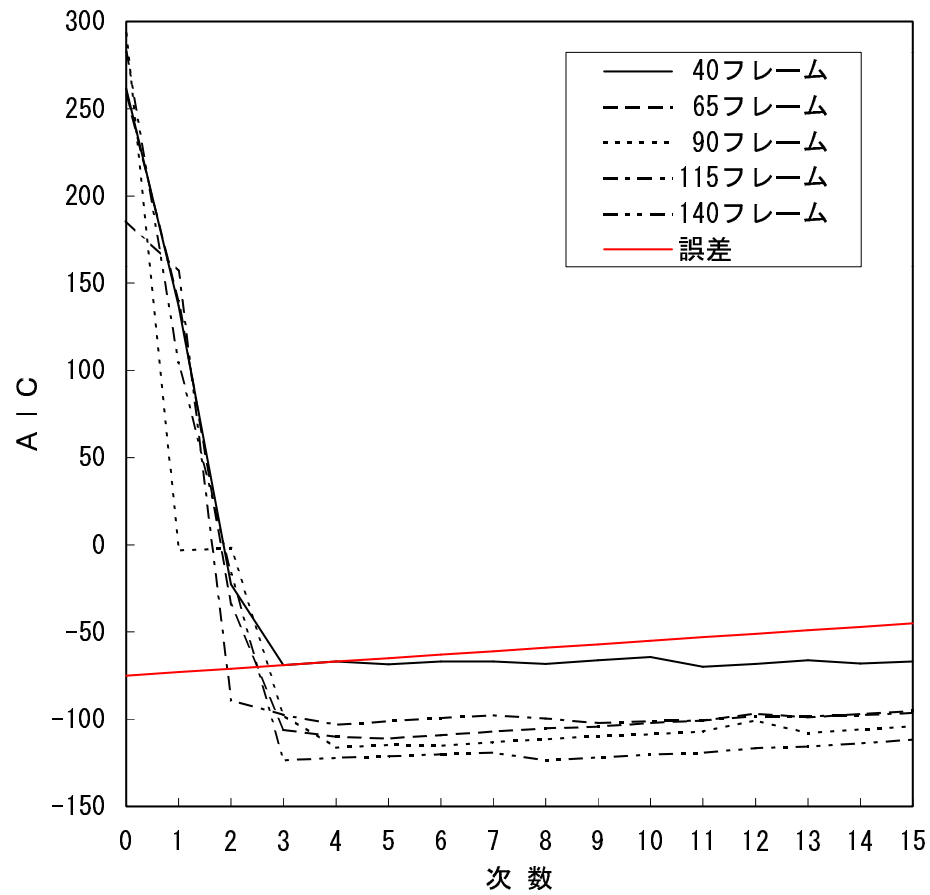


図 4.11 X 方向の多項式における次数と AIC 値の関係

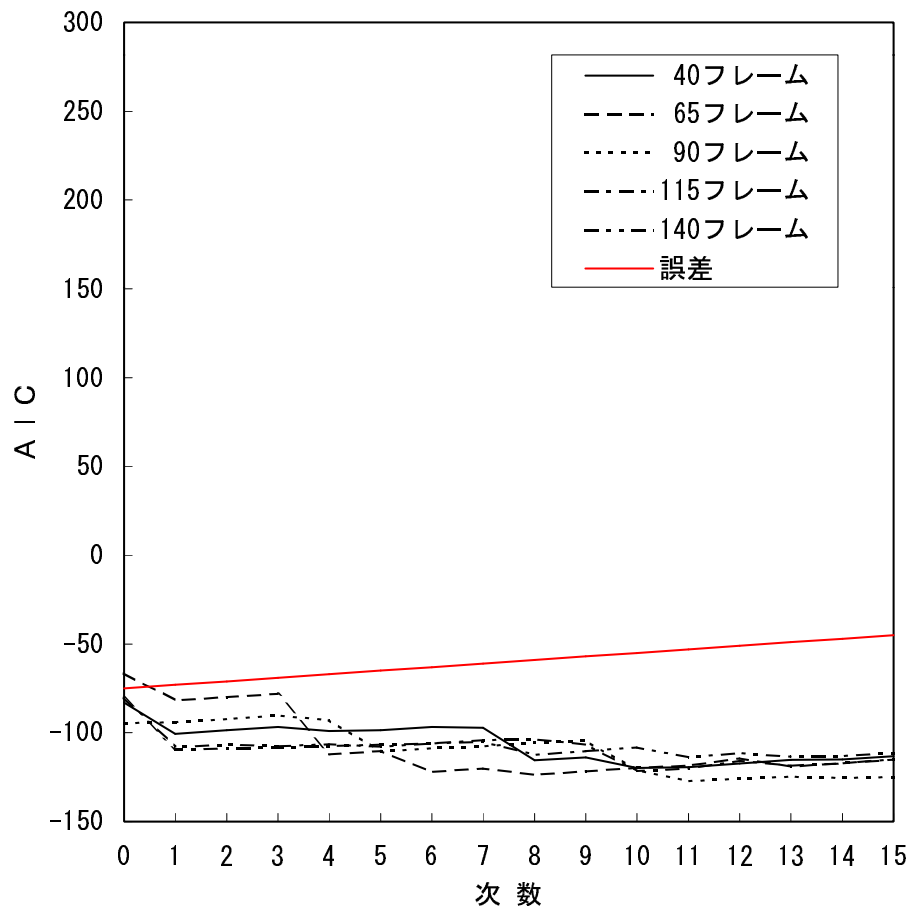


図 4.12 Y 方向の多項式における次数と AIC 値の関係

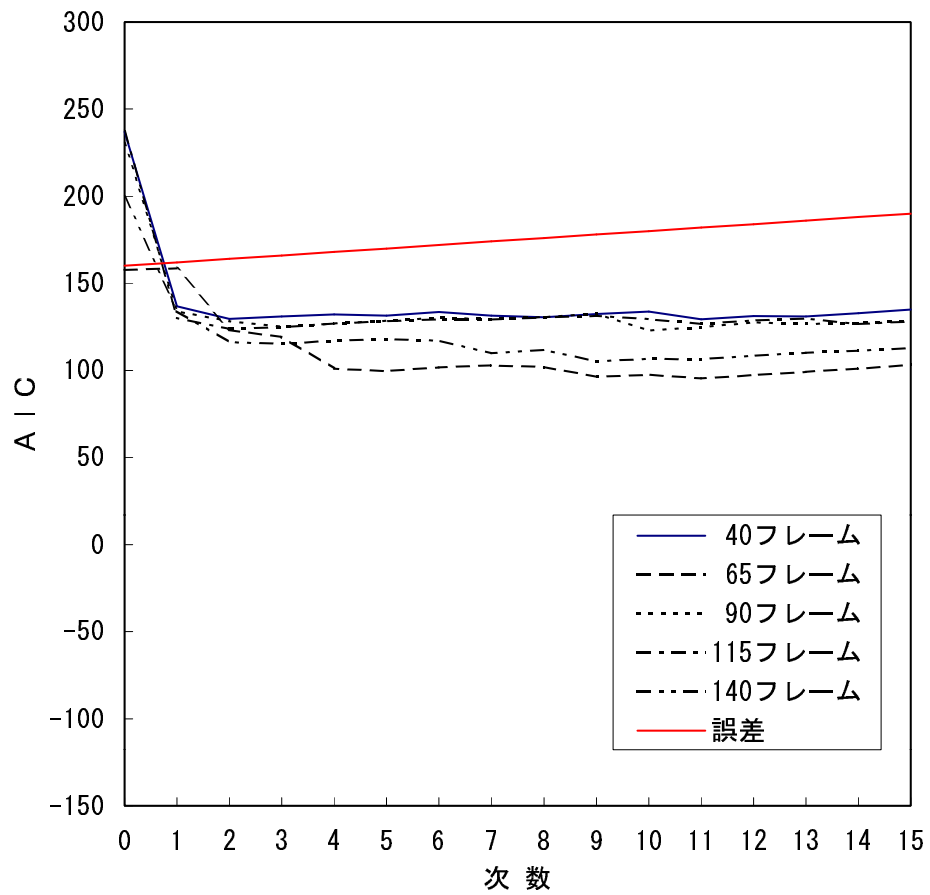


図 4.13 Z 方向の多項式における次数と AIC 値の関係

これらに共通な点は、各 AIC の値は量子化誤差のみによる AIC 曲線を下まわったところから変化が緩やかになっていることである。これは、同一画像フレームにおける次数が異なる多項式間において、AIC の計算式(式 (3.15)) が次のような特性を持つためである。

- (1) 右辺第 1 項の差が小さくなり、軌跡の特徴を反映した多項式になった。
- (2) 次数による右辺第 2 項の増加により、右辺第 1 項の減少分 (オーバーフィッティング傾向) を打ち消す。ただし、右辺第 2 項の増加分で右辺第 1 項の減少分を相殺できず、AIC の値が次数とともに減少する場合があります、AIC 最小化原理を適用できない場合も有る。

このような特性から、量子化誤差のみによる AIC 値以下になった多項式を選ぶことも考えられるが、量子化誤差のみによる AIC 値を下まわる最低次数の多項式では、AIC の値が十分に小さくならない場合が有ることから、量子化誤差のみによる AIC 値を下まわる最低次数より若干高い次数まで多項式近似を行ない選択すれば良いことになる。

以上の結果から、本研究では 0 ~ 4 次の多項式から最適次数を決定することにした。

4.3.4 軌跡の多項式近似精度と探索窓の大きさの関係

本項では、軌跡の多項式近似に用いる時系列座標数による次時刻位置の近似精度と、近似精度に左右される対応点検出に必要な探索範囲の関係について検討を行なう。

物体上の 6 つの特徴点の軌跡の多項式近似に必要な時系列座標を得るため、45 フレームまでの特徴点追跡は、直線移動の前提による移動予測を行ない、画像平面に広い探索窓を設定し追跡した。46 フレーム以降の特徴点追跡は、獲得している時系列座標から軌跡の多項式近似を行ない、予測される画像平面上の位置に $3 \times 3 \sim 11 \times 11$ 画素の 5 種類の探索窓を設定し追跡を行なった。

探索窓の大きさと軌跡の多項式近似に使用した時系列座標数による追跡結果の成否を表 4.3 に示す。探索窓の大きさ 11×11 画素、軌跡の多

項式近似用の時系列座標数 40 での結果を真の軌跡と仮定した場合の、探索窓の大きさと誤対応の割合を表 4.4 に示す。また、左右の 2 次元画像上に現れる特徴点の予測位置と実際の対応点位置のずれが、探索窓の大きさと軌跡の多項式近似に用いる時系列座標数によりどのように変化するかを図 4.14 に示す。

表 4.3 探索窓の大きさと多項式近似に使用する時系列座標数による追跡の成否

時系列座標数 探索窓 (画素 × 画素)	10	15	20	25	30	35	40
3 × 3	×	×	×	○	○	○	○
5 × 5	×	○	○	○	○	○	○
7 × 7	×	○	○	○	○	○	○
9 × 9	○	○	○	○	○	○	○
11 × 11	○	○	○	○	○	○	○

○ 成功 × 失敗

表 4.4 探索窓の大きさと多項式近似に使用する時系列座標数による誤対応の割合

時系列座標数 探索窓 (画素 × 画素)	10	15	20	25	30	35	40
3 × 3				3.8	6.1	2.3	1.7
5 × 5		1.7	0.0	0.0	0.0	0.0	0.0
7 × 7		0.0	0.0	0.0	0.0	0.0	0.0
9 × 9	0.5	0.0	0.0	0.0	0.0	0.0	0.0
11 × 11	0.0	0.0	0.0	0.0	0.0	0.0	—

[%]

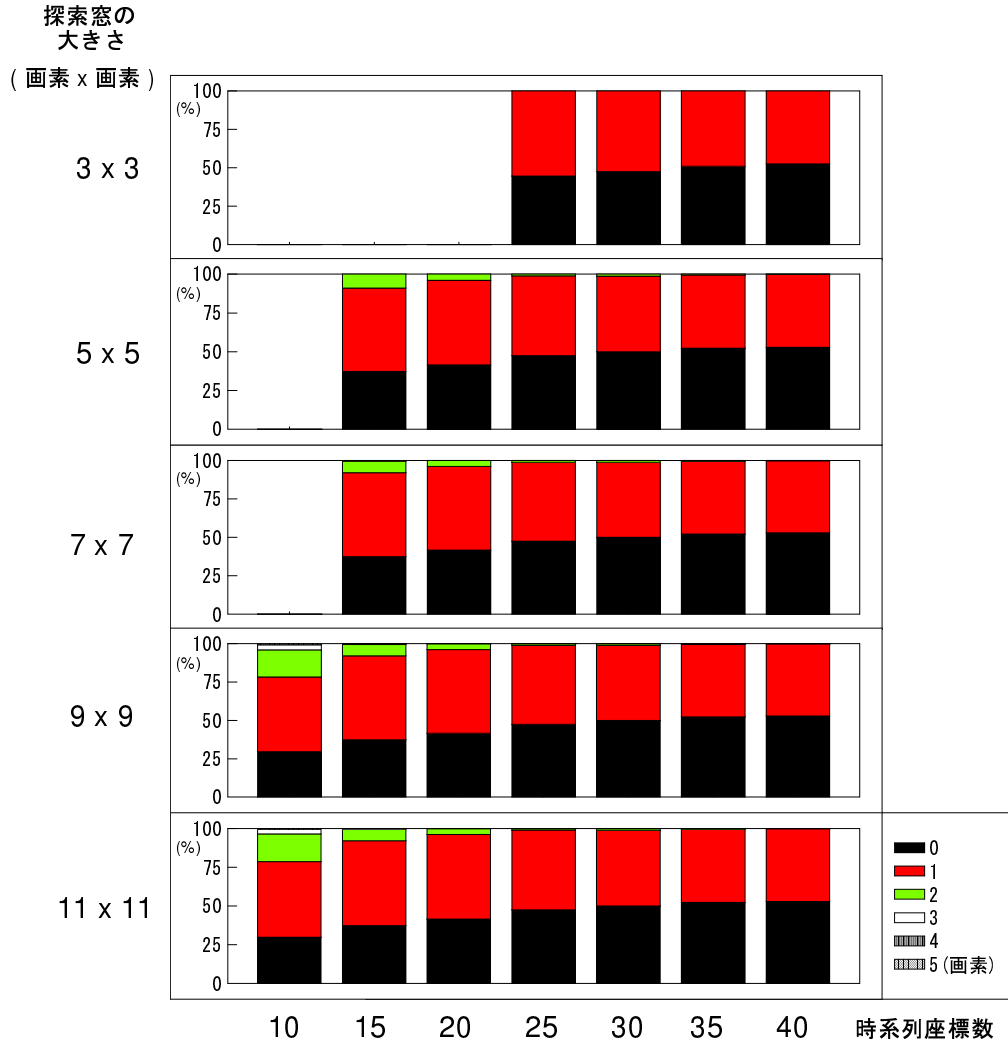


図 4.14 探索窓の大きさと多項式近似に使用する時系列座標数による予測位置と対応点のずれの割合

これらの結果から以下のことがわかる。

- 正しい軌跡を得られた最小の探索窓の大きさは 5×5 画素であった。
- 探索窓の大きさが 3×3 画素では、正常に追跡できているように見えても若干の誤対応が含まれる。

これは、画像の量子化誤差と軌跡の予測誤差を探索窓 3×3 画素では補正しきれないことを示している。

- 探索窓の大きさ 5×5 画素以上で、軌跡の多項式近似の時系列座標数 20 点以上では、次フレームの画像上の位置は 2 画素誤差以内に納まった。

また、軌跡の多項式近似の時系列座標と次フレームの画像上の位置が 1 画素誤差以内に入る確率は、時系列座標数 20 点で 96.1%、25 点で 98.9%であった。

探索窓の大きさを小さくするには、正確な軌跡の多項式を求める必要があるが、時系列座標数 25 点以上で精度の高い軌跡の多項式近似を行なえた。

これらの結果から、探索窓の大きさ 5×5 画素以上、軌跡の多項式近似用時系列座標 25 以上が最適であると言える。ただし、動画像処理に必要な計算時間と窓の大きさ・多項式近似用時系列座標数の関係については第 4.3.5 項で検討する。

また、時系列座標数を多くすると対象物体の動きが急激に変化する場合同追従が困難になることも考えられる。したがって、実画像を使用し最適な時系列座標数を求めることが必要である。

4.3.5 対応点探索処理の計算時間

軌跡の多項式近似と探索窓の大きさが、演算時間におよぼす影響について検討する。

計算機内の 1/60 秒タイマーを使用し相関係数算出・対応点決定処理に必要な平均演算時間を計測した結果を図 4.15 に示す。この図から以下のことがわかる。

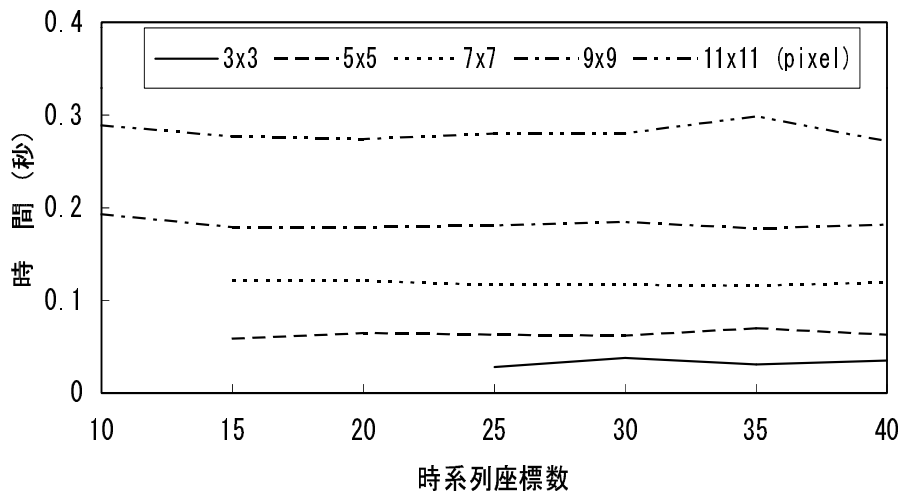


図 4.15 探索窓の大きさと軌跡の多項式近似用時系列座標数による対応点探索計算時間

- 軌跡の多項式近似に使用する時系列座標数が増加しても処理性能への影響はほとんどない。時系列座標数が増加すると軌跡の多項式近似に時間がかかり、処理時間は単調増加するはずであるが、計測用タイマーの誤差範囲に入り性能測定結果に現れなかった。

また、相関係数の演算回数から逆算して得られる、相関係数演算時間以外の軌跡の多項式近似処理時間、システム・オーバーヘッドなどの所要時間は約 0.03 秒である。

- 探索窓を大きくすると処理性能が著しく低下する。これは、相関係数の演算回数が探索窓の大きさの 3 乗に比例して増加することに起因している。
- 探索窓の大きさ 5×5 画素で動画処理に必要な 15 フレーム/秒の処理性能が得られた。

3次元情報の解析処理などとの併用を考えた場合、さらに高速な処理が必要であるが、先に述べた SSDA 法などを用いた高速なマッチング処理を行なうことにより、処理効率を上げることが可能である。

4.4 人物の実動画像による実験と評価

本節では、人物を含む室内シーンの動画像を使用した特徴点の追跡実験により、提案手法が実画像に対して有効であることを示す。

4.4.1 実験画像

実験には、衣服に特徴点用のマーカを付けた人物を表 4.5 の条件で撮影した 123 組のステレオ動画像を使用した。ステレオ画像の例を図 4.16 に、Moravec のインタレストオ・ペレータにより抽出した追跡対象の特徴点を図 4.17 に示す。

表 4.5 撮影条件

撮影間隔	1/30 秒
シャッター速度	1/250 秒
撮影距離	約 7m
カメラ間距離	400mm
原画像サイズ	720 × 486 画素
処理画像サイズ	360 × 243 画素

4.4.2 特徴点追跡結果

物体上の 5 つの特徴点の軌跡の多項式近似に必要な時系列座標を得るため、40 フレームまでの特徴点追跡は、直線移動の前提による移動予測を行ない、画像平面に広い探索窓を設定し追跡した。41 フレーム以降の特徴点追跡は、獲得している時系列座標から次数が 0 ～ 4 次の 5 種類の多項式による軌跡を求め、AIC の値が最小になる次数を逐次選択し、予測される画像平面上の位置に $3 \times 3 \sim 11 \times 11$ 画素の 5 種類の探索窓を設定し追跡を行なった。

図 4.18 に 2 つのレンズ中心の中点を原点として定め、X 軸を基線と平行な方向、Y 軸を基線と光軸に垂直な方向、Z 軸を光軸方向とした 3 次元座標で 5 箇所の特徴点追跡結果を示す。

第0フレーム



第40フレーム



第81フレーム



第122フレーム



図 4.16 人物の実ステレオ動画像の例



図 4.17 追跡対象人物の特徴点位置

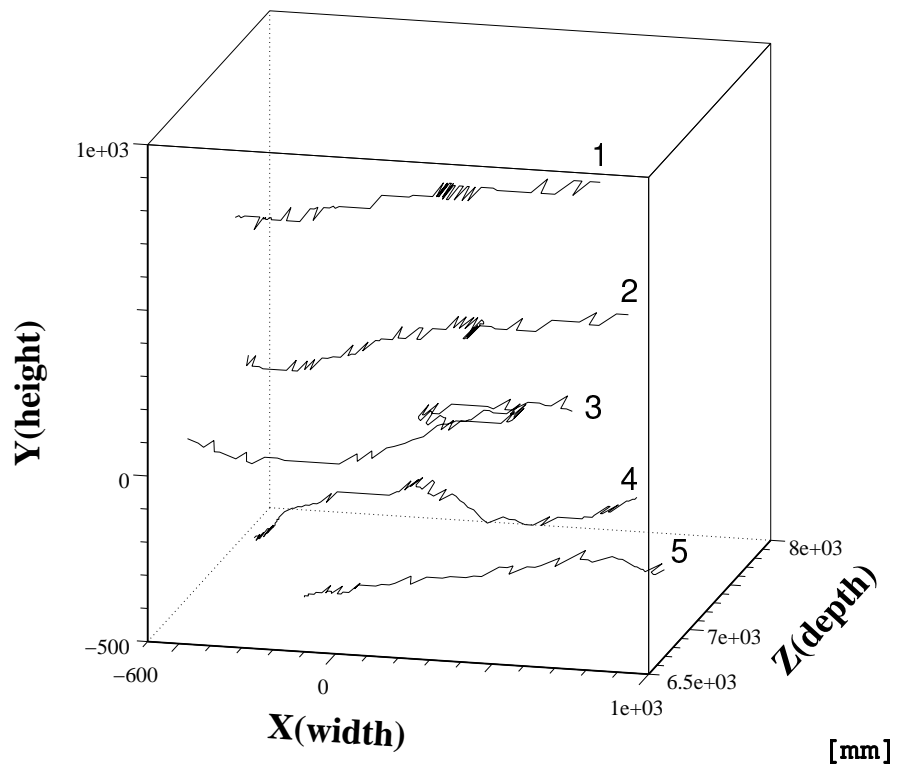


図 4.18 人物の追跡結果

4.4.3 軌跡の多項式近似精度と探索窓の大きさの関係

探索窓の大きさと軌跡の多項式近似に使用した時系列座標数による追跡結果の成否を表 4.6 に示す。探索窓の大きさ 11×11 画素、軌跡の多項式近似用の時系列座標数 40 での結果を真の軌跡と仮定した場合の、探索窓の大きさと誤対応の割合を表 4.7 に示す。また、左右の 2 次元画像上に現れる特徴点の予測位置と実際の対応点位置のずれが、探索窓の大きさと軌跡の多項式近似に用いる時系列座標数によりどのように変化するかを図 4.19 に示す。

表 4.6 探索窓の大きさと多項式近似に使用する時系列座標数による追跡の成否

探索窓 (画素 × 画素)	時系列座標数	10	15	20	25	30	35	40
3×3		×	×	×	×	×	×	×
5×5		×	×	○	×	×	×	×
7×7		×	○	○	○	○	○	×
9×9		○	○	○	○	○	○	○
11×11		○	○	○	○	○	○	○

○ 成功 × 失敗

表 4.7 探索窓の大きさと多項式近似に使用する時系列座標数による誤対応の割合

探索窓 (画素 × 画素)	時系列座標数	10	15	20	25	30	35	40
5×5				0.0				
7×7			0.5	0.0	0.0	0.5	0.5	
9×9		0.5	0.0	0.0	0.0	0.0	0.0	0.0
11×11		0.0	0.0	0.0	0.0	0.0	0.0	—

[%]

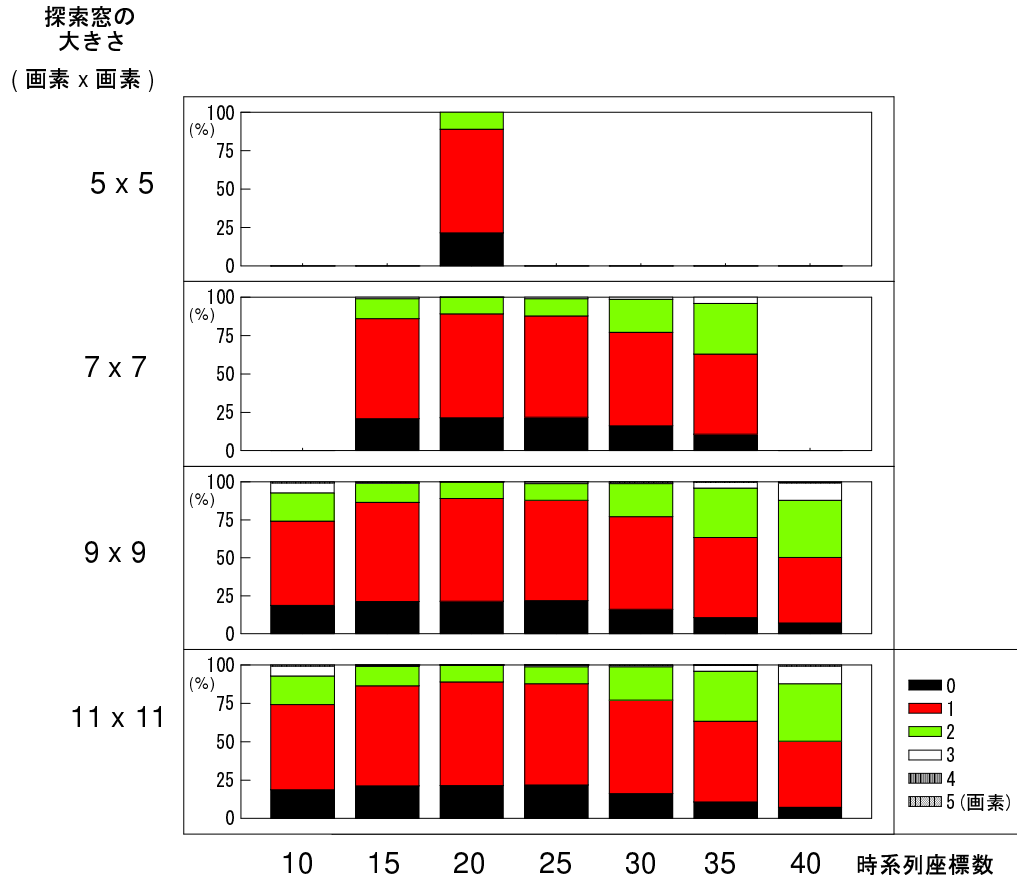


図 4.19 探索窓の大きさと多項式近似に使用する時系列座標数による予測位置と対応点のずれの割合

これらの結果から以下のことがわかる。

- 正しい軌跡を得られた最小の探索窓の大きさは 5×5 画素であった。
- 軌跡の多項式近似の時系列座標と次フレームの画像上の位置が 2 画素誤差以内に入る確率は、時系列座標数 20 ~ 30 点で 98% 以上であった。

また、時系列座標数 20 点で次フレームの画像上の位置は 2 画素誤差以内に納まった。

- 軌跡の多項式近似の時系列座標数が多くなると予測位置と対応点のずれが大きくなる。これは、足首・手首近くなど、ほとんど停止した状態から急激に動く、あるいは、急激に停止するため、軌跡の多項式近似に用いる座標数が多い場合、対象物の動きに追従できなくなることに起因する。また、追跡できた場合でも予測した位置とのずれが大きかった。

本節での実験結果は、連続静止画像を持ちいた第 4.3 節の実験に比べ悪くなっている。これは以下の原因によると考えられる。

- 静止画像に比べ画像品質が悪く対応点のずれが大きくなった。

前述したように、インタレース画像を変換したため、実験画像が本来の画像を 1 画素おきにしたものになった。また、走査線の同期ずれによる画像劣化の影響も大である。

- マーカー部分のコントラストが低く Moravec のインタレスト・オペレータの閾値を小さくし (静止画像を用いた場合の $1/2$)、5ヶ所個の特徴点を抽出した。このため、周囲の画像と分離しにくい画像パターンになり、誤対応が起きやすかった。
- 基準画像を固定したため、対象物体に回転運動が含まれると対応点を見つけにくい。

例えば、図 4.20 に示すような回転により異なる角・辺が初期画像と類似し誤対応をおこす。探索窓サイズを小さくすると、この傾向は小さくなるが、探索範囲が小さいため追跡できなくなった。

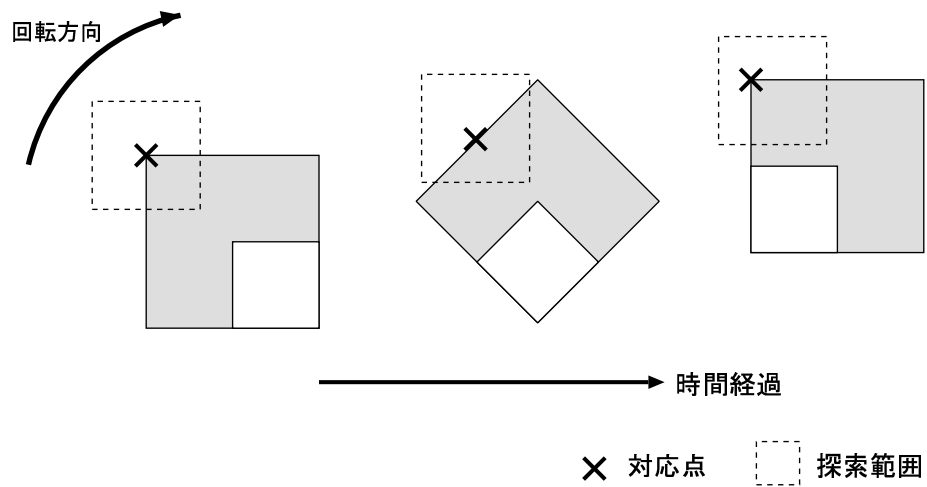


図 4.20 物体の回転の影響による誤対応

本実験において、雑音が多く含まれている画像を用いても、探索窓の大きさ 7×7 画素以上では、軌跡の多項式近似に 20 ~ 30 点の時系列座標を使用することにより、2 画素誤差以内に入る確率は 98% 以上であった。また、探索窓の大きさ 5×5 画素、多項式近似に用いる座標数 20 で正しく追跡を行なえたことは、静止画像を用いた実験結果と一致している。これらのことから、自然画像に対しても本手法が有効であると言える。

5. 結論

本論文では、ステレオ動画像から対象物体上の特徴点を効率良く追跡するための手法と実験結果について報告した。

本論文では、新たに構築したステレオ動画像入力装置から得られた実動画像を使用した実験を行ない、以下の点を明らかにした。

- (1) フレーム間の対応点探索とエピソード拘束に基づいたステレオ画像間対応点探索に相関係数を用いたマッチング処理を適用し、これらを同時並行的に行なうことでステレオ動画像の対応点探索が連続的に行なえることを示した。
- (2) ステレオ動画像から得られる物体上の特徴点の3次元軌跡を時間の関数として多項式近似し、この多項式から次時刻における3次元位置を精度良く予測できることを示した。また、情報量基準 AIC 最小化原理に基づいて多項式の次数選択が行なえることを示した。
- (3) 予測された位置を左右の2次元画像平面に投影し、画像平面に設定した狭い探索空間だけを対象とした対応点探索で連続した追跡が可能であることを確認し、探索空間を狭くすることにより高速に処理できることを示した。

このように、ステレオ動画像から獲得した物体上の特徴点を3次元空間上の動きとしてとらえ、軌跡の多項式近似や対応点探索を行なうことにより、計算量が多く困難と考えられていたステレオ動画像からの3次元情報抽出処理をリアルタイムに処理可能であることを示した。

しかし、提案手法を画像解析や追跡処理に適用するためには、フレーム間対応点探索に使用する基準画像の量子化誤差による対応点のずれを補正する手法の導入が必要である。また、追跡開始直後のフレーム数が少ない時に正確な軌跡を多項式近似できないため、広範囲の対応点探索が必要になる。このため、多くの計算時間が必要になることから、処理速度の向上策が必要である。

また、追跡対象がオクルージョンにより見えなくなった場合などの予測追跡機能の導入や、剛体仮定による複数の特徴点追跡の精度向上なども今後研究して行く必要がある。

謝辞

本研究を進めるにあたり、終始暖かく御指導を頂いたソフトウェア基礎講座 横矢 直和教授 に厚く御礼申し上げます。

副指導教官として御助言頂いた像情報処理学講座 千原 國宏教授、並びに情報科学センター 荻原 剛志助教授に深く感謝いたします。

本研究への助言や環境の整備に協力して頂いたソフトウェア基礎講座 竹村 治雄助教授、情報科学センター 馬場 始三助手、岩佐 英彦助手、ソフトウェア基礎講座 片山 喜章助手に深く感謝いたします。

本学においての研究の機会を与えて下さいました三菱電機エンジニアリング（株）鎌倉事業所 中村 泰而所長、野明 宏亙副所長、情報電子機器部 横山 繁盛部長、計算機技術部 戸根 吉彦担当部長 に感謝いたします。

最後に、ソフトウェア基礎講座の諸氏には、本研究の様々な局面にて多大なる御協力いただきました。ここに感謝の意を表します。

参考文献

- [1] 渡邊 睦, 小野口 一則, 久野 義徳, 麻田 治男: “視差予測ステレオ法を用いた障害物検出”, 信学論 (D-II), Vol.J73-D-II, No.6, pp.862-870, 1990.
- [2] 安居院 猛, 日野 知明, 長橋 宏: “ステレオ動画像による多動物体の運動認識における一手法” 信学技報, PRU93-56, pp.17-22, 1993.
- [3] 山本 正信: “連続ステレオ画像からの3次元情報の抽出” 信学論 (D), Vol.J69-D, No.11, pp.1631-1638, 1986.
- [4] B. K. P. Horn: “Closed-form solution of absolute orientation using unit quaternions”, *J. Opt. Soc. Am. A*, Vol.4, No.4, pp.629-642, 1987.
- [5] O. D. Faugeras and M. Hebert: “The representation, Recognition, and Locating od 3-D Objects”, *Int. J. Robotics Res.*, Vol. 5, No.3, pp.27-52, 1986.
- [6] 三池 秀敏, 古賀 和利: パソコンによる動画像処理, 森北出版, 1993.
- [7] 古賀 和利, 三池 秀敏: “動画像の時空間相関に基づくオプティカルフローの解析”, 信学論 (D-II), Vol.J72-D-II, No.4, pp.507-516, 1989.
- [8] 志沢 雅彦, 間瀬 健二: “多重オプティカルフロー — 基本拘束方程式と運動透明視・運動境界検出の統一計算理論 —”, 信学論 (D-II), Vol.J76-D-II, No.5, pp.987-1005, 1993.
- [9] 金谷 健一: “動画像からの3D コンピューティング”, 情処研報, コンピュータビジョン, CV 81-6, pp.43-50, 1993.
- [10] M. Jenkin and K. Tsotsos: “Applying temporal constraints to the dynamic stereo problem”, *Computer Vision, Graphics and Image Processing*, Vol.33, pp.16-32, 1986.
- [11] 山本 正信, ロベルト チポラ: “ステレオ動画像解析の一手法”, 信学論 (D-II), Vol.J72-D-II, No.6, pp.855-865, 1989.

- [12] 阿部 恭一, 横矢 直和: “探索平面での動的輪郭モデルを用いたステレオマッチング”, 第25回画像工学コンファレンス論文集, 2-6, pp.57-60, 1994.
- [13] 大崎 善彦, 池田 謙三, 山本 正信: “ステレオ動画像を使った物体の追跡”, 94 信学春季大会, D-557, 1994.
- [14] 田村 洋, 横矢 直和: “ステレオ動画像からの運動予測に基づく物体の追跡”, 情処第 49 全大会, 1F-4, 1994.
- [15] 横矢 直和: “知能処理特論 I”, 奈良先端科学技術大学院大学 知能処理特論 I テキスト, 1993.
- [16] B. K. P. Horn (NTT ヒューマンインタフェース研究所プロジェクト RVT (訳)): ロボットビジョン –機械は世界をどう見るか–, 朝倉書店, 1993.
- [17] 田村 進一: FA のための画像処理技術, 工学研究社, 1988.
- [18] 画像処理ハンドブック編集委員会: 画像処理ハンドブック, 昭晃堂, 1987.
- [19] 奥富 正敏, 吉崎 修, 富田 剛司: “カラーステレオマッチングとその視神経乳頭 3 次元計測への応用”, 信学論 (D-II), Vol.J76-D-II, No2, pp.342-349, 1993.
- [20] 赤池 弘次: “情報量基準 AIC とは何か –その意味と将来への展望–”, 数理科学, No.153, pp.5-11, 1976.
- [21] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery: *Numerical Recipes in C*, Cambridge University Press, New York, second edition, 1992.
- [22] H. P. Moravec: “Visual mapping by a robot rover”, *Proc. 6th IJCAI*, pp.598-600, 1979.