

NAIST-IS-MT9451073

## 修士論文

# 2段階アフィン変換を利用した 移動カメラ画像中の動物体抽出

寺久保 敏

1996年2月16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  
修士(工学) 授与の要件として提出した修士論文である。

寺久保 敏

指導教官： 横矢 直和 教授  
千原 國宏 教授  
竹村 治雄 助教授

# 2段階アフィン変換を利用した 移動カメラ画像中の動物体抽出\*

寺久保 敏

## 内容梗概

近年、移動ロボットの自律化や、乗物の自動操縦や安全装備の充実を目標とした研究が盛んになり、カメラの移動パラメータが取得困難な環境で、物体認識能力が要求されるようになってきた。動画像中から動物体を抽出することを考える時、カメラに動きが伴う場合は、実際の動物体以外の背景部分にもカメラの移動やカメラパラメータの変化によって見かけ上の動きが生じてしまい、動物体の特定が困難になる。

本論文では、移動カメラで撮影した動画像中から、静止背景の見かけの動きを除去し、動物体のみを抽出する方法について述べる。カメラの動きやカメラパラメータの変化についての情報は未知として、まず、連続する画像フレーム間のマッチング結果をもとに、アフィン変換を2段階に適用することで背景領域の位置合わせを行なう。次に、位置合わせ後の時間軸方向での差分画像から動物体領域の抽出を行なう。最後に実画像を用いた実験によって提案手法の有効性を検証する。

## キーワード

動画像処理, アフィン変換, 動物体抽出, 移動撮像系, 見かけ上の背景の動き

---

\*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-MT9451073, 1996年2月16日.

# Detection of Moving Objects in Moving Camera Images by Two-step Affine Transforms\*

Satoshi Terakubo

## Abstract

Recently, many researches for autonomous control of mobile robots or cars are being carried out. For such applications, it is required to detect moving obstacles from images captured by a moving camera, whose motion parameters are difficult to be estimated accurately. Image sequences captured by a moving camera include apparent motions of a static background. Such motions make it difficult to detect moving objects in a scene.

This paper describes a new approach for motion detection from image sequences captured by a moving camera with unknown parameters. Making an assumption that the apparent motion of the static background is dominant in an image, we determine affine transformation parameters between consecutive image frames in two steps for registering the static background. To detect moving object regions, the previous frame is transformed with computed affine parameters, and subtracted from the subsequent frame. We have successfully demonstrated the effectiveness of the method using real dynamic images.

## Keywords:

dynamic image analysis, affine transformation, motion detection, moving camera image, apparent motion of background

---

\*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9451073, February 16, 1996.

# 目次

# 目 次

# 表 目 次

# 1. 序論

近年、計算機の処理能力の飛躍的な向上に伴い、様々な分野において、膨大な情報を含む動画像を対象とした研究が注目されるようになってきた。特に、自動車などの乗物の安全装備の充実や自動操縦の実現を目指した研究 [1][2] や、各種作業ロボットの自律化を目標として、人間の目に相当する視覚センサーを実現しようとする研究が盛んになっている [3][4]。

これらの研究では、移動体に搭載されたカメラで撮影した動画像を利用して、周辺環境や障害物を認識することが必要になってくるが、移動体が動くと同時にカメラそのものも動くため、本来静止している背景領域にも見かけ上の動きが生じ、画像中に存在する動物体を抽出するのが困難である。ビデオカメラの微小なブレを補正して被写体の見かけ上の動きを補正する技術は、すでに市販品に実装されているものもあるが [5]、レンズのズームやカメラの回転等には原理的に対応できていない。従来から、これらの問題を解決するための試みとして、種々の方法が提案されている [6][7][8][9][10][11]。

代表的な方法として、センサー等で得られたカメラの運動情報を利用する方法が提案されている [12]。この方法では、カメラを回転運動が可能な台に固定して映像を撮影し、台に取り付けられているポテンショメータから得られる信号をもとにカメラの回転角度を検出して、計算によって背景の見かけ上の動きをキャンセルすることに成功している。この方式は比較的簡単に精度良く実現できるが、カメラの動きをモニターするセンサーが必要なため機構が大がかりになるという問題や、水中や空中、または滑らかでない路面を走行するロボットなど、カメラの姿勢を精度良く検出する術を持たない環境では適用できないという問題がある。

これに対して、カメラパラメータ等の外部情報を必要とせず、画像のみを使用する方法もいくつか提案されている。この種の方法で最も代表的なものは、背景の見かけ上の動きを画像面上の支配的な動きとみなし、その動きと異なる動きを示す領域を動物体として検出する方法である。たとえば、カメラの動きによって生じるオプティカルフローを計算し、それらが一つの消失点方向に収束する性質を利用して動物体を検出する方法 [13][14] が提案されているが、動物体の運動方向がカメラの動きと同じ場合には原理的に検出が不可能である。また、フレーム間



の背景の見かけ上の動きを画像中の支配的な動きとみなし、その動きをアフィン変換で近似して、計算で求まるアフィン変換パラメータをもとに背景部分の位置合わせを実現する方法も提案されているが [1][2][3]、これらの方式では動物体領域も含んだ全画素を対象に変換パラメータを計算しているため、背景領域のみの変換が精度良く実現できていない。そのため、動物体を特定するために複雑な後処理が必要になるという問題や [4]、反復演算で精度を上げようとしているため膨大な処理時間を要したり、初期パラメータの精度に結果が左右されるという問題がある [5][6]。

この他にも、被写体に対して特定の方向からのみ照明が当たっているという前提のもとで画像を解析し、特定された照明方向を画像フレーム間で合わせたり [7]、画像中の複数の被写体位置のトポロジー変化からカメラの移動を推定して [8]、背景部分の位置合わせを実現する方法も提案されているが、いずれも画像撮影時の環境状態によって結果が大きく左右され、安定した結果が得られていない。

本論文では、これらの問題点を解決する方法として、2段階にアフィン変換を適用して入力画像中の背景領域のみの位置合わせを実現できる、動物体抽出手法を提案する [9] [10] [11]。提案手法では、入力される1枚目の画像上に設定された複数の注目点と、2枚目の画像上で探索された対応点の組を1段階目のアフィン変換を適用して評価することで、動物体上に設定された注目点や誤対応が生じた注目点の排除を実現する。次に、こうして選択された背景上の注目点とその対応関係を用いて2段階目のアフィン変換を適用し、背景領域のみの位置合わせを可能にする。さらに、背景位置合わせ後の差分画像を時系列にまとめて評価することで、動物体領域のみの抽出を行なう。

以下、第2章では本研究の基礎となるアフィン変換を利用した画像位置合わせ法について説明し、第3章、第4章で提案手法の詳細、実験結果について述べ、第5章で本論文の結論と今後の課題について述べる。

## 2. アフィン変換を利用した画像位置合わせ

本章では、提案手法の基礎となるアフィン変換を利用した画像位置合わせ法について述べる。まず始めにアフィン変換そのものについて説明し、続いて画像位置合わせへの応用について簡単に述べる。

### 2.1 アフィン変換

画像の歪み・変形は、濃度値に関する歪み (grey level distortion) と像空間の空間的な幾何学的歪み (geometric distortion) に分けられる [?]。アフィン変換とは、後者に属する歪みである平行移動、回転、拡大縮小、およびこれらの要因が複数絡んだ幾何学的線形歪みを表すことができる一般1次変換である (図 2.1参照)。

もともとなる画像上の座標系を  $(x, y)$  で表し、変形された画像上の座標系を  $(u, v)$  で表すと、画像の変形がアフィン変換で表現できればこれらの関係は式 (1) で表すことができる。

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (1)$$

ここで、 $\{a_{ij}\}$  は画像の回転歪みおよび拡大縮小を表し、 $\{b_i\}$  は画像の平行移動を表している。両者を併せてアフィン変換パラメータという。アフィン変換は縦横縮尺比が異なる変換が可能であり、遠近法的歪みを近似的に補正する場合や、文字認識における文字領域のずれを検出する目的 [?] 等にも幅広く利用されている。

### 2.2 画像位置合わせ法

2枚の画像の歪みを補正して、画像位置を合わせる目的でアフィン変換を適用するには、2枚の画像間のアフィン変換パラメータを求めなければならない。以下、アフィン変換パラメータを求めるために必要な、画像間の対応点探索法と、アフィン変換を用いた画像位置合わせ法について簡単に述べる。

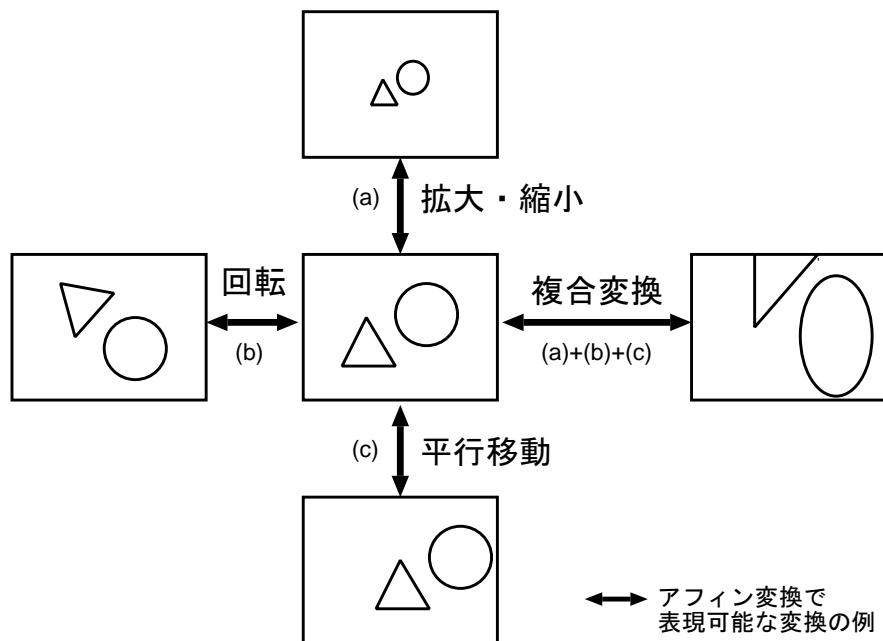


図 2.1 アフィン変換の概念

### (1) 画像間の対応点探索

2枚の画像間で、対応する点を探る代表的な手法に、テンプレートマッチングがある[?]。テンプレートマッチングは図 2.2 に示すように、既知の対象を表すテンプレート  $t(x, y)$  を動かしながら画像  $I(x, y)$  の各点での類似度を計算し、画像内でテンプレートと最も良く一致する部分を見つける操作である。画像の類似度の計算には式 (2)~(4) で示されるような画像内の各点でのテンプレートとの差異の尺度を用いるのが一般的である。

$$d_1(I, t; i, j) = \max_{(x, y) \in D} |I(x + i, y + j) - t(x, y)| \quad (2)$$

$$d_2(I, t; i, j) = \int \int_D |I(x + i, y + j) - t(x, y)| dx dy \quad (3)$$

$$d_3(I, t; i, j) = \int \int_D \{I(x + i, y + j) - t(x, y)\}^2 dx dy \quad (4)$$

ただし、 $D$ はテンプレート  $t$  の定義域である。 $d$  の値が小さいほど類似度が高いと考えられる。2枚の画像間のアフィン変換パラメータを求めるために、提

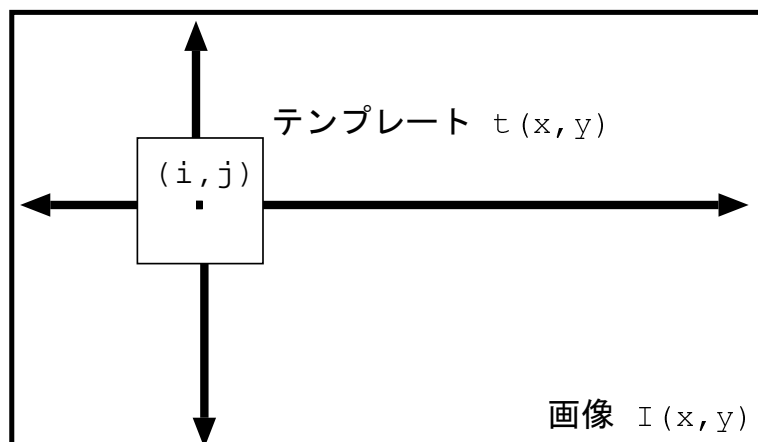


図 2.2 テンプレートマッチングの概念

案手法ではアフィン変換のもとになる画像上からテンプレートを切り出し、他方の画像上にテンプレートマッチングの定義域を設定して、対応点探索を行っている。

テンプレートマッチングは画像とテンプレートが大きい場合に計算量が大きくなるので、高速化の方法として、ピラミッド画像を用いた粗精探索法や、残差  $(d_k)$  を累積してゆく過程である閾値を越えた時点で計算を打ち切る SSDA 法 (Sequential Similarity Detection Algorithm) などが考えられている [?]。提案手法では処理時間の短縮を目指し、対応点探索手法として SSDA 法を採用している。

## (2) アフィン変換パラメータの獲得と画像位置合わせ

式 (1) に示すように、アフィン変換パラメータは全部で 6 つある。よって、もとになる画像上の注目する画素の座標  $(x, y)$  と、それに対応する変形された画像上の画素の座標  $(u, v)$  が 3 組以上求まれば、6 元 1 次方程式の解としてアフィン変換パラメータを求めることができる。実際の画像を用いた処理においては、4 組以上の対応が求まった場合は最小自乗法で計算するのが一般的である。

画像の位置合わせを行なう場合には、こうして求めたアフィン変換パラメータ、 $\{a_{ij}\}$ 、 $\{b_i\}$  を式(1)に代入して変換関数を獲得し、入力画像上の全ての画素をこの関数に入力して座標変換すれば良い。

### 3. 移動カメラ画像からの動物体の抽出

本章では下記の順番で、移動カメラで撮影した動画画像から動物体を抽出する提案手法について説明する。

#### 1. 提案手法の概要および前提条件

提案手法の概要と、提案手法が有効に機能するためのカメラの動きや被写体に対する前提条件について述べる。

#### 2. 背景上に設定された注目点の選択

入力フレーム画像上に設定された複数の注目点と、その次の時刻の入力フレーム画像上での対応点探索結果から、背景上に設定され、かつ対応点が正しく求まっている注目点を1段階目のアフィン変換を適用して選択する方法を提案する。

#### 3. 動物体領域の抽出方法

選ばれた背景上の注目点を用いた2段階目のアフィン変換により入力フレーム画像間の背景領域のみの位置合わせを行ない、背景位置合わせ後の画像列から動物体領域を切り出す方法を提案する。

### 3.1 提案手法の概要

提案手法は、第1章で述べた従来提案されている手法の問題点を全て解決することができるものであり、

- カメラの運動やレンズのズームング等の情報が不要
- 動物体領域を含まない、背景領域のみの位置合わせが可能

の2つの特徴を有している。

以下、提案手法の前提として仮定している条件について説明し、提案手法について処理のステップ毎に簡単に説明する。

## (1) 前提条件

提案手法では前提条件として以下に示す3つの仮定をおき、これらを全て満たす条件で撮影された画像のみを処理の対象としている。

1. カメラの移動速度は、フレーム間で視点が極端に変わらない程度とする
2. カメラに極端に近い被写体は無く、背景領域のフレーム画像間の対応はアフィン変換で近似できる
3. 動物体領域は背景領域に比べて小さい

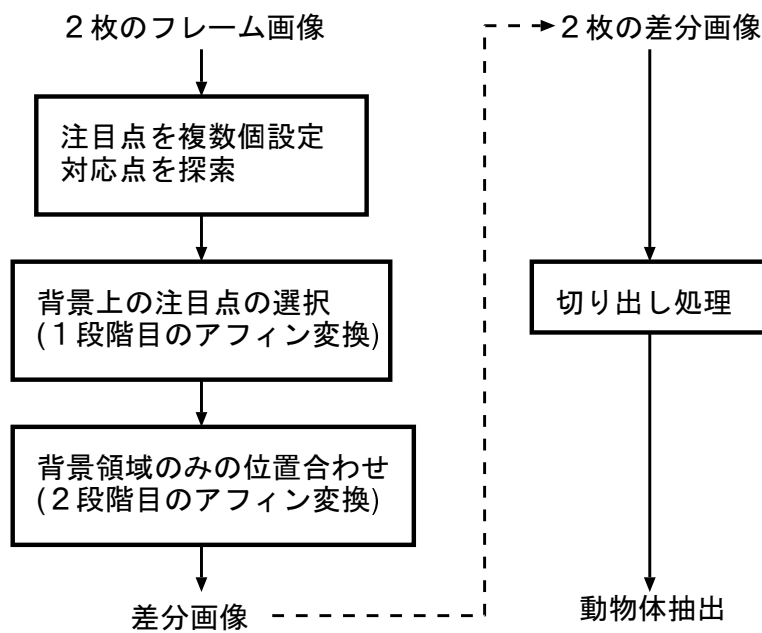


図 3.1 処理の流れ

## (2) 処理の流れ

提案手法の処理の手順を以下に示し、処理の流れの概略を図 3.1 に示す。

- (a) 入力される 1 枚目のフレーム画像上に規則的に複数個注目点を設定する。
- (b) その注目点の対応点を 2 枚目の画像上で、テンプレートマッチングを利用して探索する。
- (c) 対応点探索結果をもとに 1 段階目のアフィン変換を複数回適用して、変換誤差に基づいて、背景上に設定された注目点を選択する。
- (d) 選択結果を用いて 2 段階目のアフィン変換を適用し、背景領域のみの位置合わせを実現し、差分画像を作成する。
- (e) 得られた差分画像を時系列に 2 枚使用して、動物体領域の切り出しを行なう。

以下の節では各ステップの詳細について述べる。

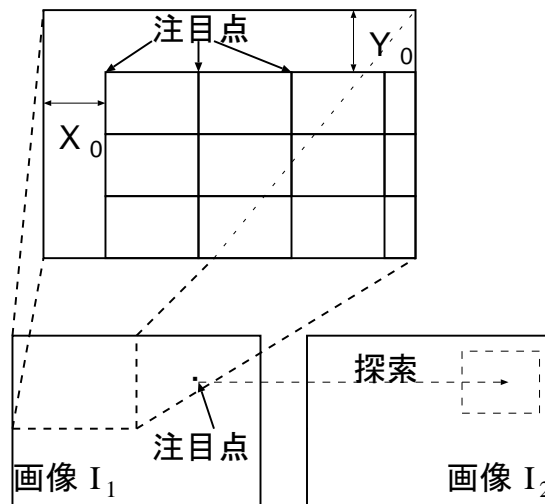


図 3.2 注目点の設定と対応点探索

### 3.2 注目点の設定と対応点探索

入力として用いる画像列を、 $I_j \{j = 1, 2, \dots\}$  とする。まず、図 3.2 に示すように  $I_1$  上に注目点を等間隔に複数個格子状に設定し、 $I_2$  上で対応する点をテンプレ



レートマッチングを利用して探索する。ただし、撮影時にカメラの移動やレンズのズームが生じることを考慮し、その様な場合にも注目点の対応点が  $I_2$  内に収まるように、 $I_1$  の外枠から  $X_0$ 、 $Y_0$  の範囲を注目点を設定しない領域 (オフセット領域) とする。

### 3.3 注目点の選択

前節で設定された注目点の中から、以下の手順により背景の位置合わせに使用できる注目点の選択を行なう。

- i. 注目点を図 3.3 に示すように規則的に三角形形状に組み合わせる。
- ii. 全ての三角形  $T_i \{i = 1, 2, \dots, n, n \text{ は三角形の数}\}$  について、その 3 頂点位置の注目点の対応結果をもとに、 $I_1$ 、 $I_2$  間のアフィン変換パラメータを計算する。
- iii. 求めたアフィン変換パラメータを用いて  $I_1$  をアフィン変換する。(1 段階目のアフィン変換)
- iv. アフィン変換された画像と  $I_2$  から、式 (5) で定義される重なり具合を評価する指標  $E_i$  を計算する。
- v.  $\{E_i\}$  をある適当な値で閾値処理し、残った  $E_i$  を求める時に用いた注目点を選出する。

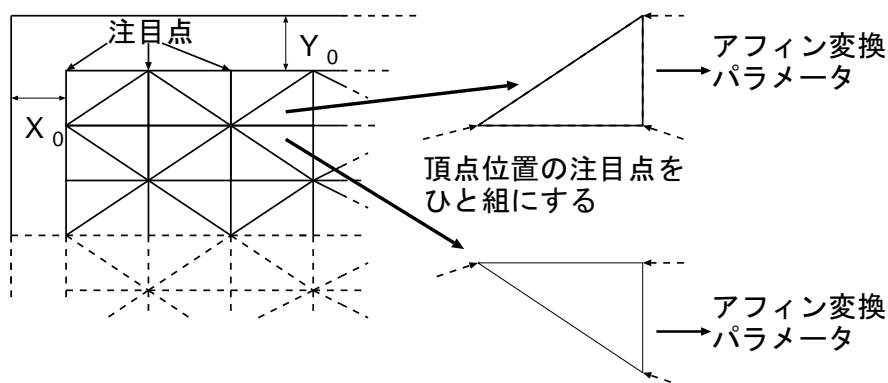


図 3.3 注目点の組み合わせ

3.1節で述べた仮定 1 ~ 3 より、 $T_i$ の 3 頂点位置の注目点のうち一つでも動物体上に設定されたり、ノイズ等の影響で誤対応が生じた場合は、全ての点が背景上に設定されて誤対応が無い場合に比べて、アフィン変換後の画像の重なり具合が悪くなると考えられる。よって、重なり具合を評価すれば背景位置合わせに使用できる注目点が選択可能である。重なり具合を評価する指標として、式 (5) で定義される  $E_i$  を使用した。

$$E_i = \sum_{(x,y) \in R} \frac{|I_1^i(x,y) - I_2(x,y)|}{r} \quad (5)$$

ただし、

$I(x,y)$ : 画像  $I$  上の座標  $(x,y)$  の画素値

$I^i$ : 三角形  $T_i$  の対応に基づく 1 段階目のアフィン変換画像

$R$ :  $I^i$  内での有効画素の集合

$r$ :  $R$  の要素数 (有効画素数)

をそれぞれ表す。ここで言う有効画素とは、アフィン変換の結果、画素値が計算できた画素のことである。 $E_i$  は、 $T_i$  をもとに計算されたアフィン変換パラメータを用いて  $I_1$  を変換 (1 段階目のアフィン変換) し、得られた画像  $I_1^i$  と  $I_2$  との画素毎の差分の絶対値の平均値を表している。

この処理の概念図を図 3.4 に示す。 $E_i$  の値は、 $I_1^i$  と  $I_2$  の重なり具合が良好な場合に小さな値を示し、逆に重なり具合が悪い場合には大きな値を示すと考えられるので、 $\{E_i\}$  を適当な値で閾値処理すれば、背景位置合わせに使用できる注目点の選択が可能である。3.1節の仮定 3 から、動物体上に設定される注目点数は最大でも全注目点の半数以下であり、また、対応点探索においては大多数が誤対応無く求まると考えられるので、ここでは  $\{E_i\}$  のメディアンを使用して閾値処理を行なうこととする。閾値処理で残った  $E_i$  を求める時に用いた注目点を選出すれば、背景上に設定され、かつ対応が正しく求まった注目点のみを選択することができる。

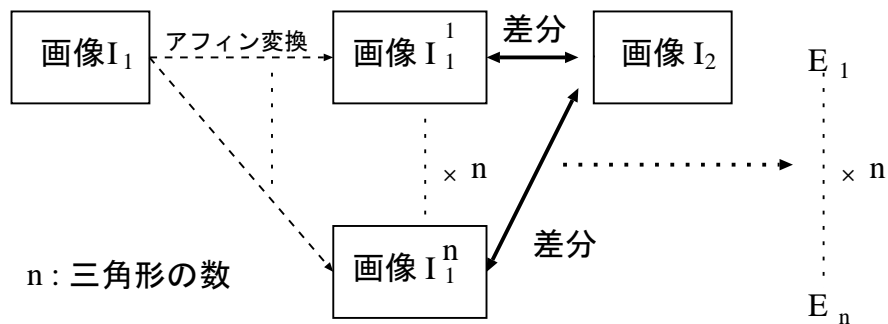


図 3.4 1 段階目のアフィン変換の適用

### 3.4 背景領域の位置合わせ

選択された全ての注目点とその対応点をもとに、再度  $I_1$ 、 $I_2$  間のアフィン変換パラメータを計算し、 $I_1$  にアフィン変換（2 段階目のアフィン変換）を施し、 $I_2$  との画素毎の差分の絶対値を画素値として持つ差分画像を作成する。こうすることで、背景領域のみの位置合わせを実現することができ、主に動物体領域の画素が大きな値をもつ差分画像が得られる。

### 3.5 動物体領域の切り出し

3.4 節で作成した差分画像の 2 値化を行ない、動物体切り出し用のマスク画像を作成する。差分画像では、大きな画素値を持つ領域は主に  $I_1$  か  $I_2$  内の動物体領域上にあると考えられるので、差分画像上の画素値を適当な値で閾値処理すれば、動物体領域を切り出すことができる。ただし、このままでは  $I_1$  内の動物体と  $I_2$  内の動物体に起因する領域が両方とも切り出されてしまうことになるので、 $I_2$ 、 $I_3$  をもとに作成された差分画像と統合して切り出し精度を上げる。以下、この方法について詳しく説明する。

### 3.5.1 動物体切り出し用マスク画像の作成

3.4節で作成した差分画像を  $S$ 、 $S$  の座標  $(x, y)$  における画素値を  $S(x, y)$ 、画素値のソート用の配列を  $X$ 、動物体切り出し用のマスク画像を  $M$  とし、以下に  $M$  を作成する手順を示す。

- i.  $S(x, y)$  を小さい順にソートして配列  $X$  に代入する。
- ii. 式 (6) に示される条件でマスク画像の画素値  $M(x, y)$  を決定し、マスク画像  $M$  を作成する。
- iii. マスク画像  $M$  に収縮・膨張処理を施し、微小な誤検出領域を消去する。
- iv. さらに、膨張・収縮処理を施し、分かれて抽出された同一の動物体領域を統合する。

マスク画像  $M$  の画素値  $M(x, y)$  は、式 (6) で定義できる。

$$M(x, y) = \begin{cases} 1, & \text{if } S(x, y) > X[q]; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

ここで、 $q$  は 2 値化のレベルを表す変数であり、式 (7) で定義する。

$$q = \left( \frac{P_b + k}{P_a + k} \right) \cdot m \quad (7)$$

ただし、

$P_a$  : 注目点の総数

$P_b$  : 選択された注目点数 (背景上の注目点数)

$k$  : 補正係数

$m$  : 画像中の全画素数

をそれぞれ表す。

3.1節で述べた仮定 2 より、透視投影歪みを含むフレーム画像間の対応をアフィン変換で近似していることから、差分画像には原理的にわずかなキャンセルエラー

が残り、マスク画像上にも誤検出領域が存在する。そこで、このエラーを除去するために収縮・膨張処理を施し、また、分かれて抽出された同一の動物体領域を統合するために膨張・収縮処理を施し、最終的な動領域切り出し用のマスク画像を作成する。

ただし、ここで言う収縮・膨張処理は、式 8) で表される収縮フィルタと、式 9) で表される膨張フィルタ [?] をこの順で施す処理を意味しており、膨張・収縮処理は逆の順で施す処理を意味している。

$$g_{ij} = \min_{(k,l) \in W_{ij}} \{f_{kl}\} \quad (8)$$

$$g_{ij} = \max_{(k,l) \in W_{ij}} \{f_{kl}\} \quad (9)$$

ただし、

$f_{kl}$  : 処理前の画像の座標  $(k, l)$  における画素値

$g_{ij}$  : 処理後の画像の座標  $(i, j)$  における画素値

をそれぞれ表し、 $W_{ij}$  は式 10) で定義される座標  $(i, j)$  を中心とする  $m \times n$  の局所領域である。

$$W_{ij} = \left\{ (k, l) \mid i - \left\lceil \frac{m-1}{2} \right\rceil \leq k \leq i + \left\lfloor \frac{m}{2} \right\rfloor, j - \left\lceil \frac{n-1}{2} \right\rceil \leq l \leq j + \left\lfloor \frac{n}{2} \right\rfloor \right\} \quad (10)$$

ただし、 $\lceil \cdot \rceil$  はガウス記号を示す。

### 3.5.2 処理結果の統合

前節までの処理によって得られたマスク画像は、入力として用いた2枚の画像  $I_1$ 、 $I_2$  の背景位置合わせ後の変化領域をもとに作成されている。すなわち、このマスク画像をそのまま使用して動物体領域を切り出そうとすると、動物体の周辺の領域も余分に切り出されてしまうことになる。そこで、時系列に得られる2枚のマスク画像を統合することで、両方のマスク画像で共通して動物体領域として

判定されている領域  $M(x, y) = 1$  の画素) のみを切り出せる新たなマスク画像を作成し、動物体の切り出し精度を上げる。この処理の概念図を図 3.5 に示す。

$I_1$  と  $I_2$  を基に計算されたマスク画像を  $M_{12}$ 、 $I_2$  と  $I_3$  を基に計算されたマスク画像を  $M_{23}$  とすると、 $I_2$  上の動物体を切り出すマスク画像  $M_2$  の画素値  $M_2(x, y)$  は式 (11) で表せる。

$$M_2(x, y) = M_{12}(x, y) \times M_{23}(x, y) \quad (11)$$

こうして求めたマスク画像  $M_2$  を  $I_2$  と画素毎に乗算することで、動物体領域のみを切り出すことができる。

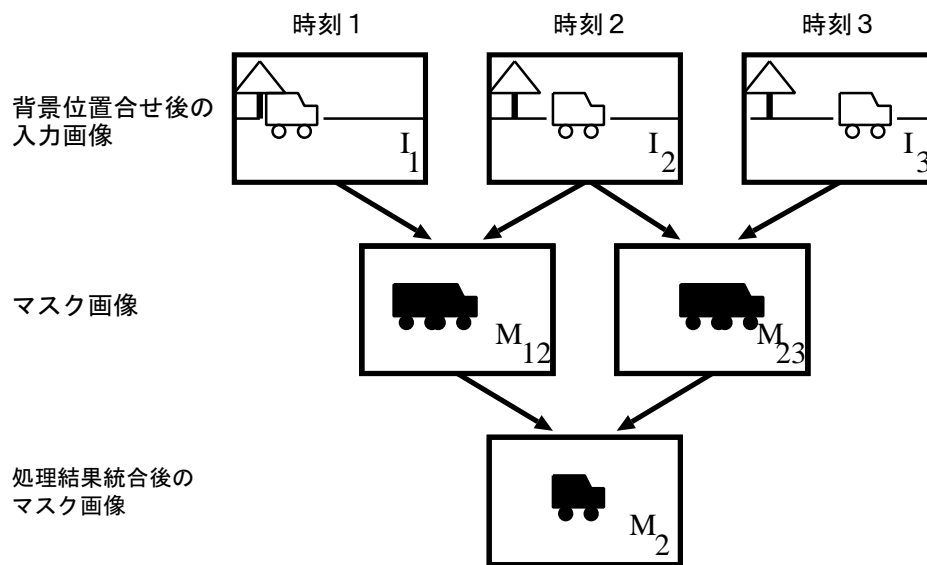


図 3.5 処理結果の統合の概念図

## 4. 実験および結果

### 4.1 入力画像の準備

#### 4.1.1 画像入力装置および撮影条件

実験用入力画像の撮影には、SONY 製のビデオカメラ (Handycam PRO、41 万画素の CCD × 3、Hi8 方式) を使用した。また、撮影した画像を電子ファイルとして取り込むのに、SGI 製の Sirius Video 装置を利用した。本装置は 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することができ、格納した画像をハードディスクに連続して記録することができる。

実験用画像の撮影にあたっては、ビデオカメラを手で持ち、首振り、光軸回りの回転、レンズのズーム等動きを意図的に与えながら動物体を撮影し、故意に背景の見かけ上の動きを生じさせた。本論文の実験で使用した 3 種類の画像の撮影条件を表 4.1 に示す。

表 4.1 実験で用いた画像の撮影条件

	レンズの動き	カメラの動き	備考
画像列 (a)	ズームイン	手ブレ程度	
画像列 (b)	なし	光軸回りの回転 首振り運動	
画像列 (c)	ズームアウト	首振り運動	動物体の交差あり

#### 4.1.2 画像の前処理

撮影された生の画像に対して以下のような前処理を行なった。

##### (1) 分割・縮小

画像入力に使用した Sirius Video 装置のアナログ・ビデオ・インタフェースは、1 フレーム内に時系列の異なる 2 枚の画像が含まれるインタレース信号で

ある。そこで、実験で使用する画像が同一時刻に撮影された画素のみで構成されるようにするため、図 4.1 に示すようなフィールド単位の画像分割と、走査線方向に 1/2 に縮小する処理を行なった。

## (2) カラー画像から濃淡画像への変換

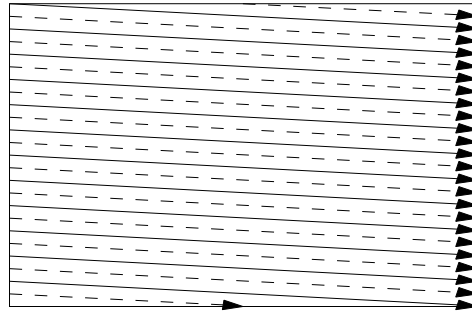
分割・縮小して得られた RGB 表色系のカラー画像を、カラーテレビ放送用の標準表色系である YIQ 表色系 [?] に変換し、Y (輝度) 値を画素値とする濃淡画像を作成して実験に用いた。式 (12) に RGB 表色系から YIQ 表色系への変換式を示す。

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.522 & 0.311 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (12)$$

上記の処理結果の例を図 4.2 に示す。実際のプログラムに画像を供する時は、フィールド毎に分割後作成される 2 種類の画像列のうち、統一してどちらか一方を選んで使用した。

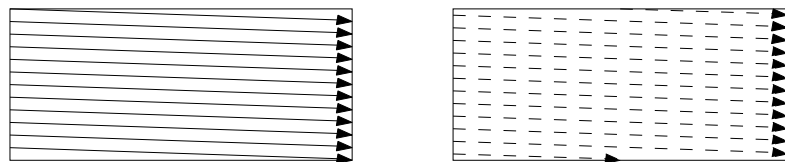


カメラで撮影された画像  
(インタレース)



分割

Sirius Video で取り  
込んだ分割画像



横幅  $\times 1/2$

実験で使  
用した画像



図 4.1 入力画像の前処理方法



Sirius Video で取り込んだカラー画像の例



実験に使用した濃淡画像の例

図 4.2 入力画像の前処理結果

## 4.2 実験概要と実行環境

今回の実験で用いた画像仕様やプログラム中で設定したパラメータ等は、表 4.2 に示す通りである。

表 4.2 画像仕様・実験パラメータ一覧

項目	パラメータ・仕様等
画像サイズ	360 × 243 画素
濃淡レベル	256 階調
初期注目点数	36 個 (6 × 6)
オフセット値 ( $X_0$ )	画像の幅の 10 %
オフセット値 ( $Y_0$ )	画像の高さの 10 %
対応点探索窓	10 × 10 画素
対応点探索範囲	30 × 30 画素
注目点選択時の閾値	$\{E_i\}$ のメディアン
マスク画像作成用補正係数 ( $k$ )	28
収縮窓サイズ	4 × 4 画素
膨張窓サイズ	6 × 6 画素
処理間隔	2 フレーム: 画像列 ( $a$ ), ( $b$ ) 1 フレーム: 画像列 ( $c$ )

今回の実験では注目点の対応点探索手法として、SSDA 法 (残差逐次検定法) を利用して高速化を図った。また、差分画像 2 値化の閾値決定時の補正係数  $k$  の値は、今回注目点を設定した 6 × 6 箇所の外側に、ひと回り多く注目点を設定した場合に余分に設定できる注目点数として 28 を採用した。

画像処理に使用した計算機は、SGI 製の ONYX (88 SPECint92, 97 SPECfp92 × 8CPU) で、並列処理は全く利用せずに作成したプログラムを実行させた。1 フレーム当りの処理に要した平均実行時間を表 4.3 に示す。

表 4.3 1 フレーム当りの平均実行時間

処理内容	必要回数	所要時間
注目点設定 対応点探索	1 回	19 秒
1 段階目アフィン変換	1 回	19 秒
2 段階目アフィン変換	2 回	0.7 秒
動物体切り出し処理	1 回	5 秒
合計		約 44 秒

### 4.3 対応点の探索および注目点の選択

図 4.3～4.5 に 3 種類の入力画像列の一部をそれぞれ示す。図 4.3 はレンズをズームインさせながらリヤカーを引いている人間を、図 4.4 はカメラを光軸周りに回転させながら右へ首振りして走行中の自動車を、図 4.5 はレンズをズームアウトさせながらカメラを右へ首振りして走行中の複数の自動車を、それぞれ撮影した画像である。

図 4.6～4.8 に、図 4.3～4.5 に示した入力画像上の注目点の設定状況と対応点探索によって求めた速度ベクトル、および 1 段階目のアフィン変換による注目点選択状況をそれぞれ示す。図中、1 段階目のアフィン変換の結果、背景上に設定され、対応点が正しく求めたと判定された注目点を白色で表示し、動物体上に設定されているか対応点探索で誤対応が生じたと判定された注目点を赤色で表示している。なお、注目点を見やすくするために図 4.6～4.8 の処理画像は暗く表示している。また、図??～?? には、3 種類の入力画像列の全フレームを処理した結果、選択された（背景位置合わせに使用する）注目点数の変化を示す。以上の結果から、各画像列ともに誤選択された注目点は無く、背景位置合わせに使用する注目点の位置や数は動的に変化していることがわかる。



第1フレーム



第8フレーム



第15フレーム



第22フレーム



第29フレーム



第36フレーム

図 4.3 入力画像 画像列(a) :ズームイン



第1フレーム



第7フレーム



第13フレーム



第19フレーム



第25フレーム



第31フレーム

図 4.4 入力画像 画像列 (b) :回転・首振り)



第1フレーム



第11フレーム



第21フレーム



第31フレーム



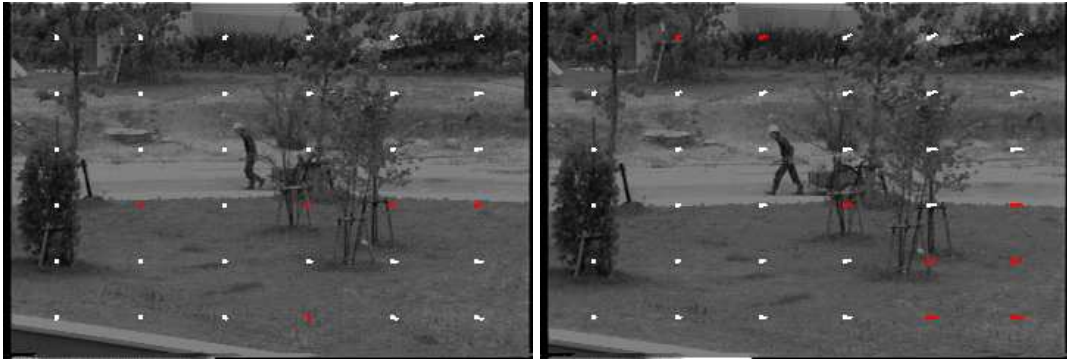
第41フレーム



第51フレーム

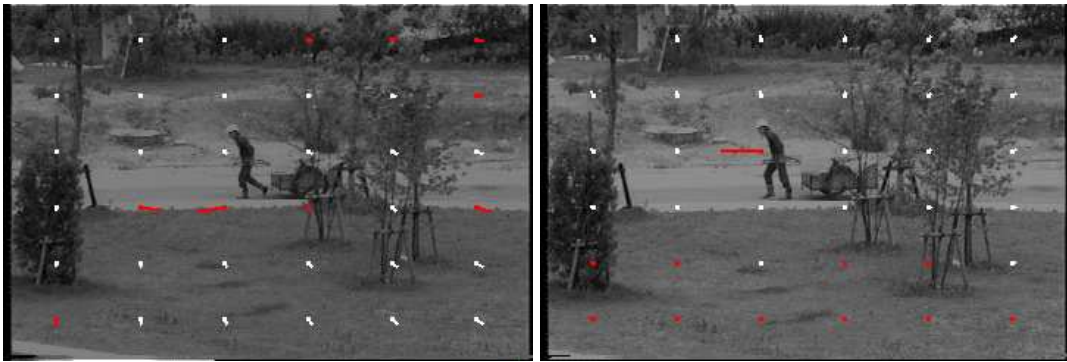
図 4.5 入力画像 画像列(c) :ズームアウト・首振り、複数の動物体)





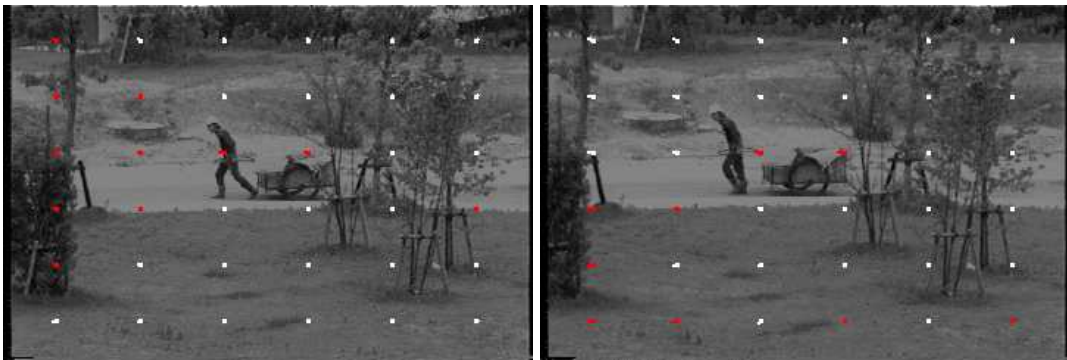
第1フレーム

第8フレーム



第15フレーム

第22フレーム



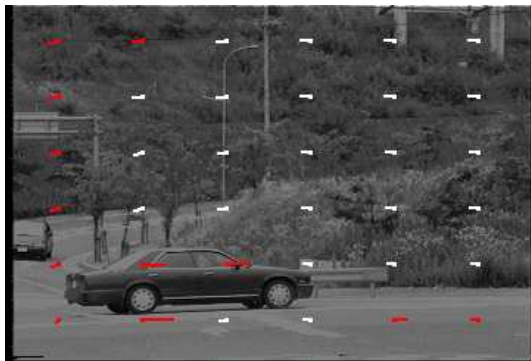
第29フレーム

第36フレーム

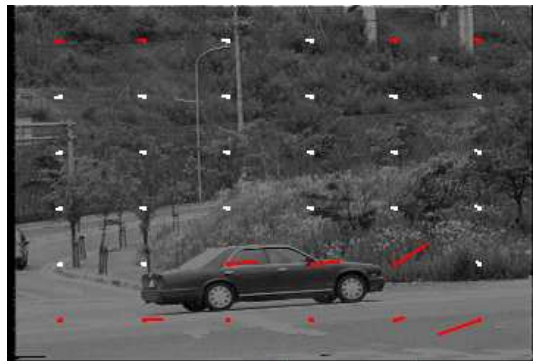
(白 : 選択注目点)

図 4.6 注目点の選択状況と速度ベクトル : 画像列 (a)





第1フレーム



第7フレーム



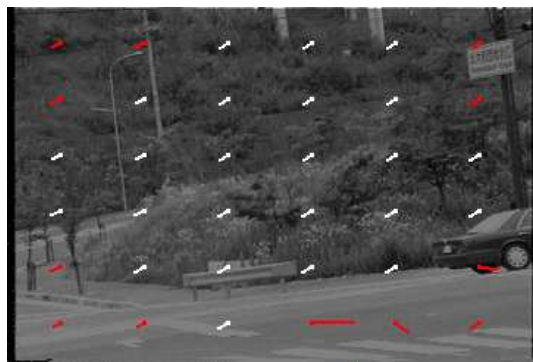
第13フレーム



第19フレーム



第25フレーム



第31フレーム

(白 : 選択注目点)

図 4.7 注目点の選択状況と速度ベクトル : 画像列 (b)



第1フレーム

第11フレーム



第21フレーム

第31フレーム



第41フレーム

第51フレーム

(白 : 選択注目点)

図 4.8 注目点の選択状況と速度ベクトル : 画像列(c)

選択された注目点数

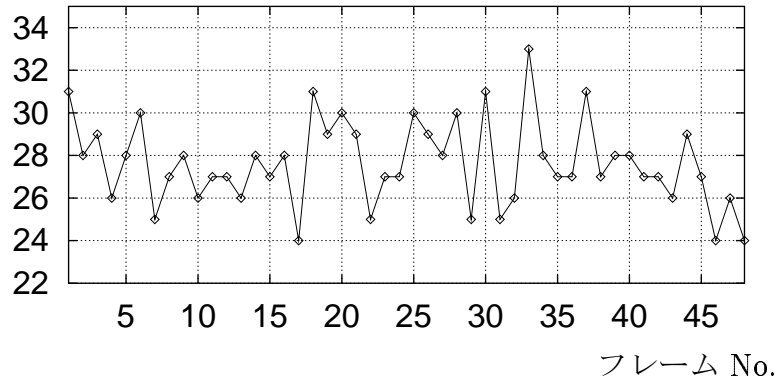


図 4.9 選択された注目点数の変化 :画像列 (a)

選択された注目点数

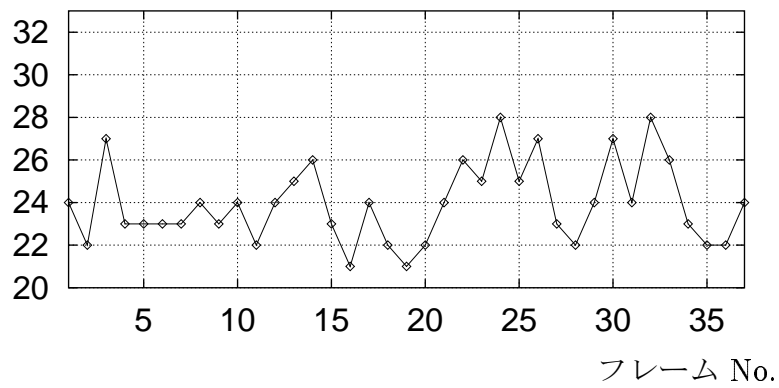


図 4.10 選択された注目点数の変化 :画像列 (b)

選択された注目点数

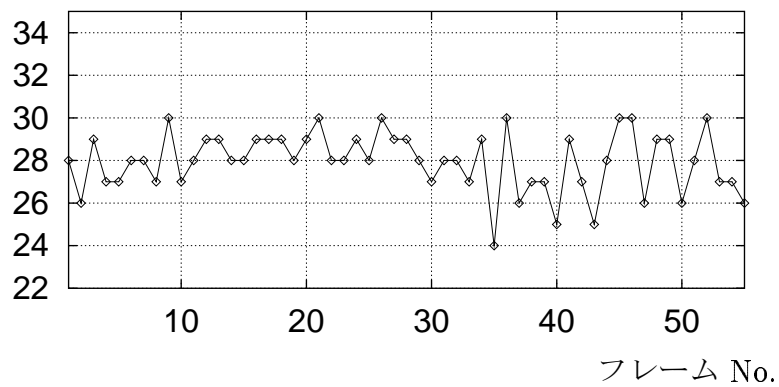


図 4.11 選択された注目点数の変化 :画像列 (c)

#### 4.4 動物体の切り出し結果

図??～?? に背景位置合わせ後の差分画像を示し、これらをもとに動物体を切り出した結果を図??～?? に示す。図??～?? では、画像中で背景部分と判定された画素をそのまま表示し、動物体として検出された画素は赤色で示している。また、注目点設定時にオフセット領域とした範囲の画素は、画像端の信頼性の無いデータによる影響を省くため、無条件で背景領域として扱っている。

図?? に示した結果から、レンズをズームインさせながら撮影しても、入力画像中で唯一の動物体である人間とリヤカーの領域がほぼ正しく抽出されていることが確認できる。

図?? に示した結果から、カメラを回転や首振り運動させながら撮影しても、入力画像中の唯一の動物体である自動車の領域が抽出されていることが分かる。

図?? に示した結果から、レンズをズームアウトさせ、カメラを首振りさせながら撮影しても、互いに交差する自動車が動物体領域としてほぼ正しく抽出されていることが分かる。

2段階アフィン変換による背景位置合わせの効果を確認するために、提案手法と、注目点の選択を行わずに1段階のアフィン変換で位置合わせを行なう方法について、差分画像と動物体切り出し結果を図?? にそれぞれ示す。図?? から、2段階アフィン変換（提案手法）の結果の方が、明らかに背景領域の位置合わせが良好であることがわかり、注目点の選択が有効に作用していることが確認できる。



第1フレーム - 第3フレーム



第8フレーム - 第10フレーム



第15フレーム - 第17フレーム



第22フレーム - 第24フレーム



第29フレーム - 第31フレーム

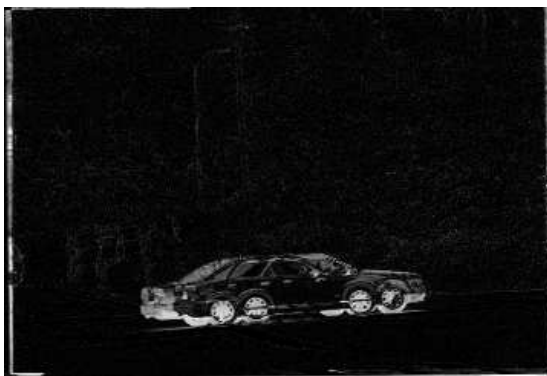


第36フレーム - 第38フレーム

図 4.12 差分画像 画像列(a) :ズームイン)



第1フレーム - 第3フレーム



第7フレーム - 第9フレーム



第13フレーム - 第15フレーム



第19フレーム - 第21フレーム

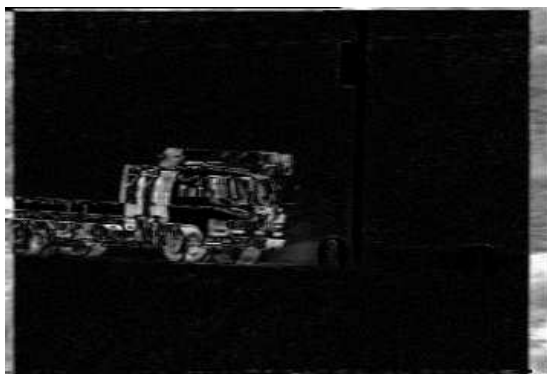


第25フレーム - 第27フレーム



第31フレーム - 第33フレーム

図 4.13 差分画像 画像列 (b) :回転・首振り)



第1フレーム - 第2フレーム



第11フレーム - 第12フレーム



第21フレーム - 第22フレーム



第31フレーム - 第32フレーム



第41フレーム - 第42フレーム



第51フレーム - 第52フレーム

図 4.14 差分画像 画像列(c) :ズームアウト・首振り、複数の動物体)





第1フレーム



第8フレーム



第15フレーム



第22フレーム



第29フレーム



第36フレーム

図 4.15 動物体抽出結果 画像列(a) :ズームイン





第1フレーム



第7フレーム



第13フレーム



第19フレーム



第25フレーム



第31フレーム

図 4.16 動物体抽出結果 画像列 (b) :回転・首振り)



第1フレーム



第11フレーム



第21フレーム



第31フレーム

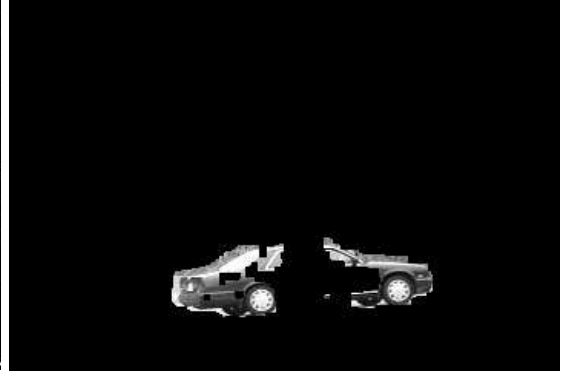


第41フレーム



第51フレーム

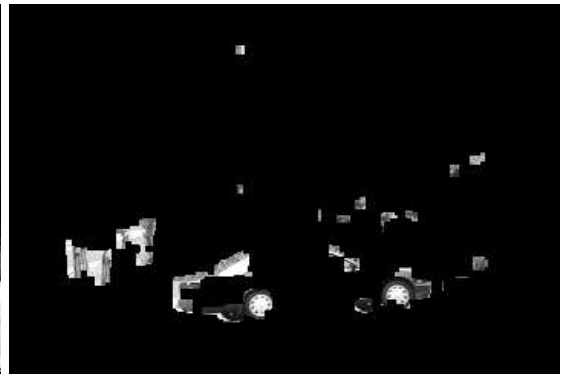
図 4.17 動物体抽出結果 画像列(c) :ズームアウト・首振り、複数の動物体)



アフィン変換後の差分画像

動物体切り出し結果

2段階アフィン変換 (提案手法)



アフィン変換後の差分画像

動物体切り出し結果

1段階アフィン変換

図 4.18 提案手法による注目点選択の効果

## 4.5 考察

図 4.6~?? に示した結果から、フレーム間での背景位置合わせに関して次のことがわかる。

- 動物体上に設定された注目点は全て、1段階目のアフィン変換で排除されている。
- 対応点探索時に明らかに誤対応が生じたと考えられる注目点は全て、1段階目のアフィン変換で排除されている。
- 背景位置合わせに選択される注目点数および位置は動的に変化している。
- 舗装された道路や空の領域など、目立ったテクスチャ特徴が存在しない領域に設定された注目点は誤対応が生じやすい。
- 背景領域のみの位置合わせを良好に実現している。

さらに、図??~?? に示した結果から、動物体抽出に関して次のことがわかる。

- カメラに与えた動きやズーム方向に依らずに、動物体領域を良好に抽出できる。
- 2段階アフィン変換による注目点の選択が効果的に作用している。
- 樹木のように背景から浮き出ている物体上に誤検出領域がわずかに存在する。
- テクスチャ特徴が存在しない動物体の内部領域が抜け落ちた様に検出される。

図?? の第 22 フレームや図?? の第 13 フレーム内の樹木上にわずかに存在する誤検出領域は、カメラの視点位置から見た距離が周辺に写っている背景領域までの距離と比べて大きく変化する部分である。このように被写体間の遠近差が大きい部分の検出エラーは原理的に完全には回避できないが、提案手法では対応点が求まった背景全域の平均的な位置合わせを行なうため、ある程度までは収縮・膨張処理で消去することができている。さらにこれらの誤検出領域は、

1. 連続した処理結果に継続的に発生していない。
2. 面積が小さい。

の2点が言えることから、抽出結果を時系列に追跡することやパラメータの調整を行なうことで、除去できる可能性があると考えられる。

動物体内部が抜け落ちてしまう現象は、該当領域が背景位置合わせ後の2枚の画像上で似た画素値を持つためである。例えば、図?? ではほぼ同一の位置にある若干大きさの異なるリヤカー同士の差分に、また、図??、?? では少し位置がずれたような自動車のボディー同士の差分になっていると考えられる。これらの抜け落ちエラーは提案手法が差分画像の作成に基づいているため、原理的に必ず生じる問題であると考えられる。

注目点の選択を行なう時に、式 5) で定義した  $\{E_i\}$  をそのメディアンで判定しているために、3.1節で述べた仮定 3 の条件を満たしていても動物体が画像面積の 50% に近い場合には、提案手法でも良好な結果が得られない可能性がある。

実行時間に関しては、現在のプログラムは汎用的利用を前提に設計しているので、対応点探索範囲や探索窓は大きめに取っており必要以上に実行時間を要している。また、さらに多くの時系列の処理結果を統合することなどで、動物体の切り出し精度を上げれば注目点の設定数を減らすことができると思われ、これらのパラメータの見直しを行なうことで、大幅な処理速度の改善が期待できると考えられる。

## 5. 結論

本論文では、2段階にアフィン変換を適用することで、移動カメラで撮影した動画像中の背景領域のみの位置合わせを実現し、動物体領域のみを抽出できるアルゴリズムを提案した。提案手法では、画像上に設定した注目点と次のフレームでのその対応点の関係から、局所的な対応関係をもとに画像全体にアフィン変換を複数回適用して、背景位置合わせに使用できる注目点が正しく選択できることを示した。また、カメラやレンズの動き情報が未知の3種類の動画像を用いた実験により、注目点の選択状況や背景位置合わせ後の差分画像、および動物体切り出し結果を示し、提案手法の実環境での有効性を確認した。これらの知見から、提案手法は例えば、自律移動ロボットなどに搭載して視覚センサーとして、自動車などの乗物に搭載して移動障害物検出装置として、さらには監視カメラに応用して侵入者の自動検出装置としてなど幅広く応用できる可能性があると考えられる。

抽出された動物体領域の内部が抜け落ちてしまう現象は、現状のアルゴリズムのパラメータ調整程度では完全に回避できない問題であると考えられる。さらに、抽出可能な動物体の画像中の大きさに関しては、画像上での最大サイズの評価、および注目点の選択時に設定している閾値との関係が明確にできていない。これらの問題点は、現在実験的に設定しているパラメータ類とともに今後さらに検討を要する課題である。

## 謝辞

本研究を進めるにあたり、終始暖かく御指導を頂きまし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。

本論文をまとめるにあたり、有益な御助言を頂きました像情報処理学講座 千原 國宏教授、並びにソフトウェア基礎講座 竹村 治雄助教授に深く感謝いたします。

本研究への助言や研究環境の維持整備に御尽力頂きましたソフトウェア基礎講座 岩佐 英彦助手、情報科学センター 片山 喜章助手に深く感謝いたします。

本学で研究する機会を与えて下さいました、住友電気工業株式会社 研究開発部門 辻 康次郎 技師長、電力事業部 古川 晃平技師長、電力システム技術研究所 福井 武所長、電力システム技術研究所 エネルギー情報システム研究部 仲野 恒雄部長に感謝いたします。

ソフトウェア基礎講座の諸氏には、本研究の多様な局面においておしめない御協力を頂きました。ここに感謝の意を表します。

最後に、本研究に没頭する筆者を陰ながら支えてくれた、妻 真由美に心から感謝致します。

## 参考文献

- [1] 武田 信之, 渡辺 睦, 小野口 一則: “消失点推定残差法を用いた移動観測系からの移動体検出”, 情報処理学会 第 51 回全国大会, 6S-5, 1995.
- [2] 羽下 哲司, 笹川 耕一, 黒田 伸一: “対象の局在性に着目した障害物認識手法”, 画像の認識・理解シンポジウム (MIRU'94), pp.II-97-104, 1994.
- [3] D. Nair and J.K. Aggarwal: “Detecting Unexpected Moving Obstacles That Appear in the Path of a Navigating Robot”, Proc. ICIP'94, Vol.2, pp.311-315, 1994.
- [4] 陳 超, 中嶋 正之, 松本 重貴: “Hough 変換を用いた海底ケーブルの自動認識”, 信学技報, PRU94-22, 1994.
- [5] 森村 淳, 魚森 謙也, 石井 浩史, 赤堀 裕志, 北村 好徳, 坂口 隆: “純電子式画像揺れ補正技術”, National Technical Report, Vol.37, No.3, pp.308-314, 1991.
- [6] D. Murray and A. Basu: “Motion Tracking with an Active Camera”, IEEE Trans. PAMI, Vol.16 No.5, pp.449-459, 1994.
- [7] J.M. Odobez and P. Bouthemy: “Detection of Multiple Moving Objects Using Multiscale MRF with Camera Motion Compensation”, Proc. ICIP'94, Vol.2, pp.257-261, 1994.
- [8] M. Ben-Ezra, S. Peleg and B. Rousso: “Motion Segmentation Using Convergence Properties”, Proc. ARPA Image Understanding Workshop, pp.1233-1235, 1994.
- [9] M. Irani, B. Rousso and S. Peleg: “Detecting and Tracking Multiple Moving Objects Using Temporal Integration”, Proc. ECCV'92, pp.282-287, 1992.
- [10] Q. Zheng and R. Chellappa: “A Computational Vision Approach to Image Registration”, Proc. 11th ICPR, pp.193-197, 1992.



- [11] 杉山 善明, 後藤 啓悟, 有木 康雄: “トポロジー変化に基づく動領域抽出”, 情報処理学会 第 49 回全国大会, 4F-8, 1994.
- [12] 寺久保 敏, 横矢 直和, 岩佐 英彦, 竹村 治雄: “移動カメラで撮影した動画像からの移動物体の抽出”, 情報処理学会 第 51 回全国大会, 6S-3, 1995.
- [13] 寺久保 敏, 横矢 直和, 岩佐 英彦, 竹村 治雄: “2 段階アフィン変換を利用した移動カメラ画像からの動物体の抽出”, 信学技報, PRU95-181, 1995.
- [14] S. Terakubo, N. Yokoya, H. Iwasa and H. Takemura: “Detecting Moving Objects from Image Sequences with a Moving Camera Using Two-step Affine Transformations”, Proc. 13th ICPR, 1996. (投稿中)
- [15] 谷内田 正彦 編: コンピュータビジョン, 丸善 (株), 1990.
- [16] 若原 徹: “局所的アフィン変換を用いた画像間のずれ検出”, 信学論 (D-II), Vol.J72-D-II, No.12, pp.2070-2079, 1989.
- [17] 田村 進一: FA のための画像処理技術, 工学研究社, 1988.
- [18] 画像処理サブルーチン・パッケージ SPIDER USER'S MANUAL, 電子技術総合研究所, 1980.
- [19] 高木 幹雄, 下田 陽久: 画像解析ハンドブック, 東京大学出版会, 1991.