

NAIST-IS-DD0361203

博士論文

画像上での見え方の変化に頑健な  
物体検出手法に関する一考察

川西 隆仁

2006年 9月 29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  
博士(工学) 授与の要件として提出した博士論文である。

川西 隆仁

審査委員：

横矢 直和 教授 (主指導教員)

千原 國宏 教授 (副指導教員)

木戸出 正継 教授 (委員)

# 画像上での見え方の変化に頑健な 物体検出手法に関する一考察\*

川西 隆仁

## 内容梗概

画像から物体を検出する技術は、画像検索、映像監視などへの応用が可能であり、より広範に適用できる手法の実現が望まれている。本研究では、照明・隠蔽などの環境の変化や物体の位置・姿勢の変化により画像上での物体の見え方が大きく変化する場合にも、効率よく物体を検出する手法の確立を目的とする。まず、1章では、物体検出法を、背景を既知として探索する背景モデルに基づく方法と物体の見え方を既知として探索する検出対象モデルに基づく方法とに分類して関連研究を概観し、本研究の目的と意義を明確にする。2章では照明条件の変化や隠蔽によって生じる物体の見え方の変化に対して頑健に検出する背景モデルに基づく物体検出手法を扱う。まず照明条件が変化し、かつ画像中の背景領域が少ない場合においても適用可能な背景差分法を提案する。次に画像情報と音情報を利用し、音情報・画像情報のどちらかが断絶した場合にも移動物体の追跡を継続する手法を提案する。3章では、物体の位置・姿勢の変化によって生じる物体の見え方の変化に対して、様々な見え方のモデルを用意する検出対象モデルに基づく物体検出手法を扱う。まず色ヒストグラム特徴を用いて高速に物体を検出する手法を提案する。次にこの手法を一般の距離特徴に拡張する。2章の手法により検出された移動物体領域に対してのみ3章の手法を適用することによって効率的に移動物体検出を行うことができる。最後に、4章で本研究を総括し今後の展望について述べる。

---

\* 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 博士論文, NAIST-IS-DD0361203, 2006年9月29日.

## キーワード

物体検出, 背景差分法, 逐次モンテカルロ法, アクティブ探索, テンプレートマッチング法, パン・チルト・ズームカメラ



# A Study on Robust Object Detection under Various Changes of Appearances in Images\*

Takahito Kawanishi

## Abstract

The techniques of finding an object from captured images are essential for many computer vision applications. However, conventional techniques cannot be applied for a wide range of complicated scenes because objects in those scenes have various appearances in captured images. This thesis proposes a novel approach to finding an object using various appearances. Chapter 1 gives a perspective of the studies in the areas of finding an object from the view points of background image subtraction methods and template matching methods. Chapter 2 describes two methods of finding moving objects in an unstable background scene captured by a fixed camera. First the proposed background image subtraction method works well even when illumination of the scene changes and an unknown object occupies most parts of the captured image. The second method tracks an object using audio and visual information. If an obstacle masks the object, this method continues to track the object successfully by using the sound of the object. Chapter 3 describes two fast template matching methods of finding a known object which has various appearances depending on the position and the orientation. First the proposed method uses color histogram feature and skips unnecessary matchings to reduce calculation costs. The second method is proposed to skip unnecessary matchings using general distance measure. Both methods in Chapter 3 can be

---

\* Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0361203, September 29, 2006.

applied to regions detected by the methods in Chapter 2. Finally, Chapter 4 summarizes the present work.

**Keywords:**

Object detection, background subtraction method, markov chain monte-carlo method, active search, template matching method, pan-tilt-zoom camera

# 目次

<b>1. 序論</b>	<b>1</b>
1.1 物体検出研究の概観	1
1.1.1 背景モデルに基づく移動物体領域検出	3
1.1.2 検出対象モデルに基づく物体検出	4
1.1.3 背景モデルと検出対象モデルの組み合わせ	8
1.1.4 画像情報と音情報の組み合わせ	8
1.2 本研究の位置づけと研究方針	9
1.3 本論文の構成	13
<b>2. 照明変化や隠蔽に頑健な移動物体領域検出</b>	<b>15</b>
2.1 序言	15
2.2 背景の大部分の隠蔽と照明変化に頑健な移動物体領域検出	16
2.2.1 背景の大部分の隠蔽に頑健な距離特徴と逐次モンテカルロ法を用いた背景差分法	17
2.2.2 実験	22
2.2.3 考察	34
2.3 画像情報と音情報の動的線形和による隠蔽に頑健な移動物体検出	37
2.3.1 画像による物体検出手法と音による物体検出手法	38
2.3.2 確率密度分布の重み付き線形和による物体追跡	40
2.3.3 実験	45
2.3.4 考察	51
2.4 結言	52
<b>3. 物体の位置・姿勢の変化に頑健な高速物体検出</b>	<b>54</b>
3.1 序言	54
3.2 色ヒストグラムを用いた物体検出の高速化	56
3.2.1 動的アクティブ探索法	58
3.2.2 複数のパン・チルト・ズームカメラによる物体の位置推定	73

3.2.3	実験	76
3.2.4	考察	87
3.3	一般距離特徴に拡張した物体検出の高速化	88
3.3.1	テンプレートマッチング	88
3.3.2	適応的ウィンドウスキップによる高速テンプレートマッチング法: Adaptive Window-Skipping Method (AWS)	90
3.3.3	複数テンプレートの高速探索のための拡張	96
3.3.4	実験	100
3.3.5	考察	115
3.4	結言	116
<b>4.</b>	<b>結論</b>	<b>118</b>
4.1	まとめと考察	118
4.2	展望と今後の課題	120
	<b>謝辞</b>	<b>123</b>
	<b>参考文献</b>	<b>124</b>

## 目 次

1	提案手法を組み合わせた場合の機能ブロック図 . . . . .	11
2	背景モデルに基づく移動物体領域検出における従来手法 [MWHT01] と提案手法の位置付け (照明の変化がある場合) . . . . .	12
3	検出対象モデルに基づく物体検出における従来手法 [MN94, MV98] と提案手法の位置付け (静止物体を検出対象とする場合) . . . . .	12
4	逐次モンテカルロ法における処理の流れ . . . . .	19
5	逐次モンテカルロ法を用いた背景画像の推定処理の流れ . . . . .	21
6	逐次モンテカルロ法を用いた背景画像推定法のフローチャート . . . . .	23
7	背景画像の生成に用いた 4 枚の蓄積背景画像: (a) 蛍光灯, (b) 白熱 灯 (部屋の右側), (c) 白熱灯 (部屋の左側), (d) 全ての照明を点灯	25
8	移動物体領域の大きな入力画像の例: 画像上で移動物体領域 (扇形 領域) が占める面積の割合 (左 20%, 中央 50%, 右 80%) . . . . .	26
9	提案手法を用いて生成した背景画像と背景差分による検出結果: (a) 推定された背景画像, (b) 検出結果 . . . . .	27
10	従来手法を用いて生成した背景画像と背景差分による検出結果: (a) 推定された背景画像, (b) 検出結果 . . . . .	28
11	提案手法と従来手法との精度比較: 入力画像と推定した背景画像間 の背景領域での二乗誤差 . . . . .	29
12	移動物体領域 (扇形領域) の面積比率を変えて提案手法を適用し た移動物体領域検出精度 . . . . .	30
13	移動物体領域 (扇形領域) の面積比率を変えて従来手法を適用し た移動物体領域検出精度 . . . . .	31
14	人物領域の検出結果: (a) 入力画像, (b) 提案手法, (c) 手作業によ る結果手法 . . . . .	32
15	白熱灯 (b) の明るさを緩やかに変化させて提案手法を適用した場合 の重み係数の追従性 . . . . .	33
16	白熱灯 (b) の明るさを急激に変化させて提案手法を適用した場合の 重み係数の追従性 (サンプル数: $N = 100$ ) . . . . .	35

17	白熱灯 (b) の明るさを急激に変化させて提案手法を適用した場合の 重み係数の追従性 (サンプル数: $N = 400$ ) . . . . .	36
18	背景差分画像から得られた確率密度分布 . . . . .	39
19	CSP 法による方向推定 . . . . .	40
20	円形マイクロホンアレイ . . . . .	41
21	音情報と画像情報を用いた物体追跡法のフローチャート . . . . .	44
22	音情報と画像情報の統合に関する実験で用いたシステムの構成 . . . . .	45
23	音情報と画像情報の統合に関する実験を行った室内環境 . . . . .	46
24	画像情報による物体方向の追跡結果: (a) 入力画像 (b) 背景差分 画像 (c) $\bar{P}_t^V(\phi)$ (d) $P_t^V(\phi)$ (提案手法) . . . . .	47
25	画像情報による物体方向の追跡の経過: (a) $\bar{P}_t^V(\phi)$ , (b) $P_t^V(\phi)$ (提案 手法) . . . . .	48
26	音源方向の追跡の経過 . . . . .	49
27	画像情報と音情報を統合した物体の追跡結果: (a) 推定した物体の 方向, (b) 尤度 $k_1$ の変化の様子 . . . . .	50
28	画像からの静止物体の検出における問題 . . . . .	54
29	パン・チルト・ズームカメラを使った探索範囲の拡大 . . . . .	56
30	予測制御によるカメラ制御回数の削減 . . . . .	58
31	動的アクティブ探索法の処理の概要 . . . . .	59
32	ヒストグラム区画の作成の例 . . . . .	61
33	ヒストグラム区画の作成のフローチャット . . . . .	62
34	重なりを利用した照合のスキップ . . . . .	63
35	複数テンプレートを用いた物体探索の処理の流れ . . . . .	65
36	並列探索の仕組み . . . . .	66
37	並列探索のフローチャート . . . . .	67
38	併合探索のフローチャート . . . . .	69
39	カメラ制御回数の削減の必要性 . . . . .	70
40	カメラ制御のフローチャート . . . . .	72
41	カメラ座標系と画像座標系 . . . . .	74

42	物体検出結果からの物体方向の導出 . . . . .	74
43	一台のカメラでの物体距離の導出 . . . . .	75
44	実験に使用したパン・チルト・ズームカメラ . . . . .	76
45	1台のパン・チルト・ズームカメラによる物体検出実験で用いたテンプレートの例 . . . . .	77
46	1台のパン・チルト・ズームカメラによる物体検出実験で用いた物体の学習位置と探索位置 . . . . .	78
47	複数台のパン・チルト・ズームカメラを使った実験環境 . . . . .	82
48	複数台のパン・チルト・ズームカメラを使った場合の物体位置に対する探索時間の比較 . . . . .	85
49	複数台のパン・チルト・ズームカメラを使った場合の物体位置に対する位置推定精度の比較 . . . . .	86
50	テンプレートマッチングによる物体検出の例 . . . . .	89
51	サブテンプレート . . . . .	91
52	テンプレート $T$ と ウィンドウ $W_A$ の距離を用いて, テンプレート $T$ と ウィンドウ $W_B$ の距離を推定 . . . . .	92
53	$T$ と $W_A$ の距離を用いて, $T$ と $W_B$ の距離を推定する代わりに, $T_{subA}$ と $W_{sub}$ の距離を用いて, $T_{subB}$ と $W_{sub}$ の距離を推定 . . . . .	93
54	三角不等式による距離下限値の導出 . . . . .	95
55	AWS のフローチャート . . . . .	97
56	異なるテンプレート間での距離下限値計算 . . . . .	98
57	複数テンプレートに対する AWS のフローチャート . . . . .	99
58	AWS の評価実験に用いた入力画像の例 . . . . .	101
59	AWS の評価実験に用いたテンプレートの例 . . . . .	101
60	マージンサイズを変えて AWS の探索時間を測定した結果 . . . . .	103
61	初期閾値を変えて AWS と他の手法の探索時間を比較した結果 . . . . .	104
62	テンプレートサイズを変えて AWS と他の手法の探索時間を比較した結果 . . . . .	105

63	拡張 AWS の実験のために物体の映る向きや大きさの異なるテンプレートを作成した物体 . . . . .	107
64	物体の向きが異なる多数のテンプレートを用いた拡張 AWS の探索時間 . . . . .	108
65	物体の向きが異なる多数のテンプレートを用いた拡張 AWS をマージン (m) を変えて実行した探索時間 . . . . .	109
66	物体の向きが異なる多数のテンプレートを用いた拡張 AWS をグループサイズ (N) を変えて実行した探索時間 . . . . .	110
67	物体の映る大きさが異なる多数のテンプレートを用いた拡張 AWS の探索時間 . . . . .	112
68	物体の映る大きさが異なる多数のテンプレートを用いた拡張 AWS をマージン (m) を変えて実行した探索時間 . . . . .	113
69	物体の映る大きさが異なる多数のテンプレートを用いた拡張 AWS をグループサイズ (N) を変えて実行した探索時間 . . . . .	114
70	擬人化エージェントと円筒ディスプレイ . . . . .	121

## 表 目 次

1	動的アクティブ探索の実験に用いた計算機およびカメラの仕様 . . . . .	76
2	ベクトル量子化を行った色ヒストグラムによる物体検出精度 . . . . .	79
3	アクティブ探索を拡張した高速探索手法の速度評価 . . . . .	80
4	1 台のパン・チルト・ズームカメラによる探索時間の比較 . . . . .	81
5	複数台のパン・チルト・ズームカメラを使った実験の計算機およびカメラの仕様 . . . . .	82
6	複数台のパン・チルト・ズームカメラを使った探索時間の比較 . . . . .	83
7	複数台のパン・チルト・ズームカメラを使った位置推定精度 . . . . .	83
8	AWS の実験に使用した計算機とソフトウェアの仕様 . . . . .	102
9	拡張 AWS の実験に使用した計算機とソフトウェアの仕様 . . . . .	106



10	物体の向きが異なる多数のテンプレートを用いた場合の拡張 AWS と他の手法の探索時間の比較 . . . . .	111
11	物体の向きが異なる多数のテンプレートを用いた場合の拡張 AWS と他の手法の探索時間の比較 . . . . .	115

# 1. 序論

## 1.1 物体検出研究の概観

どこに目的の物体があるかを求める物体検出の研究は、実世界の画像を入力とするコンピュータビジョン研究の初期から研究されてきた基本的かつ重要な問題である [SSH05]. 物体検出の問題は、検出対象の物体とそれ以外（背景）との2つの領域に分離する識別問題あるいは領域分割問題と定義できる. 物体を検出するアプローチとして、対象物体の見え方を基準に同じ見え方をする領域を対象物体とする方法と、背景の見え方を基準に背景でないものを対象物体とする方法がある. 前者を検出対象モデル、後者を背景モデルと呼ぶことにする. 入力に画像を用いることから、これらのモデルの作成も画像（見え方）に基づくことが多く、対象物体の見え方、あるいは背景の見え方の学習が行われる.

物体検出技術の歴史は古く、1970年代から、産業用ロボットビジョンの分野で利用されている [SI71, TI77, FDNI85, YIE74, KNMU73, HN83]. これらの研究は、工場のような安定した環境が作りやすい条件下で行われ、また、対象物体も形や模様が均質なものであった. 1980年代以降になると、対象物体や背景の条件が安定するオートメーションの研究から、様々な状況が想定される、移動ロボットナビゲーション・自動走行・ヒューマンインターフェース・セキュリティなどの研究へと広い範囲に展開してきている [THKS88, KCL+98]. 最近では、ITS への期待や、セキュリティへの関心の高まりなどで、オフィスや家庭などの室内環境や屋外環境などのより一般的な環境での物体検出の研究が盛んに行われるようになってきた.

このような物体検出技術の適用先の拡大に伴って、検出対象の見え方、背景の見え方の複雑性が増し、物体検出に関する問題を難しくする要因が広がっている. 問題を難しくする要因を、検出対象に関する要因、環境の変化に関する要因の2つに分けて考察する. まず検出対象に関する要因を整理すると、次のようになる.

1. 物体の位置の違いに伴う画像上の位置、スケールの変化
2. 物体の姿勢の違いに伴う画像上の見え方の変化

### 3. 人間の顔や動きなど, 検出対象自身の形状の変化

一方, 環境の変化に関する課題を整理すると, 次のようになる.

1. 照明と検出対象との位置関係によって生じる色の変化
2. 日照の変化など緩やかな照明の変化
3. 雲による日照の遮蔽や, 照明の点灯・消灯に伴う大きな照明変化
4. 極度のノイズ (豪雨などの天候の変化や遮蔽物, カメラの一時的故障など) による視界不良

上記の要因に加えて, 高視野高解像度のカメラを用いたり, 複数のカメラを用いることによる探索する領域が拡大し, 計算量が増加する難しさもある.

上記物体検出に関する問題の難しさに対して, 背景モデルに基づく方法, 検出対象モデルに基づく方法のそれぞれで研究が進められている. 背景モデルに基づく方法には, 物体の見え方の変化を考慮する必要がないため, 計算量が少ないという長所があるが, 移動物体が存在する領域全てが検出されてしまうため, 物体が何であるかの識別や, 物体の位置・姿勢といった物体の状態を推定することができないという短所がある. 一方で検出対象モデルに基づく方法には, 物体の識別が可能であり, 位置・姿勢といった物体の状態も知ることができる長所がある反面, 物体の多様な見え方と入力画像の部分領域との膨大な回数の照合を行う必要があるため, 計算量が大きいという短所がある. このため, 移動物体に対して物体の識別が必要な場合, 背景モデルに基づく方法を用いて移動物体が存在する領域をあらかじめ絞り込み, その領域に対して検出対象モデルに基づく方法を実行する方法がよく用いられている. また, 画像情報だけでは十分な情報が得られない場合には, 音情報を利用する手法の研究が行われている.

以下では, 物体検出に関する従来研究を, 検出対象モデル, 背景モデル, 検出対象モデルと背景モデルの組み合わせ, 画像情報と音情報の組み合わせの4つのタイプに分けて概観し, 次に従来手法の問題点について述べ, 本研究の位置づけと研究方針を明確にする.

### 1.1.1 背景モデルに基づく移動物体領域検出

画像内で観測される背景は、見かけ上様々な変化が生じる。それらの変化への対応方法として、下記の 3 つの方式が一般的である。

1. 照明変化に不変な特徴を利用
2. 背景画像を更新
3. リアルタイムに背景画像を生成

本節では、以上の分類に従って背景モデルに基づく物体検出技術について述べる。

**照明変化に不変な特徴を利用** 緩急の照明変化に不変な (あるいは頑健な) 画像特徴として正規化距離が利用されることが多い。最も単純には、画像をベクトル表記し、ベクトルの大きさを正規化する。部分画像ごとの正規化ベクトルで背景をモデル化し、入力画像と背景画像における正規化ベクトル間の距離を測ることで背景差分を実現している [NMF+96]。この距離は正規化距離 (Normalized Vector Distance) と呼ばれる。さらに、画像を小領域に分割し、それぞれの領域で照明光強度を用いた正規化距離を求め、背景と物体との判別を行った事例がある [HWM99]。これらの手法は、照明の種類が 1 つのみの場合には有効であるが、白熱灯と蛍光灯など光源色が異なる照明が混在する場合で、特に局所的な照明変化に対しては誤検出率が高くなることがある。

**背景画像を更新** 背景差分の結果を基に、入力画像を用いた背景画像の更新を行う手法が数多く提案されている [HKIK85, TKK97, AF98]。しかし、得られた背景画像が前景となる移動物体領域の影響を受ける、処理に複数枚の入力画像を要するために時間的な遅延が生じるなどの問題が指摘されており、急激な照明変化が起きた直後や断続的な照明変動の最中においては、対象物体の検出精度が低下する。

**リアルタイムに背景画像を生成** 一般的に、照明変化などをともなった画像系列は、画像空間において、低次元の部分空間中に分布する [Hal94, BJ01]。例えば、

- 物体表面を完全拡散とみなせる
- 光源が無限遠とみなせる
- 影が無視できる

の以上3つの性質が成り立つとき、任意照明下でとりうる画像はたかだか3次元の部分空間に分布することが知られている [Sha92, Sha97].

前記の性質を利用して各時刻での入力画像から背景画像を推定し、その結果を用いて照明が変化する状況下でも背景差分を行うことができる [ORP00, MWHT01]. 背景画像は、様々な照明条件下であらかじめ取得・蓄積された背景画像（以下、蓄積背景画像と呼ぶ）から求めた固有画像の重み付き線形和として生成される。各固有画像の重み係数は、入力画像を固有画像に線形投影することで求められる。この手法では、画像中に複数の人物が写っている場合など前景となる移動物体領域が画像中で占める面積の割合が大きいと、その影響により正しい背景画像を求めることができない。この問題を解決するために、先の述べた正規化距離を使った手法 [HWM99] と組み合わせた手法も提案されている [MWHT01].

以上で述べてきた従来技術は、照明光の明るさの変化や照明方向の変化に対してはある程度の有効性が示されているが、前景となる移動物体領域が画像上の大部分を占め、かつ照明光の色が変化する状況への対応は困難である。

### 1.1.2 検出対象モデルに基づく物体検出

対象物体の見え方が既知で、その知識を元に検出処理を行う場合、適用先が一般化するほど、対象物体の見え方の多様性が増加し、網羅することが難しくなる。このため、与えられた教師パターン（検出対象物体の実例）を多数用意することが必要になり、与えられた教師パターンに含まれないものでも検出可能な汎化能力の高い手法が必要になる。

1.1 節に述べたように、検出対象モデルを使用した場合には、検出対象・環境の課題双方の影響を受ける。このため下記の問題を難しくする要因を解決する必要がある。

1. 画像上の検出対象の位置が未知
2. 物体の位置・姿勢の変化による画像上の物体の見え方の変化
3. 視野の拡大・解像度の向上

本節では以上の分類に従って従来研究を概説する。

**画像上の検出対象の位置を探索** 検出対象の画像を既知として、その画像が入力画像上のどこにあるかを調べる技術をテンプレートマッチング法 [Ono87] という。テンプレートマッチング法は、既知の画像パターン（テンプレート）と入力画像上から切り出した多数の部分領域（ウィンドウ）とを照合して物体の場所を求める手法である。要素技術として人物や物体などのコンテンツの検索、動画像からの移動物体の追跡、ステレオ測量における対応点探索、MPEG 圧縮における動き補償符号化などに広く用いられている。

テンプレートマッチング法の問題点の1つは、テンプレートとウィンドウとの距離計算にかかる計算量が大きいほか、入力画像上でウィンドウを走査して照合を行うため、距離計算回数が多くなり、計算時間が膨大になることである。このため、これまで多くのテンプレートマッチングの高速化手法が提案されてきた。

テンプレートマッチングの計算量を削減する手法には、大きく分けて、テンプレートとウィンドウの距離計算の計算量を削減する手法と、距離計算を行うウィンドウの数を削減する手法がある。前者の手法のうち、代表的な手法には、残差逐次検定法 (Sequential Similarity Detection Algorithm : SSDA) [BS72] がある。SSDA では、距離計算の途中で閾値と随時比較し、閾値を超えたとき、その処理を終了する。SSDA をさらに高速に行うために、CPU の特性を元に比較を行うタイミングを変える手法 [HMK03] などが提案されている。

照合計算を行うウィンドウの数を削減する手法には、2段階テンプレートマッチング法が古くから用いられている [VR77]。2段階テンプレートマッチング法では、テンプレートの一部であるサブテンプレートを用いて、各ウィンドウでの距離下限値を計算し、距離下限値が閾値以下のウィンドウのみテンプレート全体での照合を行なうことによって、計算量を削減している。また、Li らは、テンプレート

$\mathbf{T}$  のベクトルの大きさ  $|\mathbf{T}|$  と各ウィンドウのベクトルの大きさ  $|\mathbf{W}|$  をあらかじめ計算し,  $|\mathbf{W}|$  が式 (1) を満たさないとき照合を省略することによって計算量を削減するアルゴリズム (Successive Elimination Algorithm : SEA) [LS95] を提案した.

$$|\mathbf{T}| - SAD_{\min} < |\mathbf{W}| < |\mathbf{T}| + SAD_{\min} \quad (1)$$

ここで,  $SAD_{\min}$  は, これまで照合したウィンドウとの距離計算により求めた絶対距離の中で最も小さい絶対値距離である. さらに Lee らは, 多重解像度の画像を効率よく生成し, 粗密探索のアプローチで精度を保障したまま, 探索の高速化を行うアルゴリズム (Block Sum Pyramid Algorithm : BSPA) [LC97] を提案した. BSPA は  $2^n \times 2^n$  画素のブロックの合計値を,  $n=0$  から順にテンプレートサイズまで段階的に計算することによって平均画像を効率よく生成する. 平均画像間の絶対値距離は, 必ず元の画像の絶対値距離よりも小さいため, 大きなブロックの平均画像間の絶対値距離が閾値を上回った場合には, 細かいブロックの平均画像間の絶対値距離を計算する必要はない. この性質を利用して, 粗い平均画像の段階で不要な照合を省略し, 計算量を削減している. さらに絶対値距離以外の一般的な距離, 任意の形状のブロックを用いた手法を Alkhansari が提案している [GA01]. Alkhansari の手法は絶対値距離, 大きさが  $2^n \times 2^n$  画素の多段階ブロックを使う場合, BSPA と全く同じアルゴリズムとして動作する. これまで述べた上記の手法は, 入力画像上の全てのウィンドウへの前処理を必要とし, また, MPEG 圧縮を想定している場合では, 閾値が十分小さくないと働かないなどの問題がある. このため, 多くの入力画像から数枚のテンプレートを探すような, 物体探索や, 画像検索のアプリケーションには適用できない場合がある.

一方で, 極めて高速な物体探索手法としてアクティブ探索法 [MV98, KSM99, KKM00a, KKM00b] が提案されている. アクティブ探索法は特徴照合の際にヒストグラムを用い, ヒストグラムの性質を用いて精度を落とさないことを保証しながら探索に不要な照合を大幅に削減する手法である. 更に, ヒストグラム照合を利用するため多少の形状変動に対しても頑健である. しかしながら, 照明条件や物体の向きの違いにより特徴が大きく変化する場合には, 単一のテンプレートのヒストグラムでは正しく探索できない. この変化を吸収するには, さまざまな条

件で物体を撮影した複数のテンプレートのヒストグラムを用意し探索する必要があるが十分な検討は行われていなかった。

**物体の位置・姿勢の変化による画像上での物体の見え方の変化への対応** 物体の位置や姿勢の変化を考慮にいたった場合には画像上での物体の見え方は大きさと姿勢が変化する。物体の位置によって照明条件が異なる場合も多い。この変化に合わせて多くのテンプレートを用意して照合することによってこれらの問題を解決することができる [MN94, TSO00]。しかしながらテンプレートを多数用意する必要があるうえ、その全てのテンプレートと全てのウィンドウとを照合するため、次元圧縮などで高速化を図ったとしても、探索時間はいまだ大きい。

**視野の拡大・分解能の向上への対応** 広域監視, 遠隔監視, 自律移動ロボット, テレプレゼンス, 拡張現実感などのシステムのためには, 広い視野を獲得できるカメラを用いて広範囲にわたって画像を取得する必要がある。このためには, 専用の光学系を用いる全方位センサ [YYY95] や, 多方向に向いた複数台のカメラからなる画像入力系 [Ish97] や周囲を旋回撮影するパン・チルト・ズームカメラ [WM96] が用いられている。それぞれに得失があり, 例えば, 全方位カメラは, 一度の撮影でシームレスな全方位パノラマ画像が得られるという特長を持っており, その点では物体検出に有利であるが, 物体に対する解像度が十分に確保できないという欠点がある。これに対し, 複数台のカメラの利用は解像度を十分に確保できるが, 装置構成が大掛かりになるうえ, カメラ間で入力情報をやりとりするためにカメラ間の性能誤差や視点の違いに対する処理を付加する必要がある。一方, パン・チルト・ズームカメラは, 全方位を同時には撮影できないものの, 旋回中心と視点を一致させておけば, 比較的容易にパノラマ画像を生成できるという特長をもっている。

これらのカメラを用いた物体検出のうち, 検出対象モデルを使う手法には, ロボットの誘導のために室内環境中のランドマークを検出する方法 [ASF+98, JC98] が提案されている。これらの方法は特定の人工的なマーカーにしか適用できず, 一般の物体を物体固有の特徴に基づいて検出できない。



### 1.1.3 背景モデルと検出対象モデルの組み合わせ

移動物体が何であるかを求めたい場合、背景差分法で移動物体領域をあらかじめ抽出し、その範囲内のみを対象として検出対象モデルに基づく手法により検出を行う手法が一般的であり、この組み合わせを利用した研究が数多く提案されている。[WADP97, HK05, HK03, HSY04b, HSY04a]. とくに、人間の顔や動きなど、検出対象自身が、形状の変化を伴う場合には、動きを伴うため、背景モデルと検出対象モデルを組み合わせたアプローチが用いられていることが多い[Aka97, ILYH05, KFH05].

### 1.1.4 画像情報と音情報の組み合わせ

雑音や照明等の周辺環境が様々に変化する状況においてもある程度の精度を維持した状態で、物体の位置を自動的かつ連続的に把握し続けるためにこれまでに画像情報と音情報を用いた様々な追跡手法が提案されている。

画像情報を用いた技術では、カルマンフィルタ [ZF01] や逐次モンテカルロ法 [IB98a, IB98b] を用いた物体追跡手法や、フレーム間差分法、背景差分法に基づいた物体検出手法が提案されている。しかし画像情報のみを用いた手法の多くは、隠蔽が生じた場合には継続的な追跡が困難である、また照明が暗い場合には検出精度が低下するといった問題がある。

音情報を用いた処理としては、CSP法 (Cross-power Spectrum Phase analysis: 白色化相互相関法 [NYNS00]) に基づいた音源位置推定手法が用いられることが多い。この手法は連続音に対しては比較的安定した性能を発揮する。しかし人物を対象とした場合では通常、語句間や会話間で音の途切れ (無声区間) が生じる。周辺雑音が存在する環境においては、無声区間で雑音源を物体と誤認識してしまう場合がある。

上記の課題を克服するため、画像情報と音情報を統合した物体の追跡手法の開発が進められている。その中の一つのアプローチとして、追跡対象の位置を確率密度分布で表現し、各情報の観測結果から求めた確率密度分布を積算により統合し、位置情報を得る手法が提案されている [MKN97, ZDD02, BAJ02, SRS01]. し

かし確率密度分布の積算を行うため、一方の情報の信頼性が著しく低い、もしくは観測されなかった場合には、状態が良い一方の情報のみを用いた時よりも精度が低下してしまう。例えば、背景差分法を画像情報に適用した場合、隠蔽発生時には画像上で移動物体が観測されないために確率密度分布はゼロとなり、積算後の確率密度分布もゼロとなってしまう、追跡を継続できない。

## 1.2 本研究の位置づけと研究方針

本研究では、画像照合を使った物体検出技術において問題となる、画像上での見え方の変化に伴う問題を扱う。背景モデルに基づく方法、検出対象モデルに基づく方法に関してそれぞれ下記の課題がある。

**照明の変化・隠蔽** 照明の点灯・消灯により、背景の見え方が大きく変化すると背景モデルでの移動物体検出が困難になる。とくに、照明が大きく変化し画像上で移動物体領域が背景領域よりも多くの面積を占める状況で正しく動作する背景差分法はなかった。より劣悪な環境として、隠蔽などで移動物体が画像上に捉えられない状況もある。このような状況下では音情報を組み合わせることが考えられるが、画像情報が完全に断絶することがあるような劣悪な環境下で正しく動作するものはなかった。

**物体の位置・姿勢の変化による画像中の物体の見え方の変化** 物体の位置や姿勢の変化を考慮にいれた場合には画像上での物体の見え方は大きさと姿勢が変化する。物体の位置によって照明条件が異なる場合も多い。広い環境から静止物体を検出する場合には、物体の位置・姿勢の変化に応じた多数のテンプレートを用いて探索する必要がある。しかしながら、その膨大な計算量のため、室内環境中の位置・姿勢が未知の静止物体を対象にした物体検出法はこれまで提案されていない。

本論文では上記の問題を解決し、照明の変化や隠蔽のある状況下でも背景モデルに基づく方法で物体を検出できること、広い環境からも位置・姿勢が未知の静止物体を検出できることを課題とする。

上記 2 つの問題に対して、それぞれ 2 つのステップで解決する。

## a) 照明変化・隠蔽に頑健な移動物体領域の検出

**ステップ 1. (2.2 節)** 照明の点灯・消灯・調光などの照明条件が変化する室内環境下でも適用可能な背景モデルに基づく移動物体領域検出手法を提案する。本手法は、一定数の背景候補画像を生成し、それらから選び出した最も適切な背景画像を用いて移動物体領域検出を行う。提案手法では入力画像上で前景領域となる移動物体領域が背景領域よりも広い面積を占める状況下で照明の点灯・消灯などが発生する場合においても継続的に一定数の背景候補画像を生成し続け、移動物体領域の検出を継続する機能を有する。

**ステップ 2. (2.3 節)** 隠蔽などの影響により画像では物体が捉えられない状況下においても画像情報と音情報に基づき、移動物体を追跡し続けることができる手法を提案する。画像情報・音情報から得られる移動物体の候補に対して確率密度分布を求め、それらを動的に組み合わせることによって、画像情報、音情報のいずれかの断絶がある場合にも、正しい候補を選択する。また断絶した情報の復旧後にも両方の情報を使った移動物体の検出を継続することを可能になる。

## b) 物体の位置・姿勢の変化により画像上での見え方が変化する物体の検出

**ステップ 1. (3.2 節)** 物体の位置・姿勢の変化によって多様な見え方をする物体を室内環境から探索可能な手法を提案する。提案手法では、位置・姿勢の変化に応じて生成した多数のテンプレートに対してアクティブ探索法による高速探索を行うとともに、色ヒストグラム特徴の類似値の上限値の概念をテンプレート間にも拡張している。この結果、照合すべきウィンドウ候補の数を大幅に削減し、計算時間を大幅に削減している。

**ステップ 2. (3.3 節)** 色ヒストグラムに限定しない一般の距離特徴において照合すべきウィンドウ候補の数を削減する手法を提案し、距離特徴で識別できる物体を高速に検出することを可能にする。

**提案手法の関係** 各提案手法の関係を表す機能ブロック図を図 1 に示す。

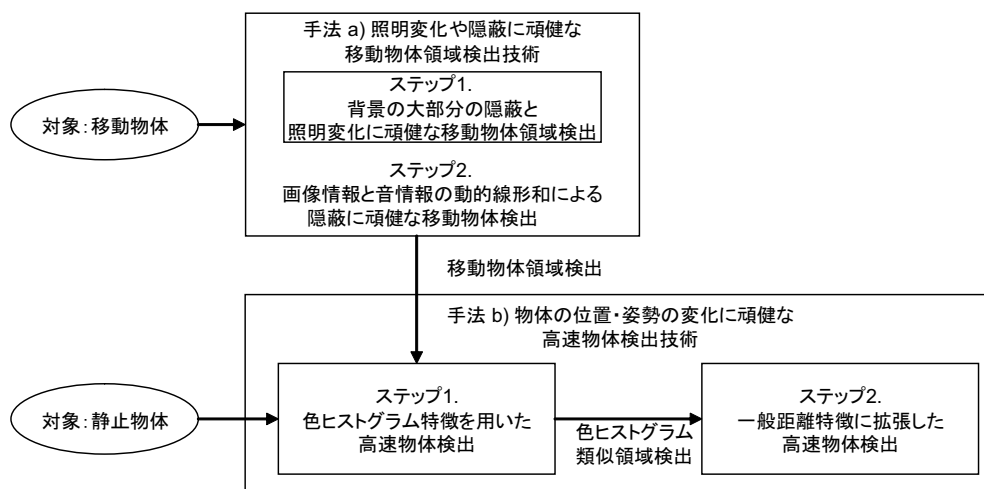


図 1 提案手法を組み合わせた場合の機能ブロック図

まず手法 a) と手法 b) の関係について述べる。手法 a) は背景モデルに基づく方法であるため、移動物体領域の検出にしか用いることができない。このため移動物体に対して物体が何であるかの識別を行う場合には手法 a) と手法 b) を組み合わせることが効率的である。手法 a) を前処理に用いて移動物体領域を絞り込み、その移動物体領域に対してのみ、物体の識別を行う手法 b) を行うことで高速に物体を検出することが可能になる。一方、静止物体の検出には手法 a) を用いることができず、手法 b) だけを用いて画像中の物体のある可能性のあるすべての部分領域に対して網羅的に照合を行う必要がある。このため、高速な物体検出手法が要求される。

手法 a) と手法 b) 内のステップ 1. とステップ 2. との関係について述べる。手法 a) におけるステップ 1. の手法は、ステップ 2. の手法の画像情報として用いることができる。この結果、照明変化に強く、隠蔽にも強い手法とすることができる。次に、手法 b) におけるステップ 1. の手法とステップ 2. の手法の関係を述べる。ステップ 1. の手法が速度が速い反面、ヒストグラム特徴しか利用できないため、精度が十分ではない場合がある。このため、ステップ 1. の手法を前処理に利用し検出した色ヒストグラムが類似した領域に対してステップ 2. の手法を用いるこ

	従来手法	提案手法 (手法 a))
画像上の移動物体領域がわからない	×	○
画像上の移動物体領域が大	×	○
画像上の移動物体領域が小	○	○

○：移動物体領域が検出できる．×：移動物体領域が検出できない．

図 2 背景モデルに基づく移動物体領域検出における従来手法 [MWHT01] と提案手法の位置付け (照明の変化がある場合)

	従来手法	提案手法 (手法 b))
画像上の物体領域が大	○	○
画像上の物体領域が小	△	○

○：実時間で検出できる．△：現実的な時間では検出できない．

図 3 検出対象モデルに基づく物体検出における従来手法 [MN94, MV98] と提案手法の位置付け (静止物体を検出対象とする場合)

とで検出精度を改善できる．

**提案手法の位置付け** 手法 a), 手法 b) のそれぞれについて, 従来手法との位置付けを図 2, 図 3 に示す. 手法 a) のステップ 1. の手法により従来解決していなかった照明が変化する状況下において, 背景領域が小さい場合にも正確な背景推定を実現する. さらにステップ 2. の手法により, 画像情報から物体が見えない場合においても音情報により物体を検出することができる.

手法 b) のステップ 1. の手法により, 従来手法では現実的な時間で検出できなかった, 物体が遠くにありズームしなければ検出できないような状況下でも実時間で検出することができる. b) のステップ 2. の手法を用いれば b) のステップ 1. の手法で検出した領域に対してさらに詳細な特徴で照合することも可能になる.

図 2, 図 3 に示すように, 本研究では物体検出技術に関わる照明の変化と隠蔽の

課題と物体の位置・姿勢の変化の課題をそれぞれ背景モデルに基づく手法, 検出対象モデルに基づく手法において解決し, 物体検出技術の適用範囲を広げている.

### 1.3 本論文の構成

本論文の構成は以下のとおりである.

2章では照明の変化や隠蔽に対しても頑健に, 背景モデルの方法を用いて移動物体領域を検出することを目的とする. 次に3章では, 物体の位置・姿勢の変化に伴い物体の見え方が多様になった場合についてそれら全ての見え方を用意して検出対象モデルに基づく手法により探索し, 静止物体を含め目的の物体を識別して高速に検出することを目的とする.

2章では, まず室内環境での照明の点灯・消灯のような照明が変化する環境下でも移動物体領域を検出する背景モデルに基づく移動物体領域検出手法について述べる. 全方位画像から生成した人工画像で評価実験を行い, 入力画像上で前景領域が背景領域よりも広い面積を占める場合, 照明の点灯・消灯などが発生する場合について頑健に背景モデルを生成し, 移動物体領域を検出できることを示す. 本章で示した手法は様々な背景画像を仮定して, その中から現状に適合した尤度の高い背景画像を生成して移動物体領域の検出を行う. しかしながら画像情報のみから追跡を行うために画像情報が断絶した場合, 移動物体検出を継続できない.

次に, 画像情報が断絶した場合にも音情報を用いて移動物体検出を継続する手法について述べる. 音情報と画像情報を組み合わせ, 隠蔽などの影響により画像では移動物体が捉えられない状況下においても物体を追跡し続ける. 実験により, 画像情報, 音情報の完全な断絶がある場合にも, それらの信頼性が高い情報を利用することによって移動物体を検出し続けることを示す.

3章では, まず, 室内環境において, 近くから, 遠くまで, 様々な位置にある物体を実時間で探索可能な手法について述べる. 本手法はヒストグラムの代数的性質を利用して未照合のテンプレートとウィンドウに対して類似度の上限値を使って照合の省略を行い大幅な探索の高速化を実現している. 実際に室内環境中にカメラを4台設置し, いくつかの物体を探索する定量的な評価実験を行い, 本手法の特徴と有効性について考察する. 本手法では利用できる特徴が色ヒストグラム特徴

に限定されているため、背景色が物体色に類似する場合には物体を見分けることができず、適用範囲が色が鮮やかな物体に限定されていた。

次に上記問題を解決するため色ヒストグラム特徴に限定しない一般の距離特徴を用いた物体探索の計算時間を大幅に削減する手法について述べる。この手法は、一般の距離を用いているため、距離が定義された特徴で識別できる任意のアルゴリズムを利用でき、適用範囲が広い。テンプレートが1枚の場合、複数の場合それぞれについて定量的な評価実験を行い、距離特徴で識別できる物体を高速に検出できることを示す。

4章で、本論文のまとめと、展望と今後の課題について述べる。

## 2. 照明変化や隠蔽に頑健な移動物体領域検出

### 2.1 序言

本章では照明の変化や隠蔽に対しても頑健に、背景モデルに基づく方法を用いて移動物体領域を検出することを目的とする。

まず室内環境での照明の点灯・消灯のような照明が変化する環境下でも移動物体領域を検出する背景モデルに基づく移動物体領域検出手法について述べる。提案手法は背景画像を仮定して、その中から現状に適合した尤度の高い背景モデルに基づいて移動物体領域の検出を行う。移動物体領域が画像中の大部分を占める場合においても移動物体領域と背景を分離できる頑健な距離尺度を導入する。さらに逐次モンテカルロ法を用いて一定の計算時間内に必要な数の背景候補画像を生成する。照明の点灯・消灯・調光が行われる室内において、実験を行い、移動物体領域が50%以上を占める場合にも正しく背景を推定し、移動物体領域を検出できることを示す。しかしながら画像情報のみから検出を行うために画像情報が断絶した場合、検出を継続できない問題がある。

検出対象が、人や車など音を発する物体の場合、その発する音の情報も利用できれば、追跡を継続できる状況も多く存在する。そこで、次に、画像情報が断絶した場合にも音の情報をを用いて継続する手法について述べる。提案手法では、移動物体の存在に関する尤度分布を、画像情報および音情報のそれぞれから求めた移動物体の存在に関する確率密度分布の尤度による重み付き線形和とする。その尤度は周辺環境の変動に応じて変化させる。実験により、画像情報、音情報いずれかの断絶がある場合にも、尤度の高い情報を利用することによって移動物体を追跡し続けることを示す。

以下、2.2節において、照明が変化する環境下でも移動物体領域を検出する背景モデルに基づく移動物体領域検出手法を述べ、2.3節において、隠蔽や音の途切れなどの影響により情報の断絶が生じるような状況下においても、画像情報と音情報を合わせて用いることによって頑健に移動物体を追跡し続ける手法を提案する。最後に2.4節で本章をまとめる。



## 2.2 背景の大部分の隠蔽と照明変化に頑健な移動物体領域検出

本節では背景差分法による高精度な移動物体領域検出を目標に、照明光の明るさおよび色が変化し、かつ、検出対象の移動物体領域が画像上の大部分を占める場合にも対応が可能な背景画像の推定手法を提案する [TKMT03a, TKMT03b, TKMT04b]。各照明条件下における背景画像は、あらかじめ用意した様々な照明条件下の背景画像の重み付き線形和として生成する。同様のアプローチに基づく固有空間への線形投影による従来手法 [ORP00, MWHT01] では、入力画像と背景画像の間での平均二乗誤差を評価値として用いるため、移動物体領域が入力画像上で占める面積の割合が大きい場合には、その影響により背景画像の推定精度が低下するという問題があった。それに対し提案手法では、入力画像と背景画像の同じ座標の画素に注目し、その値が一致した画素の数を評価値として導入し、前述の問題を解決する。しかし、この評価値を用いる場合には、各蓄積画像の重み係数の推定に最小二乗法のような解析的な手法が使えないため、全ての重み係数の組み合わせについて評価値を計算する必要があるため、計算コストが膨大となる問題が生じる。そこで、精度を維持しつつ計算量の削減を実現するため、確率モデルに基づくサンプリングおよびその更新を行う逐次モンテカルロ法を導入し、各蓄積画像の重み係数の追跡問題として扱う。

提案手法は 1.1.1 節で述べたリアルタイムに背景画像を生成する従来手法 [MWHT01] と次の三点において大きく異なる。一点目は、蓄積背景画像の重み付き線形和として各時刻の背景画像を生成することであり、二点目は、各時刻において複数の背景画像を候補画像として生成し、その中から最も確からしい画像を一枚選択することである。三点目は、蓄積背景画像の重み係数の求め方である。本手法では、入力画像と背景候補画像とを比較してその差分値が閾値以下である画素の数を求め、その値を用いて背景候補画像の確からしさを評価する。しかし、この評価法を用いる場合には最小二乗法のような解析的な解法が使えず、また重み係数の全ての組み合わせを調べると計算量が膨大となり現実的ではない。膨大な重み係数の組み合わせの中から最適解を高速に見つけるためには、統計情報などを用いて解の探索範囲を絞り込む必要がある。組み合わせ問題における最適解の探索技術の一つに、逐次モンテカルロ法 (Condensation / Markov Chain Monte Carlo:MCMC)

がある。逐次モンテカルロ法では確率モデルに基づいたサンプリングにより、対象の状態推定を、高速かつ高精度に行うことが可能であることが知られており、移動物体の追跡にしばしば用いられてきた [IB98a, IB98b]。本手法ではこの技術を、照明環境の状態予測に用いる。

本節は次の順に議論を進める。まず、手法の詳細を 2.2.1 節に述べる。2.2.2 節では、白熱灯および蛍光灯照明下において移動物体領域の検出を行う実験を実施し、入力画像の 50%以上が移動物体領域により占められている場合でも、照明変化に対し頑健な移動物体領域検出が可能なことを示す。2.2.3 節で本手法について考察する。

### 2.2.1 背景の大部分の隠蔽に頑健な距離特徴と逐次モンテカルロ法を用いた背景差分法

複数照明下で観察される物体の明るさは、拡散反射以外が無視できる場合、各光源下での物体の明るさの線形和で表現できることが知られている [Hal94, BJ01]。そこで本手法では各時刻における背景画像を、あらかじめ用意された様々な照明条件下での背景画像（以下、蓄積背景画像と呼ぶ）の重み付き線形和として生成する。蓄積背景画像の重み付き線形和による背景画像の生成を可能とするためには、(1) 蓄積背景画像と各時刻における入力画像の撮影領域が一致すること、(2) 画像情報が欠損していないこと、(3) カメラへの入力と出力の関係が既知もしくは線形であること、が必要である。そこで本手法では、以下に示す 3 条件を仮定する。

- カメラ位置と視野が固定されている
- センサの飽和を避けるよう、露光条件が適切に調節されている
- カメラの自動補正機能（自動ゲイン制御やホワイトバランス）は使用しない

各蓄積背景画像の重み係数は照明変化に合わせて、時間とともに変化する。この照明変化を予測することができれば、少ない計算コストで各蓄積背景画像の重み係数を求めることが可能である。そこで提案手法では、統計的な枠組みに基づいた逐次モンテカルロ法を、各蓄積背景画像の重み係数の変化予測に適用する。逐

次モンテカルロ法については 2.2.1.1 節で説明する。また各蓄積背景画像の重み係数の決定方法については 2.2.1.2 節で述べる。以下の説明や実験では全方位画像を用いているが、提案手法は他の種類の画像にも適用可能である。

### 2.2.1.1 逐次モンテカルロ法

物体の状態に関する追跡問題は確率的な枠組みにおいて、状態確率密度の伝播過程の追跡として定式化することができる [IB98a]。時刻  $t$  における物体の状態を  $\mathbf{X}_t$ 、時刻  $t$  におけるセンサによる観測結果を  $\mathbf{C}_t$ 、時刻  $\mathbf{C}_{t-1}$  から  $\mathbf{C}_t$  にかけての一連の観測結果を  $C_t = \{\mathbf{C}_1, \dots, \mathbf{C}_t\}$  とする。また、 $C_{t-1}$  が得られた後の時刻  $t$  における事前確率を  $p(\mathbf{X}_t|C_{t-1})$ 、物体の状態が  $\mathbf{X}_t$  であるときに観察された  $\mathbf{C}_t$  の尤度  $\pi_t$  を  $p(\mathbf{C}_t|\mathbf{X}_t)$  とすると、時刻  $t$  において  $C_t$  が得られた後の事後確率はベイズ則に基づき以下のように計算される。

$$p(\mathbf{X}_t|C_t) = k_t p(\mathbf{C}_t|\mathbf{X}_t) p(\mathbf{X}_t|C_{t-1}) \quad (2)$$

ここで  $k_t$  は正規化定数である。推移確率  $p(\mathbf{X}_t|\mathbf{X}_{t-1})$  が与えられており、またマルコフ性が仮定できる場合には、事前確率は (3) 式に示すように計算される。

$$p(\mathbf{X}_t|C_{t-1}) = \int p(\mathbf{X}_t|\mathbf{X}_{t-1}) p(\mathbf{X}_{t-1}|C_{t-1}) d\mathbf{X}_{t-1} \quad (3)$$

(2) 式, (3) 式を用いて物体の状態変化を追跡するためには、あらゆる  $\mathbf{X}_t$  に関して尤度を全て計算する必要があるが、これは現実的ではない。そこで逐次モンテカルロ法では、確率モデルに基づき状態空間上での離散的なサンプリングを行い、その結果に基づきベイズ推定を近似的に行う。

時刻  $t$  において状態空間上で  $N$  点のサンプリングを行い、 $\mathbf{S}_t = \{\mathbf{s}_{t,1}, \dots, \mathbf{s}_{t,N}\}$  を得る。そして (2) 式, (3) 式の関係に基づき確率密度  $p(\mathbf{X}_t|\mathbf{S}_t)$  を近似的に計算する。さらに、 $\mathbf{S}_t$  と推移確率を用いて、時刻  $t+1$  における  $N$  点のサンプル  $\mathbf{S}_{t+1} = \{\mathbf{s}_{t+1,1}, \dots, \mathbf{s}_{t+1,N}\}$  を新たに生成する。以下にその手順を示す (図 4)。

**手順 1** 各サンプル  $\mathbf{S}_t = \mathbf{s}_{t,1}, \dots, \mathbf{s}_{t,N}$  をそれぞれの尤度  $\pi_{t,1}, \dots, \pi_{t,N}$  で重み付けする。 $n$  番目の尤度  $\pi_{t,n}$  は、 $n$  番目の観測結果  $\mathbf{s}_{t,n}$  と物体の観測モデルの比較から計算される。この重み付けされた各サンプルを用いて、時刻  $t$  における事前確率を近似的に求めることができる。

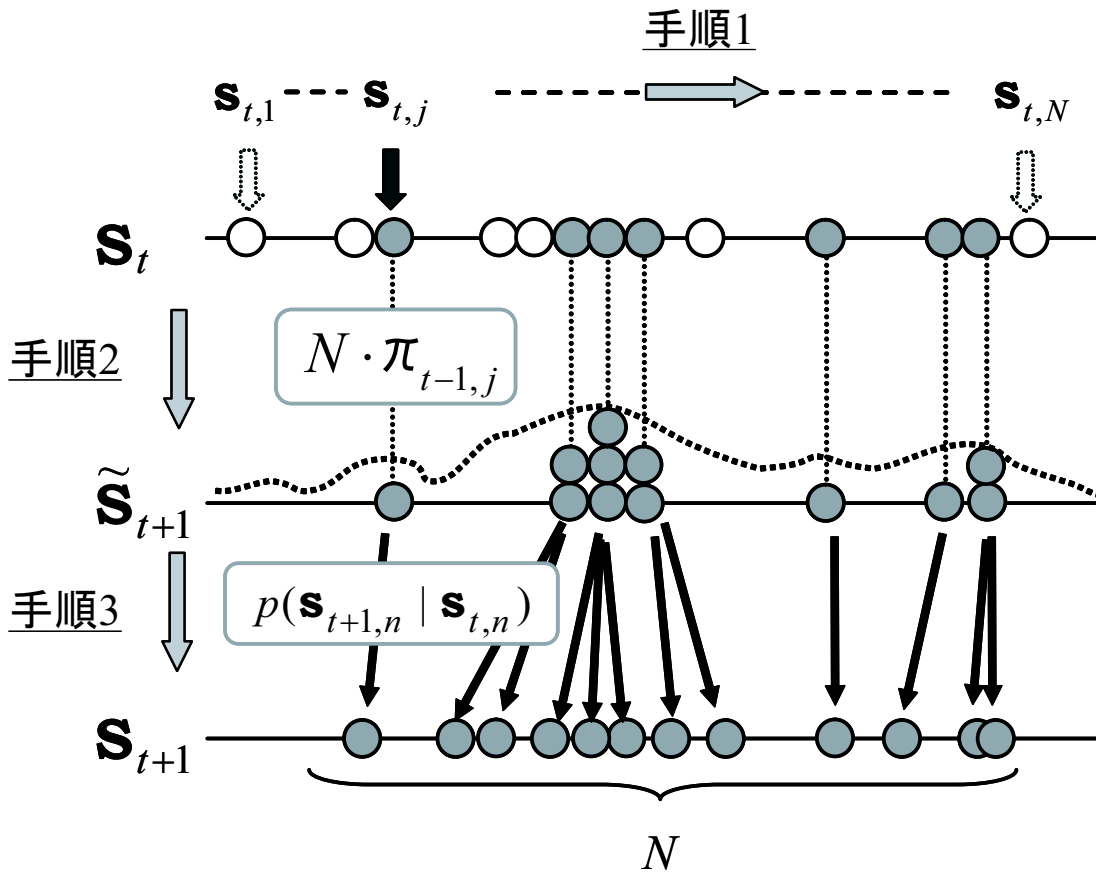


図 4 逐次モンテカルロ法における処理の流れ

手順 2  $\mathbf{s}_t$  の各サンプル  $\mathbf{s}_{t,n}$  をそれぞれ  $N \cdot \pi_{t,n}$  回選択し, 新しいサンプル  $\tilde{\mathbf{s}}_{t+1} = \{\tilde{\mathbf{s}}_{t+1,1}, \dots, \tilde{\mathbf{s}}_{t+1,N}\}$  を得る. 尤度が相対的に小さなサンプルは, この時点で追跡が打ち切られる.

手順 3 各  $\{\tilde{\mathbf{s}}_{t+1,n}\}$  をそれぞれ推移確率  $p(\mathbf{s}_{t+1,n} | \mathbf{s}_{t,n}) = p(\mathbf{X}_t | \mathbf{X}_{t-1})$  に従って変化させ, 時刻  $t+1$  におけるサンプル  $\mathbf{s}_{t+1} = \{\mathbf{s}_{t+1,1}, \dots, \mathbf{s}_{t+1,N}\}$  を得る.

### 2.2.1.2 逐次モンテカルロ法を用いた背景画像推定

図5に提案手法における背景画像の推定処理の流れを示す.

初めに, さまざまな照明条件下における背景画像 (蓄積背景画像) を撮影し, 蓄積する. ここで, 蓄積背景画像の集合を  $\mathbf{H} = \{H^1, \dots, H^L\}$  と表わす. また, 時刻  $t$  における重みベクトルの集合を,  $\mathbf{S}_t = \{\mathbf{s}_{t,1}, \dots, \mathbf{s}_{t,N}\}$  各重みベクトルの要素を  $\mathbf{s}_{t,n} = (s_{t,1}^1, \dots, s_{t,n}^L)$  とする. 各  $\mathbf{s}_{t,n}$  は考えられる全ての組み合わせパターンの中から, 尤もらしい  $N$  通りの組み合わせが逐次モンテカルロ法によりサンプリングされ, それぞれ更新される. そして得られた  $N$  通りの  $\mathbf{S}_t$  と  $\mathbf{H}$  を用いて,  $N$  枚の背景画像の候補 (背景候補画像)  $\tilde{H}_{t,N} (n = 1, \dots, N)$  が生成される. このとき, 時刻  $t$  において  $n$  番目に生成される背景候補画像は次式のように定義される.

$$\tilde{\mathbf{H}}_{t,n} = \sum_{k=1}^L s_{t,n}^k \mathbf{H}^k \quad (4)$$

正しい背景画像が得られたとき, 背景領域に関して, 入力画像と背景候補画像の同じ座標の画素値が一致するはずである. そこで, 入力画像と背景候補画像の差分を計算し, その値が閾値以下になる領域の面積 (画素数) が最大となる背景候補画像は正しい背景画像にほぼ等しいと考え, 背景候補画像  $\tilde{\mathbf{H}}_{t,n}$  の尤度  $\pi_{t,n}$  を以下のように定義する.

$$\pi_{t,n} = k_n \sum_y \sum_x f(i_{x,y} - \tilde{h}_{x,y}) \quad (5)$$

$k_n$  は正規化定数であり, また  $i_{x,y}$  と  $\tilde{h}_{x,y}$  はそれぞれ, 座標  $(x, y)$  における入力画像と背景候補画像  $\tilde{\mathbf{H}}_{t,n}$  の画素値を示す. ここで閾値を  $\rho$  とし, (5) 式における関数  $f(i_{x,y} - \tilde{h}_{x,y})$  を次式のように定義する.

$$f(i_{x,y} - \tilde{h}_{x,y}) = \left( \rho - |i_{x,y} - \tilde{h}_{x,y}| \right)^2 \quad (6a)$$

when  $|i_{x,y} - \tilde{h}_{x,y}| \leq \rho$

$$f(i_{x,y} - \tilde{h}_{x,y}) = 0 \quad (6b)$$

when  $|i_{x,y} - \tilde{h}_{x,y}| > \rho$

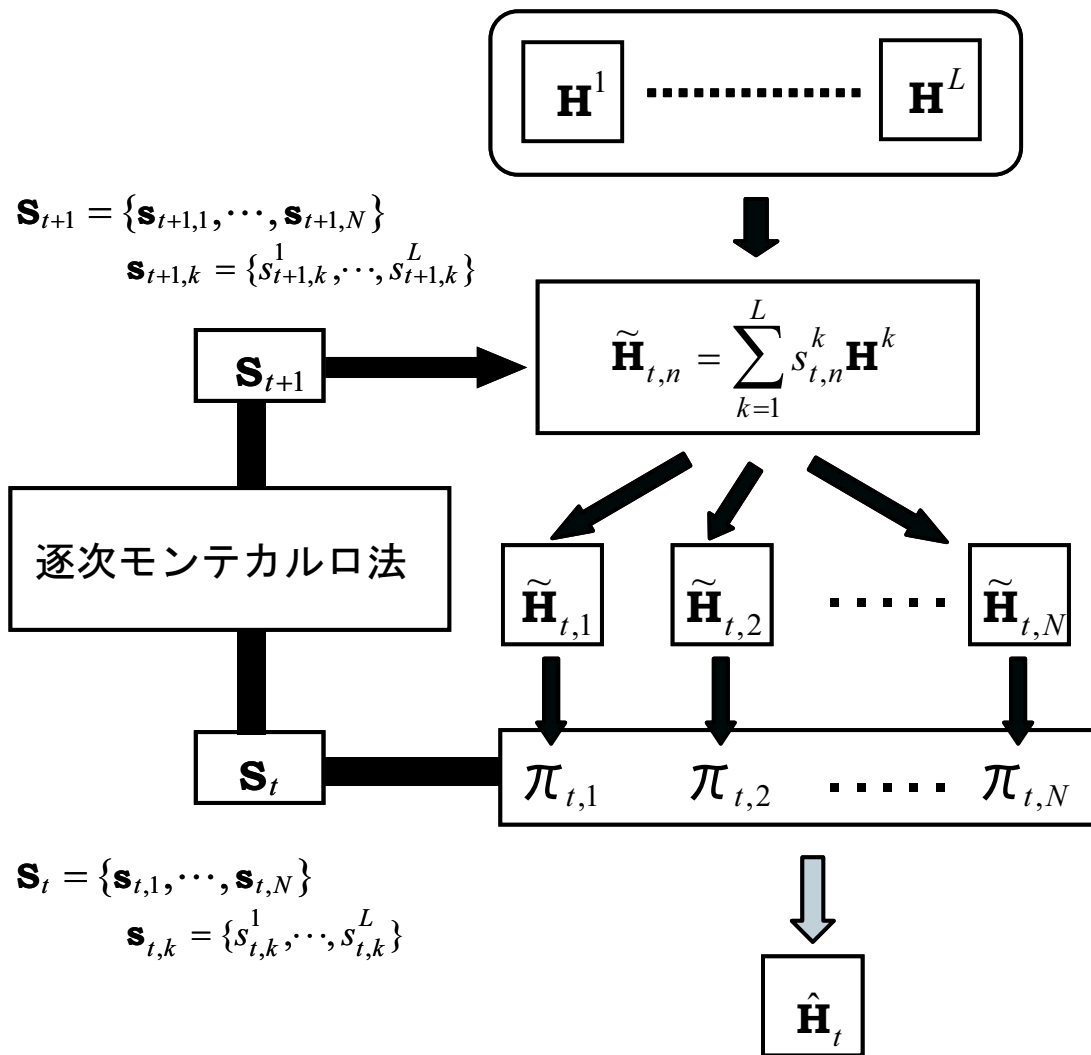


図5 逐次モンテカルロ法を用いた背景画像の推定処理の流れ

そして、生成された背景候補画像の中から尤度が最大となる画像を一枚選択し、それを時刻  $t$  における背景画像の推定結果とする。時刻  $t+1$  における重みベクトル  $\mathbf{S}_{t+1}$  は、時刻  $t$  における重みベクトル から以下の手順により生成される。

1.  $\mathbf{s}_{t,n}$  をその尤度  $\pi_{t,n}$  に比例して選択する。
2. 推移モデルに従い  $\mathbf{s}_{t,n}$  を変化させ、その結果を時刻  $t+1$  における重みベクトルとする。
3. 1) および 2) の処理を  $N$  回、繰り返す。

1) では尤度が大きい  $\mathbf{s}_{t,n}$  ほど重複して選択される回数が多く、逆に尤度が小さい  $\mathbf{s}_{t,n}$  は消失する。その結果、サンプル数は常に  $N$  個に保たれる。また 2) における推移モデルとして今回はランダムな変化を仮定し、その変化量  $r$  は正規分布である推移確率  $p(\mathbf{s}_{t+1,n}|\mathbf{s}_{t,n})$  に基づき決定した。

$$s_{t,n}^i = s_{t-1,n}^i + r \quad (7)$$

なお、時刻  $t=0$  における重みベクトルはランダムに生成するが、数回のサンプリングと更新を経て、自動的に尤もらしい値へ収束する。

逐次モンテカルロ法を用いた背景画像推定法のフローチャートを図6に示す。

### 2.2.2 実験

照明の変化範囲の予測がある程度限定でき、また想定外の外乱がほとんどない状態で撮影された画像を蓄積背景画像とし、室内で撮影した全方位画像を入力画像とした場合に対し提案手法を適用した。ここでは、異なる光源色をもつ複数の照明光源の明るさが変化する環境において、入力画像上の多くの領域が移動物体領域により占められている場合でも、提案手法により安定した移動物体領域検出が実現可能であることを確認する。

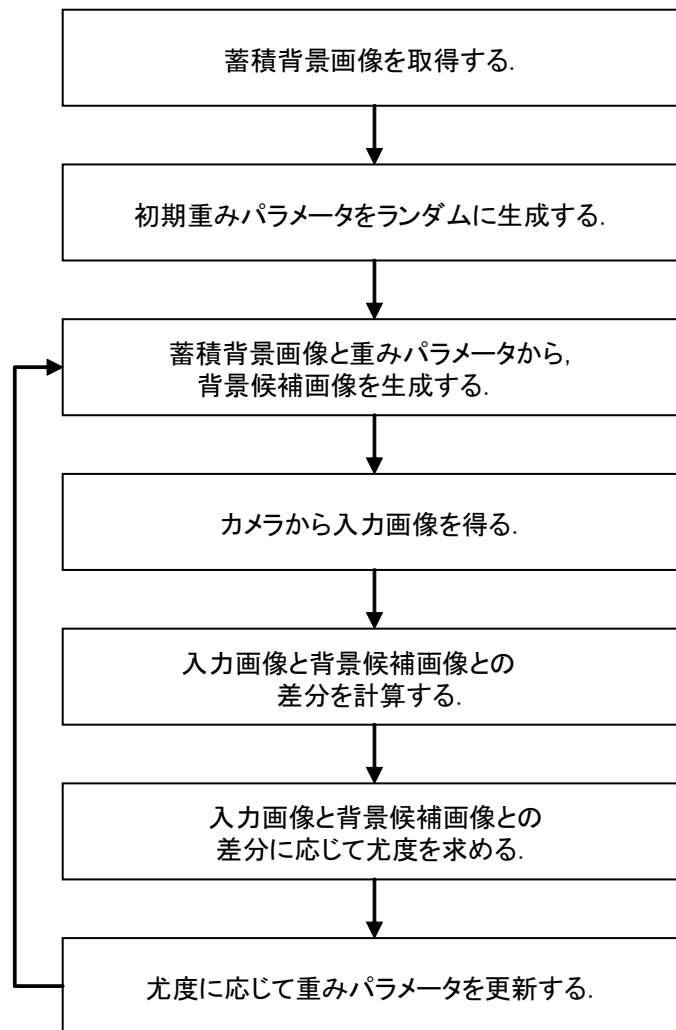


図 6 逐次モンテカルロ法を用いた背景画像推定法のフローチャート



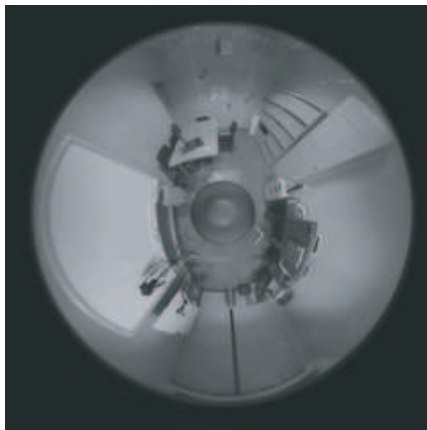
### 2.2.2.1 実験条件

蓄積背景画像および入力画像の撮影では、3つの光源 a, b, c (a: 蛍光灯, b, c: 白熱灯) を照明に使用した。白熱灯 b, c はそれぞれ独立に明るさを連続的に変化させることができる。今回の実験で使用した全方位カメラはデジタルカラー CCD カメラ (SONY 製 DFW-SX900) と凸面の双曲面ミラー (アコウル製) で構成され、RGB 各 8bit の階調を持つ画像が出力される。また、RGB のそれぞれについて入力・出力特性 ( $\gamma$  曲線) をあらかじめ計測しておき、その結果を基に取得画像の階調補正を行った。なお、撮影中はカメラの絞り、シャッター速度は変化させず、一定の露光条件下での画像を取得した。取得した画像はフレームレート 3.5 フレーム/秒 の動画画像である。取得した画像を記憶装置に一度蓄積し、その画像に対し各処理を行った。蓄積背景画像として、3つの光源のうち1つずつ点灯した状態で1枚ずつ、加えて全てを最大の明るさで点灯させた状態で1枚で、計4枚の画像を収集した (図7)。3つの光源の全点灯の状態は推定の安定性を増すために用いている。

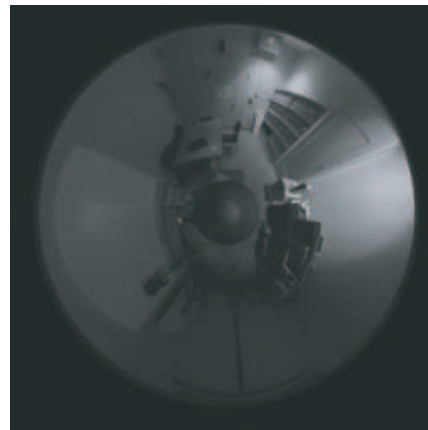
### 2.2.2.2 画像に占める移動物体領域の割合に対する検出精度

蛍光灯のみを点灯させた状態から白熱灯の明るさを徐々に変化させ、150枚の画像を撮影した。室内には1名の人物がおり、入力画像は (1) 人物が移動、(2) 人物が静止し照明変化が開始、(3) 照明変化が終了し人物が移動再開、といった3つの状況から構成される。本実験では移動物体領域が画像上で占める面積の割合を制御するために、画像の一部を白く塗りつぶした。入力画像の例として図8に3枚の画像を示す。移動物体領域が占める面積の割合は、それぞれ約20%、50%、80%である。

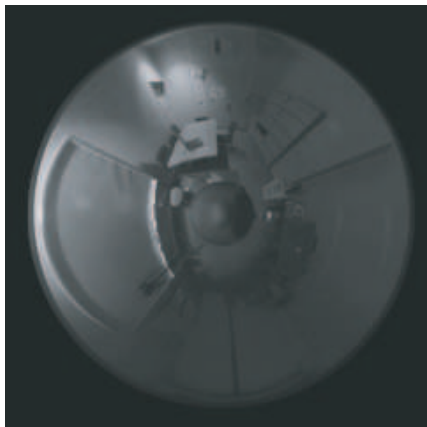
本実験では、照明の変化パターンは事前に分からないものとする。そこで、推移確率を正規分布と仮定し  $p(\mathbf{s}_{t+1,n}|\mathbf{s}_{t,n}) = N(\mathbf{s}_{t,n}, \sigma)$ 、またサンプル数  $N = 100$  とした。重み係数  $\mathbf{s}_{t,n} = (s_{t,n}^1, \dots, s_{t,n}^L)$  の範囲を  $0 \sim 2.0$  とし、その範囲で尤度  $\pi_{t,n}$  が最大となる重み係数の探索を行い、背景候補画像の生成を行った。その際の画像サイズは  $320 \times 240$ [pixel] とし、背景差分処理を含めた処理時間は Pentium



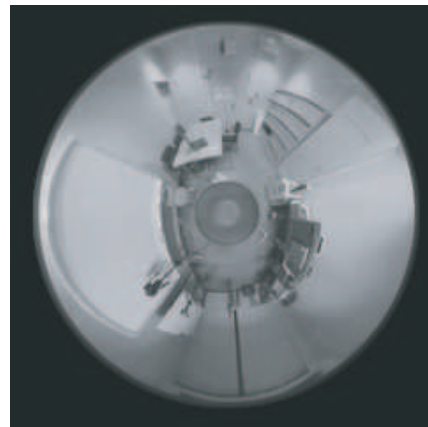
(a)



(b)



(c)



(d)

図 7 背景画像の生成に用いた 4 枚の蓄積背景画像:(a) 蛍光灯, (b) 白熱灯 (部屋の右側), (c) 白熱灯 (部屋の左側), (d) 全ての照明を点灯

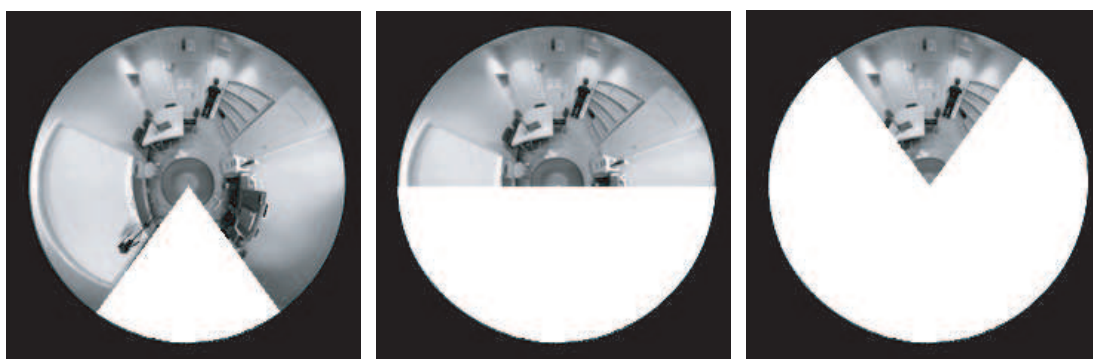


図 8 移動物体領域の大きな入力画像の例: 画像上で移動物体領域 (扇形領域) が占める面積の割合 (左 20%, 中央 50%, 右 80%)

4 1.8GHz(RAM:1GB) の PC を用いて約 330 ミリ秒/フレームであった。

図 8 の入力画像に対し, 提案手法により推定された背景画像を図 9(a) に, そして背景差分画像を図 9(b) に示す. 図 9(b) において, 扇形の白色領域は人為的に塗りつぶした部分を, その上部に見られる小さな白色領域は室内の人物を示している. 背景差分画像から, 図 8 の全ての場合について, 背景の推定がほぼ正しく行われたと思われる.

同じ入力画像に対し, 部分空間への線形投影 [MWHT01] (以下, 従来手法とする) により推定した背景画像を図 10(a) に, そして背景差分画像を図 10(b) に示す. 従来手法では, 移動物体領域部分の面積の割合が大きい場合 (50%, 80%) には, 扇形領域の影響により, 背景画像が全体的に明るすぎる画像となってしまう, また背景差分画像からも背景の推定に失敗したことが確認できる.

図 9 と図 10 の比較から, 従来手法では推定できない, 背景領域が 50%未満の場合でも正しく背景が推定できることがわかる.

図 11 に, 入力画像と推定された背景画像の間での, 背景領域に関する平均二乗誤差 (RMSE) を求めた結果を示す. 提案手法による結果の RMSE は全体的に小さく, その値はほぼ一定である. それに対し従来手法による結果の RMSE では, 移動物体領域の割合が 40%を超えたあたりからその値が大きくなっている.

これらの結果から, 入力画像の 50%以上が移動物体領域に占められている場合

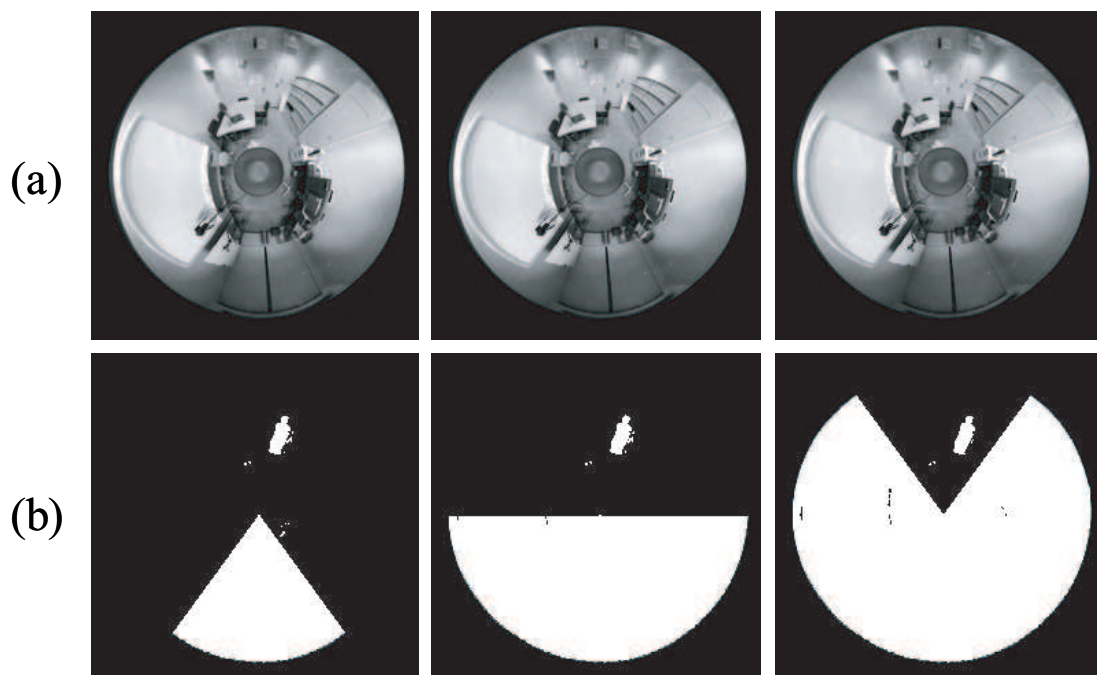


図9 提案手法を用いて生成した背景画像と背景差分による検出結果: (a) 推定された背景画像, (b) 検出結果

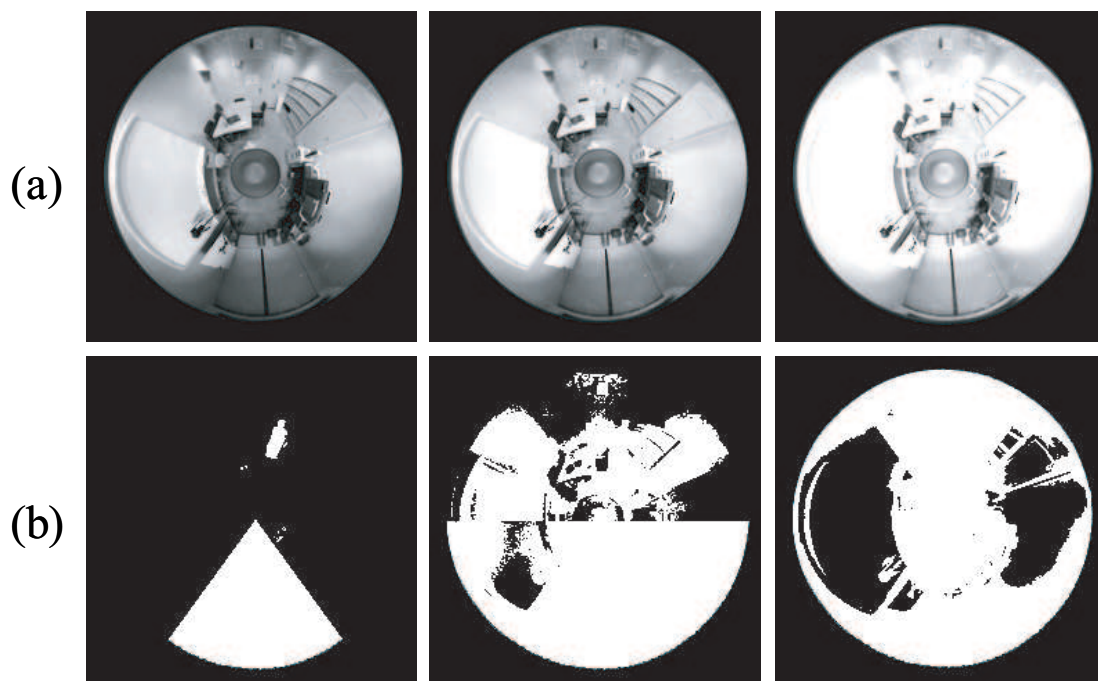


図 10 従来手法を用いて生成した背景画像と背景差分による検出結果: (a) 推定された背景画像, (b) 検出結果

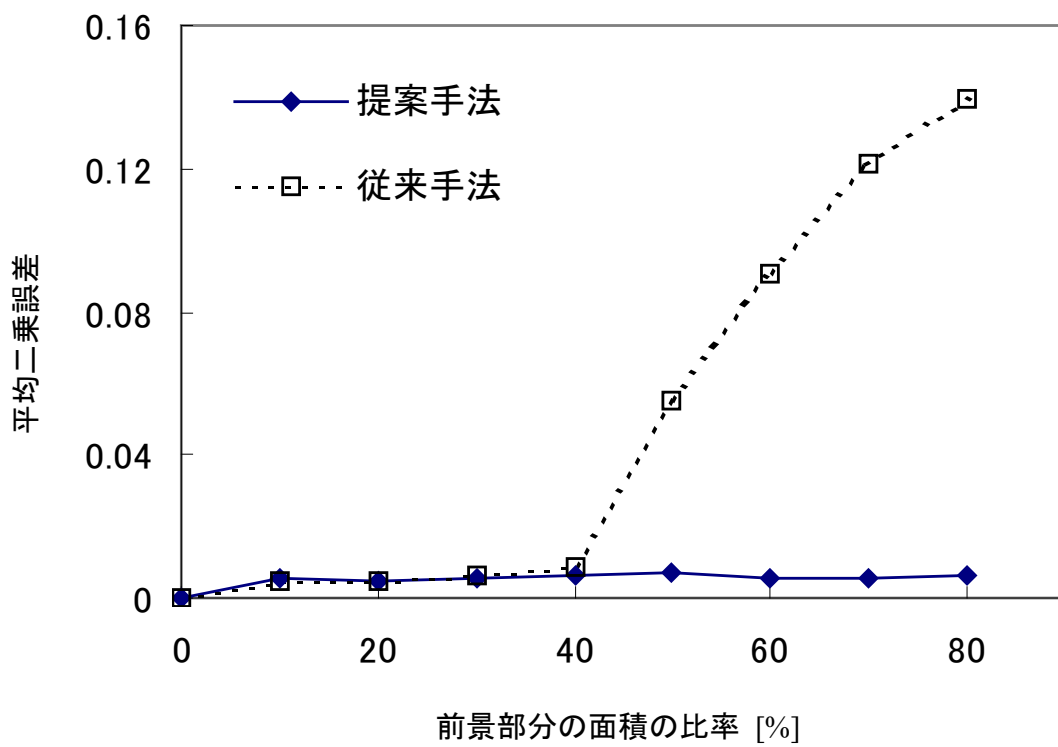


図 11 提案手法と従来手法との精度比較: 入力画像と推定した背景画像間の背景領域での二乗誤差

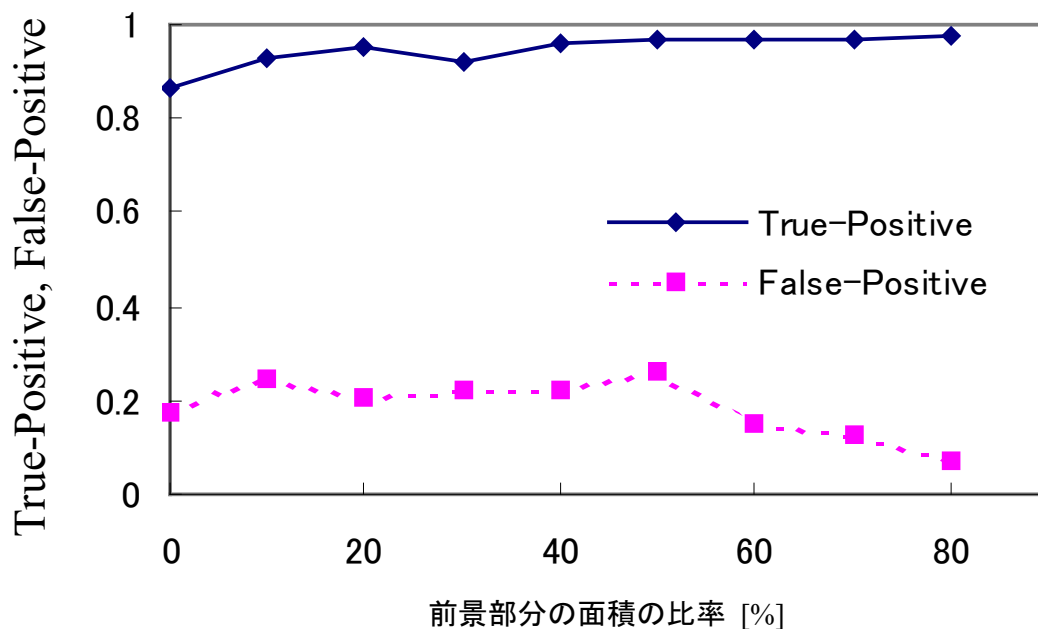


図 12 移動物体領域（扇形領域）の面積比率を変えて提案手法を適用した移動物体領域検出精度

でも、提案手法が良好に機能していることが分かる。次に、図 9 および図 10 に示した移動物体領域の検出精度を評価した。まず、画像中の移動物体領域を手作業により抽出した。次に、手作業による抽出結果と背景差分画像とを比較しながら、移動物体領域の検出結果に関して True-Positive (TP) および False-Positive (FP) を調べた。さらに同様の処理を、扇形領域の面積を変化させながら、それぞれの場合について行った。提案手法と従来手法のそれぞれに関する計算結果を図 12 および図 13 に示す。実線が True-Positive を、点線が False-Positive を表わしている。

提案手法（図 12）では、TP は 0.9 前後、FP は 0.2 前後の値で、移動物体領域の割合によらず安定している。それに対し従来手法（図 13）では、移動物体領域の割合が 30% を超えたあたりから FP が大きくなり、50% の時にはその値が 0.9 以上となった。これは背景画像の推定に失敗し、背景領域でも入力画像との差分が検出されたことを意味する。これらの結果から、従来手法ではできなかった画像

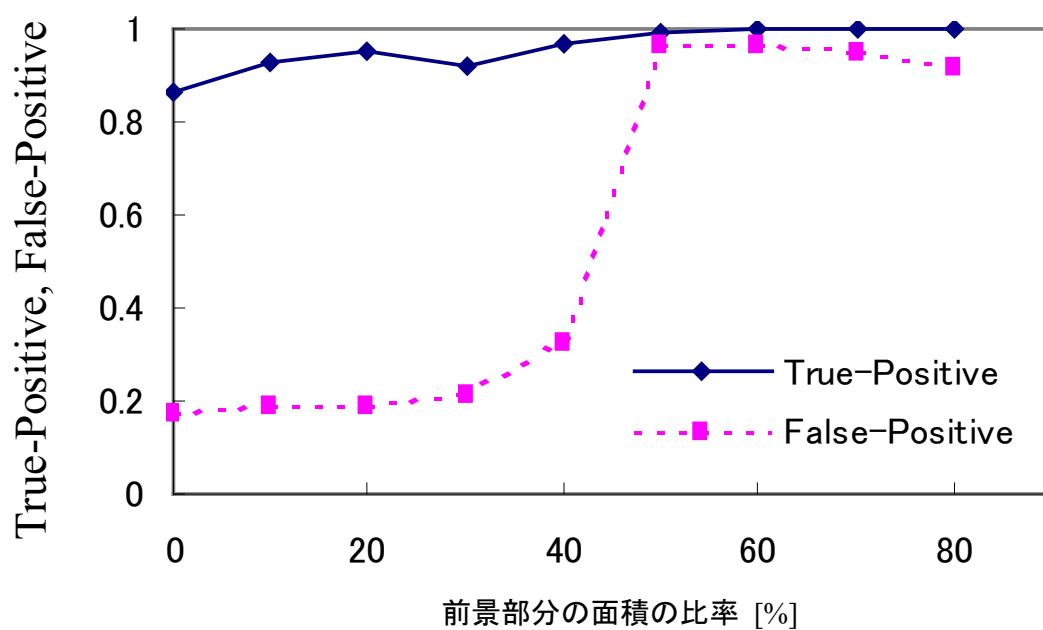


図 13 移動物体領域（扇形領域）の面積比率を変えて従来手法を適用した移動物体領域検出精度



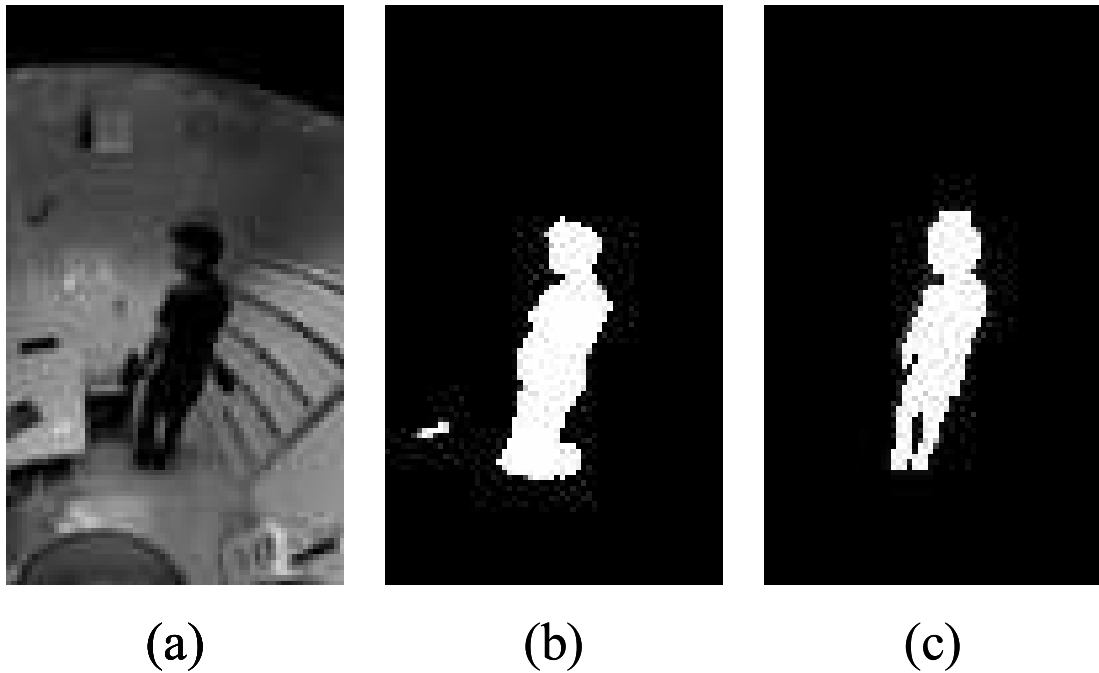


図 14 人物領域の検出結果: (a) 入力画像, (b) 提案手法, (c) 手作業による結果手法

上で背景以外の移動物体領域が占める面積の割合が大きな画像においても, 提案手法では, 移動物体領域の検出精度を落とすことなく, 背景差分法による照明変化に頑健な移動物体領域検出が実現できたことが確認できる.

### 2.2.2.3 照明変化に対する追従性の評価

2.2.2.2 節の実験では, 画像中で背景以外の領域 (扇形領域) が占める割合が小さな場合でも FP が約 0.2 となっている. これは主に次の 2 つが原因と考えられる.

- 壁や床にできた人物の影の誤検出 (図 14)
- 背景画像を推定する際の, 照明変化に対する追従性の不足

後者に関して評価するために, 照明を変えた際の重み係数の適応結果を図 15 に示す. シミュレーションに用いた入力画像は, 図 7 に示した背景画像 (a) および (b)

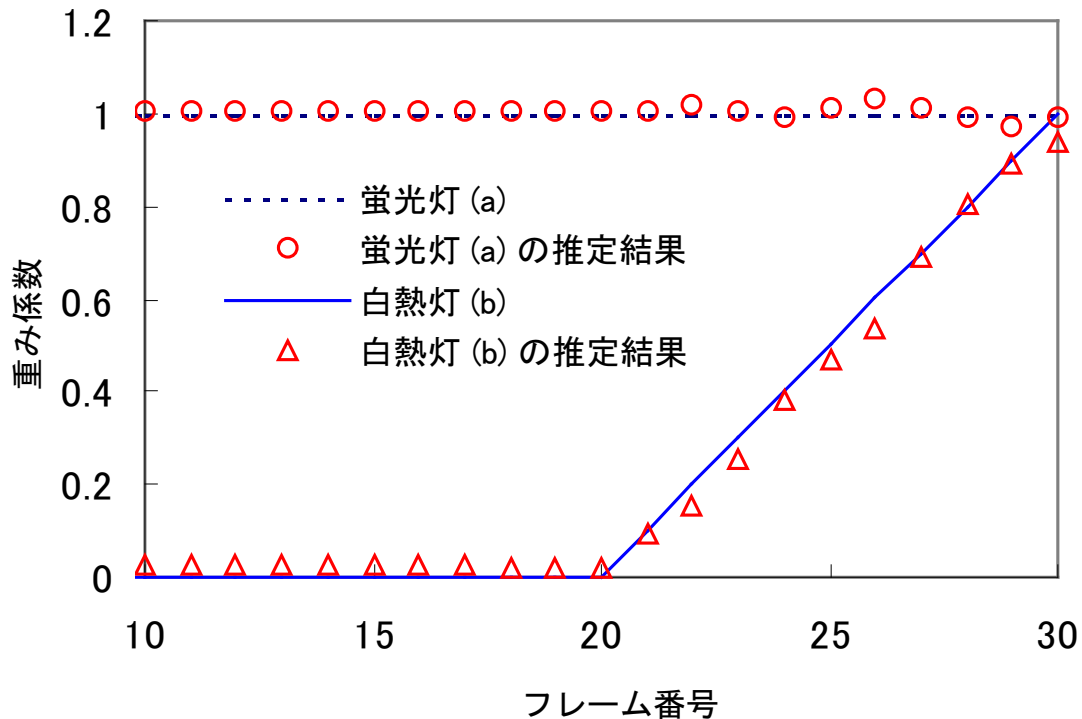


図 15 白熱灯 (b) の明るさを緩やかに変化させて提案手法を適用した場合の重み係数の追従性

を用いて人為的に合成した画像を用いた。初めに蛍光灯 (a) のみが点灯している状態から、白熱灯 (b) を徐々に明るくした状況の画像を合成した。図 15 のグラフ上の点線および実線が、それぞれ設定した照明 (a)(b) の明るさの推移を示している。白熱灯 (b) の変化の速さは 2.2.2.2 節の実験での状況に合わせ、約 3 秒で明るさ最大となるように線形に変化させた。合成した画像に対して提案手法を適用し、照明条件を推定した結果 ((4) 式の重み,  $s_{t,n}^1, s_{t,n}^2$ ) を○および△として図 15 上にプロットした。

推定時の条件は 2.2.2.2 節の実験と同じで、 $N = 100$  である。この結果より、シミュレーションで設定した程度の状況では多少の誤差は見られるが、照明変化に対してほぼ追従できていることが分かる。2.2.2.2 節の実験結果と合わせて考える

と、雲の動きにより窓からの入射光の明るさが変化するという状況においても、提案手法を適用することでその効果が得られると思われる。照明が急激に変化する場合（例えば、蛍光灯を点灯・消灯したとき）、探索空間での解の移動距離は大きくなる。

図 16,17 は図 15 の 2 倍の速さで照明が変化したときの状況を再現した結果である。背景候補画像のサンプル数が少ないと追跡精度が低下してしまう場合がある（図 16,  $N = 100$ ）。カルマンフィルタなどを用いた場合には、解を一度見失ってしまうと追跡の継続が困難である。しかし、提案手法は逐次モンテカルロ法に基づき常に複数の解の候補を保持し、また解を見失った場合には探索範囲が自動的に広げられるため、途中で追跡に失敗しても、自動的に回復することができ（図 16 の 20 ~ 30 フレーム）、継続的な移動物体検出が可能である。

高い追跡精度を維持するためには、推移確率の標準偏差を大きくし、またサンプル数を増やす必要がある（図 17 は  $N = 400$  の場合）。

### 2.2.3 考察

2.2 節では、背景の大部分の隠蔽と照明の変化に頑健な移動物体領域検出を目的として、背景の大部分の隠蔽に頑健な距離特徴と逐次モンテカルロ法を用いた背景差分法を提案した。

移動物体領域が 50% を超える場合では従来手法では正しく背景が推定できなかったが、提案手法ではほぼ正しく背景が推定できた。提案手法では移動物体領域が 80% に達しても正しく背景が推定できた。

また 100 通りの背景候補画像を用いた場合、照明変化に対しても白熱灯を 3 秒程度かけて消灯の状態から明るさ最大となるように変化させたときはほぼ照明変化に追従して背景を推定できた。同様に 100 通りの背景候補画像を用いた場合に、先ほどの 2 倍の早さ、1.5 秒程度で白熱灯を消灯の状態から明るさ最大となるように変化させたときは照明変化に追従できなかった。ただし逐次モンテカルロ法の効果により、解を見失った場合には探索範囲が自動的に広げられるため、途中で追跡に失敗しても、照明の変化が緩やかになれば、自動的に回復することができる。背景候補画像の数を 400 通りにした場合には 1.5 秒程度で白熱灯を消灯の状態か

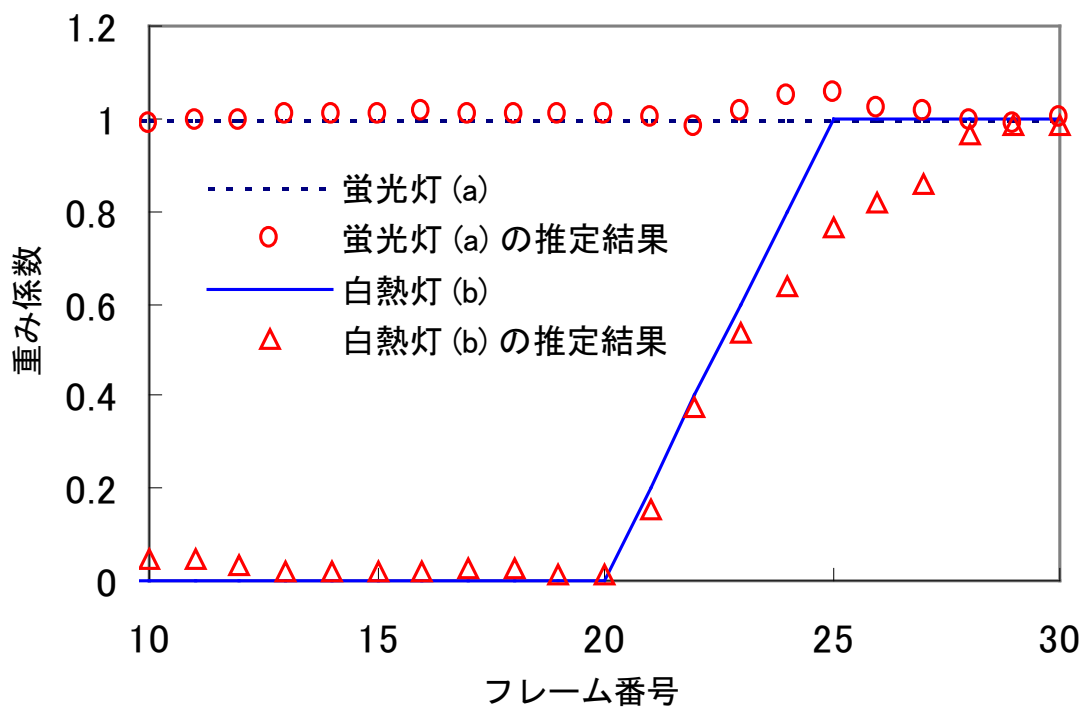


図 16 白熱灯 (b) の明るさを急激に変化させて提案手法を適用した場合の重み係数の追従性 (サンプル数:  $N = 100$ )

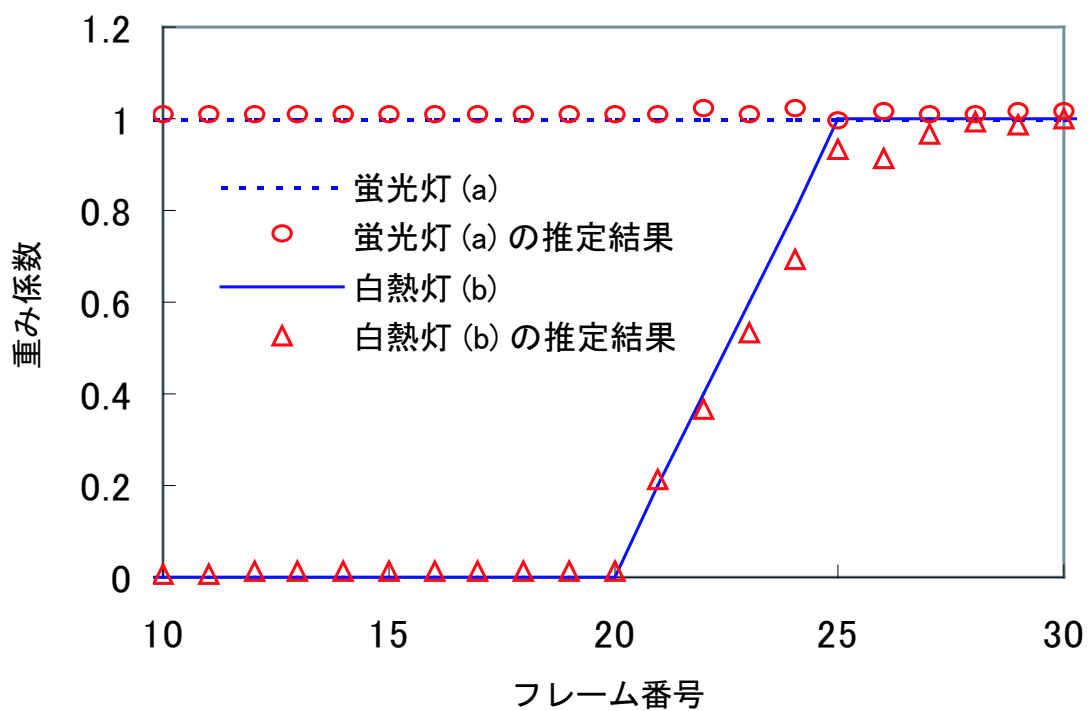


図 17 白熱灯 (b) の明るさを急激に変化させて提案手法を適用した場合の重み係数の追従性 (サンプル数:  $N = 400$ )

ら明るさ最大となるように変化させたときにも照明変化に追従できた。高い追跡精度を維持するためには、推移確率の標準偏差を大きくし、またサンプル数を増やす必要がある（図 17 は  $N = 400$  の場合）。解の追跡精度が低下した場合の回復性能、推移確率とサンプル数の最適化については、今後検討すべき課題である。

提案手法では逐次モンテカルロ法を用いているため、100 通りの重み係数の候補を使って追跡する場合、1 フレームあたりの背景画像推定に要する計算時間も探索範囲内の全組み合わせを計算する場合（1 フレーム 10 分以上）に比べ約 1/2000 の毎秒 3 フレームに短縮された。

本手法によりある程度の照明の変化に対して頑健に背景を生成し移動物体領域を検出することができるようになった。しかしながら、全照明の消灯や隠蔽など、画像情報が使えない場合には追跡が継続できない問題がある。次節では画像情報だけでなく、音情報も利用し、いずれかの情報が断絶した場合にも継続して追跡を継続する手法を述べる。

## 2.3 画像情報と音情報の動的線形和による隠蔽に頑健な移動物体検出

2.2 節では、様々な背景を仮定して、その中から現状に適合した背景モデルを選択して移動物体領域の検出を行った。しかしながら画像情報のみから検出を行うために、隠蔽などの影響により画像情報が断絶した場合、検出を継続できない。人物や車といった音を発する移動物体の場合、その発する音の情報も利用できれば、移動物体の検出を継続できる状況も多く存在する。

本節では、隠蔽や音の途切れなどの影響により情報の断絶が生じるような状況下においても、複数のモーダル（本節で扱うのは音と画像）を用いることによって頑健に物体を追跡し続ける手法を提案する [TKMT04a]。提案手法では、物体の存在に関する尤度分布を、画像情報および音情報のそれぞれから求めた物体の確率密度分布の尤度による重み付き線形和とし、その尤度は周辺環境の変動に応じて変化させる。また、各時刻における尤度分布と確率密度分布は、その時刻の観測結果から算出されたものと、過去の分布との重み付き線形和で表現する。これによ

り、雑音や照明の短期的な変動、隠蔽等による追跡対象の一時的な不検出に対して頑健な追跡が可能となる。今回は処理を簡便にするため、画像情報に関しては背景差分法を、音情報に関しては白色化相互相関法（CSP 法: Cross-power Spectrum Phase analysis）[NYNS00] を用いて、確率密度分布を求める。背景差分法には 2.2 節の手法を用いることも可能である。実験では、円形マイクロホンアレイと全方位カメラからなるシステムに提案手法を適用し、声の途切れや隠蔽が生じる状況においても、安定かつ継続的な物体の追跡が可能であることを示す。

2.3.1 節では、従来手法である背景差分法による物体検出手法、CSP 係数加算法による音源検出手法について述べる。2.3.2 節では、確率密度分布の尤度による重み付き線形和により画像情報と音情報を統合した物体の追跡手法を提案する。そして、2.3.3 節で実験により提案手法の有効性を示した後、2.3.4 節で提案手法について考察する。

### 2.3.1 画像による物体検出手法と音による物体検出手法

#### 2.3.1.1 背景差分法による物体検出

背景差分法は物体の実時間検出における簡便かつ効果的な手法の一つであり、ビデオ監視システム等で用いられている。通常、各時刻での背景画像が既知である状況で用いられるが、最近では照明変化に対しても頑健な手法 [MWHT01, TKMT03b, TKMT04b] も開発されており、その適用範囲が広がりつつある。背景差分法では、あらかじめ物体が写っていない背景画像を取得しておく。そして各時刻での入力画像と背景画像を比較し、有意な差が確認された画像上の領域を移動物体領域として検出する。以下では、時刻  $t$  における物体の存在確率密度分布  $\bar{P}_t^V(\phi)$  の求め方を述べる。初めに、背景画像  $I^{back}$  と時刻  $t$  における入力画像  $I_t$  の差分を計算し、2 値の背景差分画像  $I_t^{sub}$  を作成する。座標  $(x, y)$  の画素に注目し、その差分が閾値  $\sigma$  以上の場合には入力画像の画素値を 1、それ以外の場合には 0 とする。

$$\begin{aligned} I_{t,(x,y)}^{sub} &= 1 \quad \text{when} \quad |I_{t,(x,y)} - I_{(x,y)}^{back}| \geq \sigma \\ I_{t,(x,y)}^{sub} &= 0 \quad \text{when} \quad |I_{t,(x,y)} - I_{(x,y)}^{back}| < \sigma \end{aligned} \quad (8)$$

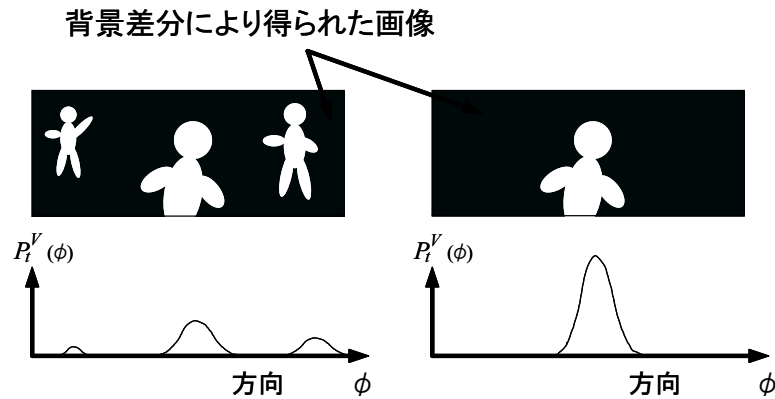


図 18 背景差分画像から得られた確率密度分布

以上の処理を全ての画素について行うことで、背景以外の物体領域を検出することができる。一般に水平に設置されたカメラの場合には、背景差分画像の  $x$  軸,  $y$  軸方向が、実空間上での方位, 高さに対応する。そこで全ての  $x$  座標 (方位  $\phi$ ) に関し、縦方向の画素値を全て足し合わせ、全方位に関して正規化したものを確率密度分布  $\bar{P}_t^V(\phi)$  とする。

$$\bar{P}_t^V(\phi) = k_v \sum_y I_{t,(\phi,y)}^{sub} \quad (9)$$

$k_v$  は正規化定数である。

### 2.3.1.2 円形マイクロホンアレイによる音源検出

CSP 法 (Cross-power Spectrum Phase analysis: 白色化相互相関法) [NYNS00] は図 19 のように 2 本のマイクロホン  $i, j$  で受信した信号  $s_i(n), s_j(n)$  より CSP 係数および到達時間差  $\tau$  を推定し、音源の方向  $\phi$  を推定する手法である。CSP 係数, 到達時間差および音源の方向は次式のように求められる。

$$C_{i,j}(k) = DFT^{-1} \left[ \frac{DFT[s_i(n)]DFT[s_j(n)]}{|DFT[s_i(n)]||DFT[s_j(n)]|} \right] \quad (10)$$

$$\tau = \arg \max_k (C_{i,j}(k)) \quad (11)$$

$$\phi = \sin^{-1}(c\tau/d) \quad (12)$$



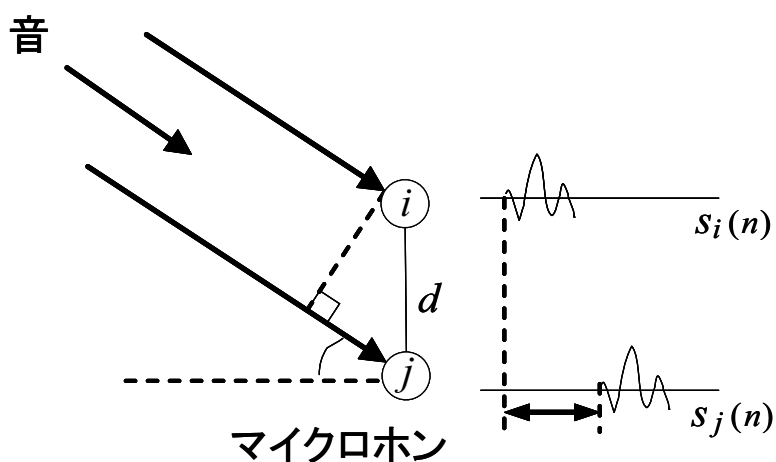


図 19 CSP 法による方向推定

$C_{i,j}(k)$  は CSP 係数,  $k, n, \tau$  は時間,  $c$  は音速,  $d$  はマイクロホン間の距離である.  $DFT[\cdot]$  および  $DFT^{-1}[\cdot]$  はそれぞれ, 離散フーリエ変換および逆変換を表わす.

この手法を複数のマイクロホンペアに適用し推定精度を向上させた例の一つに加算 CSP 法があり, マイクロホンを円形状に配置した円形マイクロホンアレイ (図 20) への適用も行われている [NNL+02]. ここでは精度の最も高い直径方向に並んだマイクロホンペアの中で異なった角度を持つペアの CSP 係数を加算し, 得られた新たな CSP 係数 ( $\bar{P}_t^A(\phi)$ ) を用いて音源の方向を推定する.

### 2.3.2 確率密度分布の重み付き線形和による物体追跡

提案手法は, 画像情報と音情報の統合と, 過去の情報を用いた確率密度分布の更新により構成される. まず, 画像情報および音情報から算出した確率密度分布 ( $P_t^V(\phi), P_t^A(\phi)$ ) の尤度による重み付き線形和として統合し, 各尤度を状況に合わせて動的に変化させる手法について述べる. 次に, ある時刻における確率密度分布と尤度分布を, その時刻での観測結果から算出された分布と過去の分布の重み付き線形和として更新する手法について述べる. このアプローチは, ビデオ監視システム等における背景画像の更新においてしばしば用いられる手法であり [HT00],

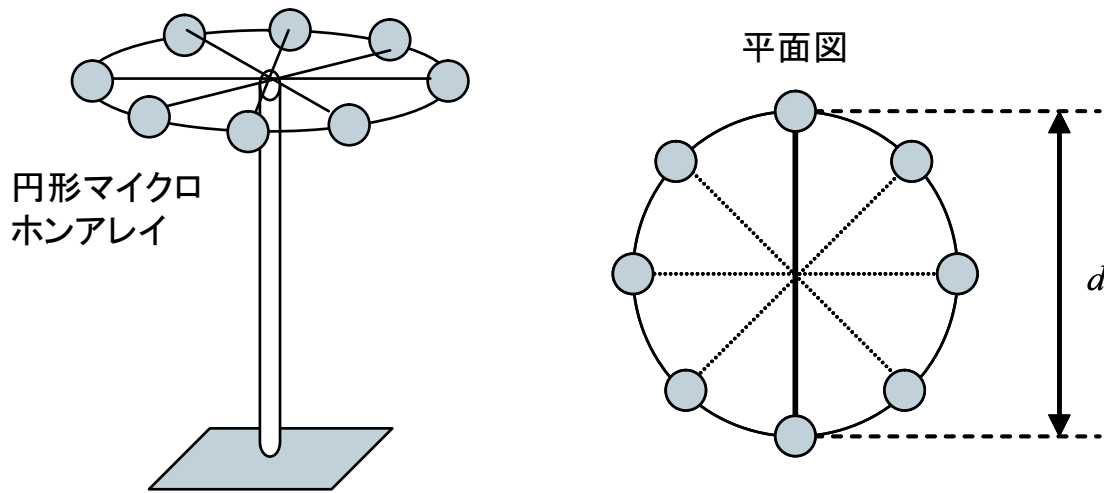


図 20 円形マイクロホンアレイ

過去の情報を一定時間保持していることから、一時的かつ突発的な観測結果の乱れに対して頑健であり、推定結果の S/N 比が向上する、といった利点を持つ。

### 2.3.2.1 画像情報と音情報の統合

時刻  $t$  における尤度分布  $\bar{F}_t(\phi)$  を  $P_t^V(\phi)$  および  $P_t^A(\phi)$  の重みつき線形和として次式のように定義する。

$$\bar{F}_t(\phi) = k_1(t)P_t^V(\phi) + (1 - k_1(t))P_t^A(\phi) \quad (13)$$

$k_1(t)$  は尤度を表す変数で、 $1 \geq k_1(t) \geq 0$  の範囲で各情報の状態に応じて変化する。尤度  $k_1(t)$  の求め方を以下に述べる。まず、画像情報から移動物体の方向検出が困難な状況について考える。照明変動により一時的に背景の様子が変化した場合には背景の一部が移動物体領域と誤検出され、その結果  $P_t^V(\phi)$  の分散が大きく、最大値は小さくなる。そして照明変動が収まり誤検出が減少すると、最大値は大きくなる。また異なる状況を考えると、室内にいる物体の数が多いほど  $P_t^V(\phi)$  の最大値は小さくなる（図 18）。すなわち、画像情報から移動物体の方向を特定することが困難な状況では、 $P_t^V(\phi)$  の最大値が小さくなる。同様に、雑音等の影響により音情報から音源方向を特定することが困難な状況では、 $P_t^A(\phi)$  の最大値が小

さくなる. したがって,  $P_t^V(\phi)$  および  $P_t^A(\phi)$  の最大値に注目することで, 各情報の信頼性を推定することができる. 尤度  $k_1(t)$  を次式のように定義する.

$$k_1(t) = \frac{\max(P_t^V(\phi))}{\max(P_t^V(\phi)) + \max(P_t^A(\phi))} \quad (14)$$

これにより, 確率密度分布の最大値が大きい入力情報ほど信頼性が高いとみなされ, その情報に重みをおいた処理が行える.

### 2.3.2.2 時間情報を用いた尤度分布と確率密度分布の更新

a) **尤度分布の更新** ある時刻  $t$  における尤度分布  $F_t(\phi)$  を, 時刻  $t-1$  における尤度分布  $F_{t-1}(\phi)$  と, 時刻  $t$  における観測結果から求めた尤度分布  $\bar{F}_t(\phi)$  を用いて次式のように定義する. なお  $\phi$  は方位を表わす.

$$F_t(\phi) = k_2 F_{t-1}(\phi) + (1 - k_2) \bar{F}_t(\phi) \quad (1 \geq k_2 \geq 0) \quad (15)$$

$k_2$  は定数である. これにより  $\bar{F}_t(\phi)$  が値をもたなかった場合でも, 過去の尤度分布  $F_{t-1}(\phi)$  が保持されているため,  $F_t(\phi)$  は値を持ち続ける. これは, 時間的に尤度分布を平滑化することを意味し, 突発的なノイズ等の影響より  $\bar{F}_t(\phi)$  に誤差が含まれる場合にはその影響を抑制することができる. また急激な変化には反応がやや遅れるものの,  $k_1(t)$  および  $k_2$  の値を適当に設定することで, 対象物体が一時的に検出できない期間での継続的な追跡に応用することができる.

b) **画像情報に関する確率密度分布の更新** 背景差分法の結果を用いた物体追跡は, 物体の検出が安定して行われる限りは良好に動作する. しかし, 隠蔽の発生, 一時的な照明変動などにより, 物体が検出できない, もしくは検出精度が低下した場合にはそれが短時間であっても, 動作が不安定になる. そこで  $\bar{P}_t^V(\phi)$  および時刻  $t-1$  の確率密度分布  $P_{t-1}^V(\phi)$  を用いて, 時刻  $t$  における確率密度分布  $P_t^V(\phi)$  を次式のように定義する.

$$P_t^V(\phi) = v P_{t-1}^V(\phi) + (1 - v) \bar{P}_t^V(\phi) \quad (1 \geq v \geq 0) \quad (16)$$

$v$  は定数である。  $P_t^V(\phi)$  を利用し、  $v$  の値を適切に設定すれば、 隠蔽や照明変動による画像情報の信頼性の低下が短時間である場合には、 以前の時刻の推定結果に基づき継続的な追跡が可能である。

**c) 音情報に関する確率密度分布の更新** 各マイクロホンからの出力信号の処理では、 既存技術である加算 CSP 法 [HT00] を適用し、 加算後の CSP 係数を用いて物体位置の確率密度分布  $P_t^A(\phi)$  を用いて、 物体の追跡を行う。 周辺雑音が存在する環境において物体の追跡を行う場合、 語句間や会話の切れ目の音が途切れる区間において雑音の影響を受け  $P_t^A(\phi)$  が不連続に変化し、 安定した追跡が困難である。  $P_t^A(\phi)$  を時間の経過とともに滑らかかつ連続的に変化するように定義できれば、 より安定した物体の追跡が可能になる。 そこで、  $P_t^A(\phi)$  を以下のように定義する。

時刻  $t$  において観測された音信号値から求められた、 加算後の 360 度分の CSP 係数をパワーで正規化したものを  $\bar{P}_t^A(\phi)$  とする。 そして確率密度分布  $P_t^A(\phi)$  を、  $\bar{P}_t^A(\phi)$  および時刻  $t-1$  における確率密度分布  $P_{t-1}^A(\phi)$  を用いて次式のように表わす。

$$P_t^A(\phi) = aP_{t-1}^A(\phi) + (1-a)\bar{P}_t^A(\phi) \quad (1 \geq a \geq 0) \quad (17)$$

$a$  は定数である。  $P_t^A(\phi)$  を利用し、  $a$  の値を適切に設定すれば、 声の途切れや周辺雑音の増大等による音情報の信頼性の低下が短時間である場合には、 以前の時刻の推定結果に基づき継続的な追跡が可能である。

### 2.3.2.3 処理の流れ

フローチャートを図 21 に示す。 提案手法では、 音情報・画像情報のそれぞれについて、 前時刻での推定結果と現時刻での観測結果を元に現時刻での確率密度分布を推定する。 この推定結果を尤度を用いて足し合わせて音・画像情報での観測結果とする。 音・画像情報を組み合わせた現時刻での物体方向の確率密度分布は、 前時刻での推定結果と現時刻での観測結果を元に推定する。

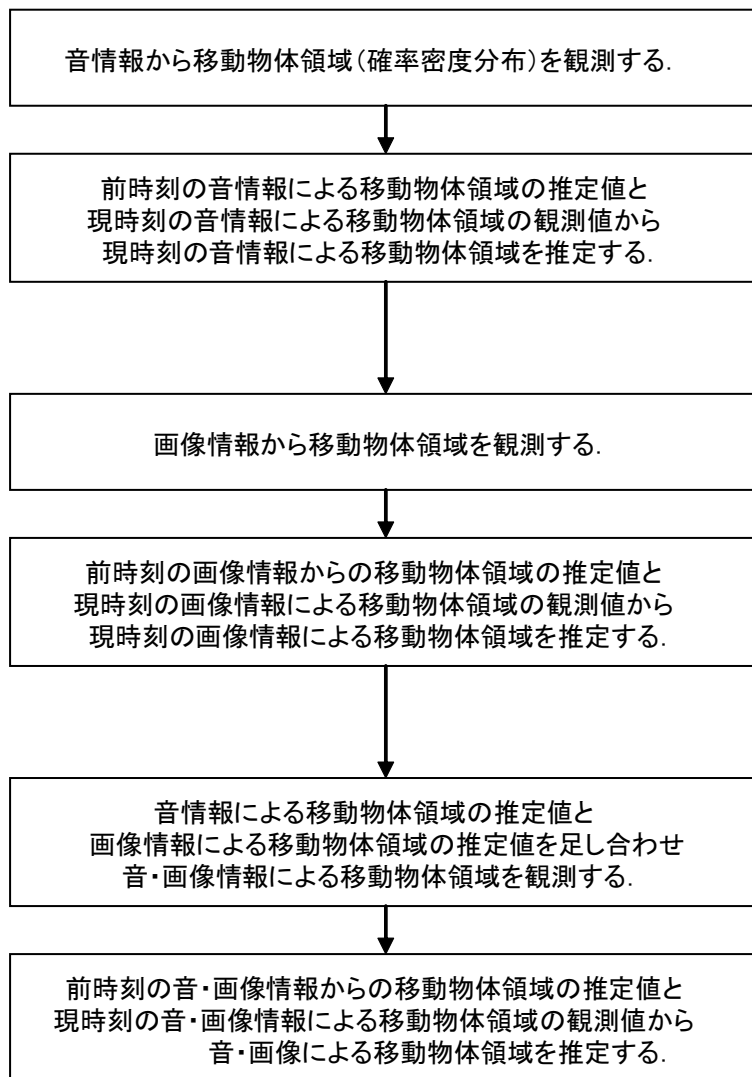


図 21 音情報と画像情報を用いた物体追跡法のフローチャート

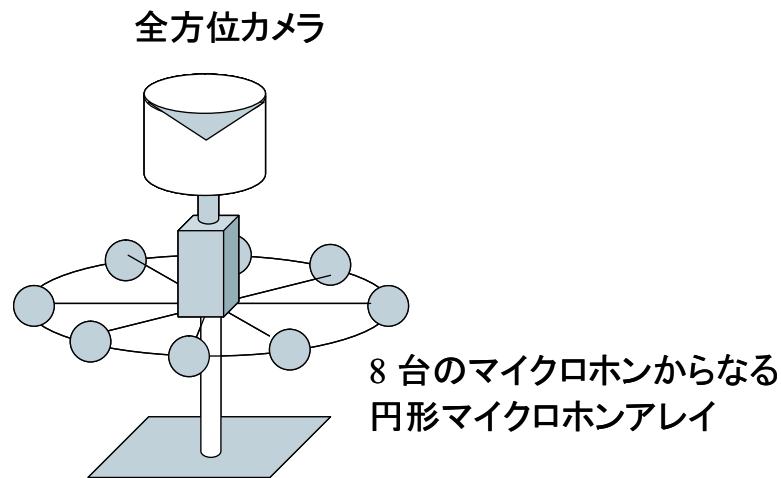


図 22 音情報と画像情報の統合に関する実験で用いたシステムの構成

### 2.3.3 実験

一時的な照明変動や周辺雑音の大きさが変化する環境での物体の追跡に提案手法を適用し、隠蔽や音声の途切れが生じた場合でも安定かつ継続的な追跡が可能なることを確認するために実験を行った。画像情報と音情報の取得には、図 22 のように配置した全方位カメラと円形マイクロホンアレイを用いた。全方位画像の撮影に用いた全方位カメラはデジタルカラー CCD カメラ (SONY 製 DFW-SX900) と凸面の双曲面ミラー (アコウル製) で構成され、毎秒 10 フレームで画像を取得した。また音情報の取得には、8 本のマイクロホンを均等に配置した円形マイクロホンアレイ (直径  $d = 25[\text{cm}]$ ) を使用し、カメラに同期させながら 48 kHz のサンプリングレートで音を収集した。なお、各装置の制御および処理は Intel Pentium 4 2.4GHz を搭載した PC 1 台で行い毎秒約 10 フレームで動作する。実験を行った室内の環境を図 23 に示す。実験中には照明条件は変化しないものとし、背景差分法で用いる背景画像は予め取得しておいたものを用いる。雑音源としては室内空調の他にオーディオスピーカを設置し、そこから音楽を小音量で流した。ついでの大きさは横幅 1.3m、高さは 1.6m である。また、マイクロホンアレイとついでとの距離は 1.5m、ついでと雑音源との距離は 2.0m とした。対象となる移

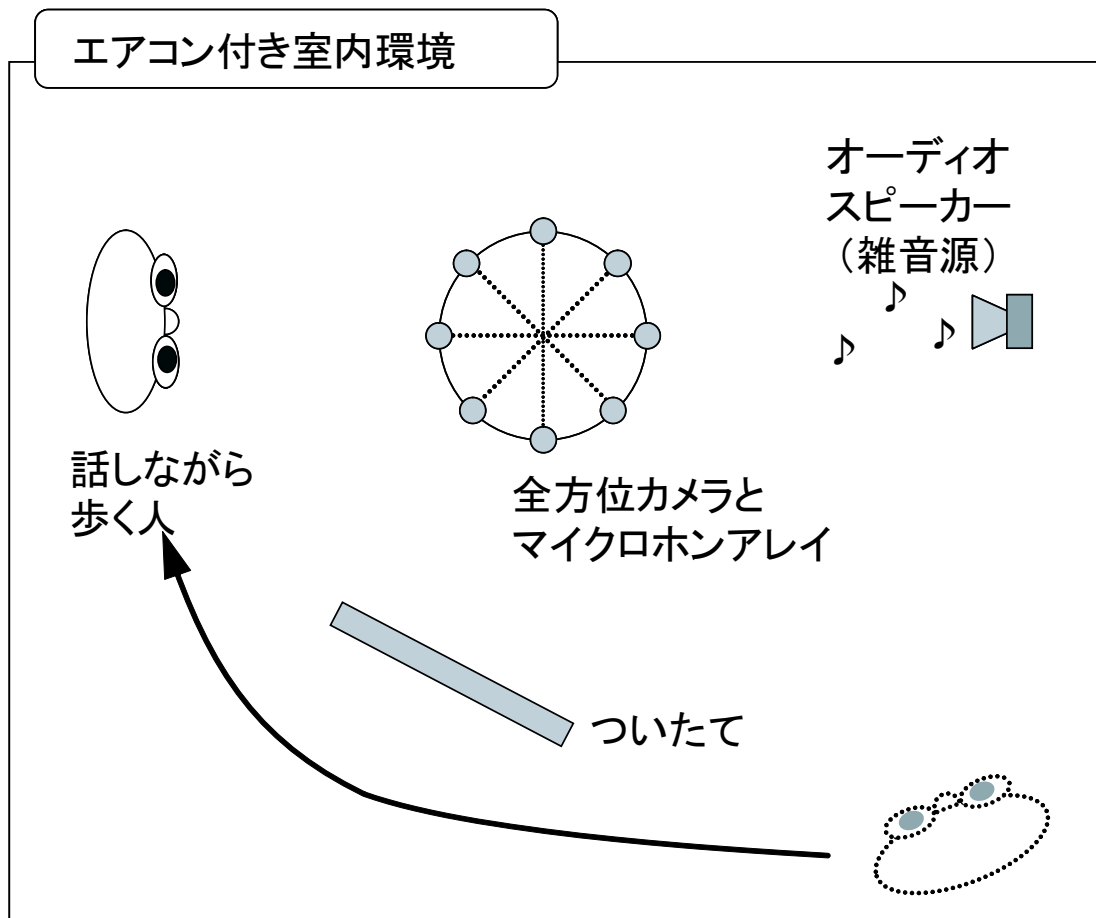


図 23 音情報と画像情報の統合に関する実験を行った室内環境

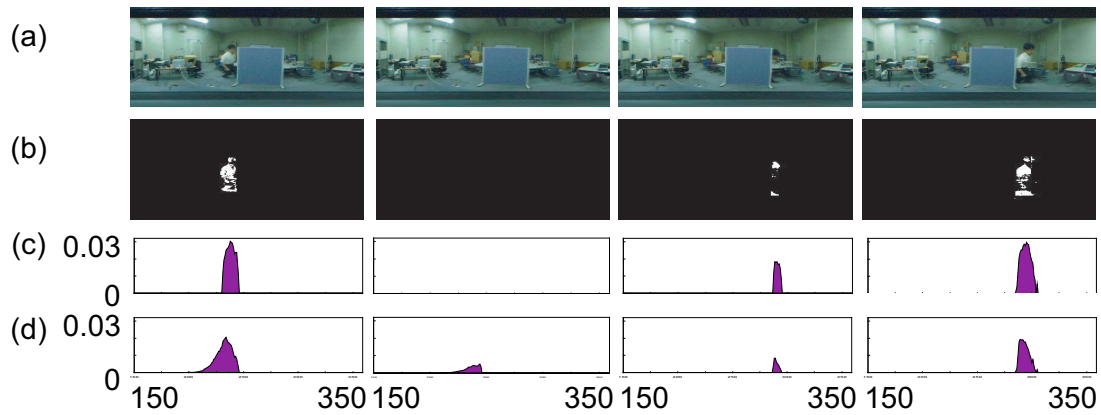


図 24 画像情報による物体方向の追跡結果: (a) 入力画像 (b) 背景差分画像  
(c)  $\bar{P}_t^V(\phi)$  (d)  $P_t^V(\phi)$ (提案手法)

動物体には話しながら歩く人物を用いた。移動物体はマイクロホンアレイを中心に約 2m の距離を等速（毎秒約 1m）で移動させた。

実験は 3 段階に分けて行う。初めに、2.3.3.1 節において画像情報のみを用いた物体追跡を行い、隠蔽もしくは照明変動が生じた場合に対して、2.3.2.2 b) 節で述べた手法の効果を確認する。次に、2.3.3.2 節において周辺雑音がある環境において音情報のみを用いた物体追跡を行い、声の途切れが生じた場合に対する 2.3.2.2 c) 節で述べた手法の効果を確認する。そして最後に、2.3.3.3 節において画像情報と音情報を統合した物体の追跡を行い、2.3.2.2 節 a) で述べた提案手法の有効性を確認する。

### 2.3.3.1 画像情報による物体追跡

初めに、隠蔽により一時的に物体が背景差分法により検出されない場合について、2.3.2.2 b) 節の提案手法を用いた物体追跡を行った。処理において、(16) 式の定数は  $\nu = 0.8$  とした。図 24 に実験で使用した入力画像列、背景差分画像および、画像から計算により求めた  $\bar{P}_t^V(\phi)$  と  $P_t^V(\phi)$  の一部を示す。

左から 2 列目では物体がついたての後ろを通過しており、隠蔽が発生している



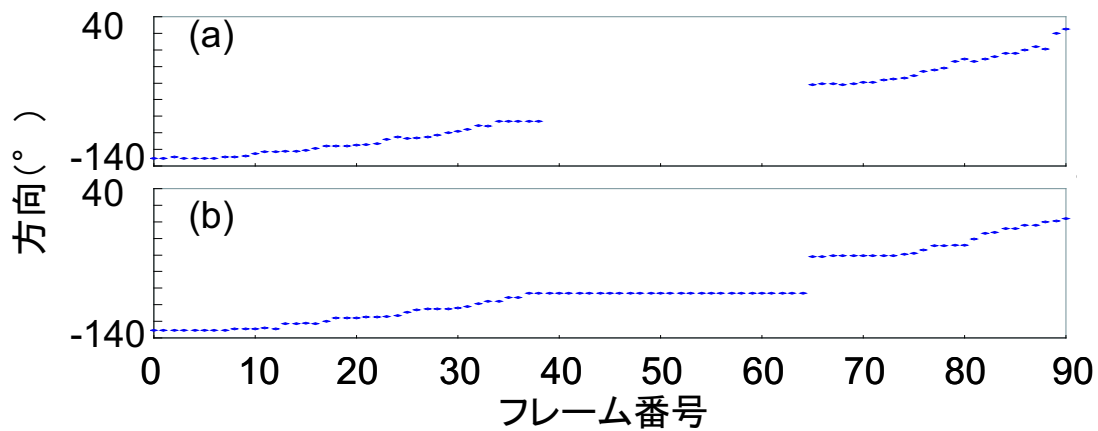


図 25 画像情報による物体方向の追跡の経過: (a)  $\bar{P}_t^V(\phi)$ , (b)  $P_t^V(\phi)$  (提案手法)

状況で、隠蔽の発生から 6 フレーム目の画像である。なお図 24(a)(b) の画像は、全方位画像から作成したパノラマ画像の一部である。図 24(c) に示した  $\bar{P}_t^V(\phi)$  を見ると、隠蔽発生中には値を持っておらず、物体の位置を特定することができない。一方で、図 24(d) に示した  $P_t^V(\phi)$  は隠蔽発生中でも値を持っており、これにより物体の位置が提示されていることが確認できる。図 25 に物体追跡の結果を示す。

フレーム番号 40 ~ 65 の間で隠蔽が生じている。(a) は  $\bar{P}_t^V(\phi)$ , (b) は  $P_t^V(\phi)$  による検出結果を示している。(a) では隠蔽発生時に物体が検出できないため、その間は位置情報が得られていない。一方で (b) では、隠蔽発生直前の値ではあるが、物体の位置情報が得られ、(a) と異なり、物体の位置についての情報を完全に失うことはない。

### 2.3.3.2 音情報による物体追跡

周辺雑音がある環境において 2.3.2.2 c) 節に述べた音情報を用いた物体追跡を行った。雑音としては室内空調の他、物体と異なる位置 (120 度) に設置したスピーカーから音楽を流した。物体にはカセットプレーヤを携帯してもらい、テープに記録された英会話を再生した。データのサンプリング間隔を 0.1 秒、(17) 式の定数

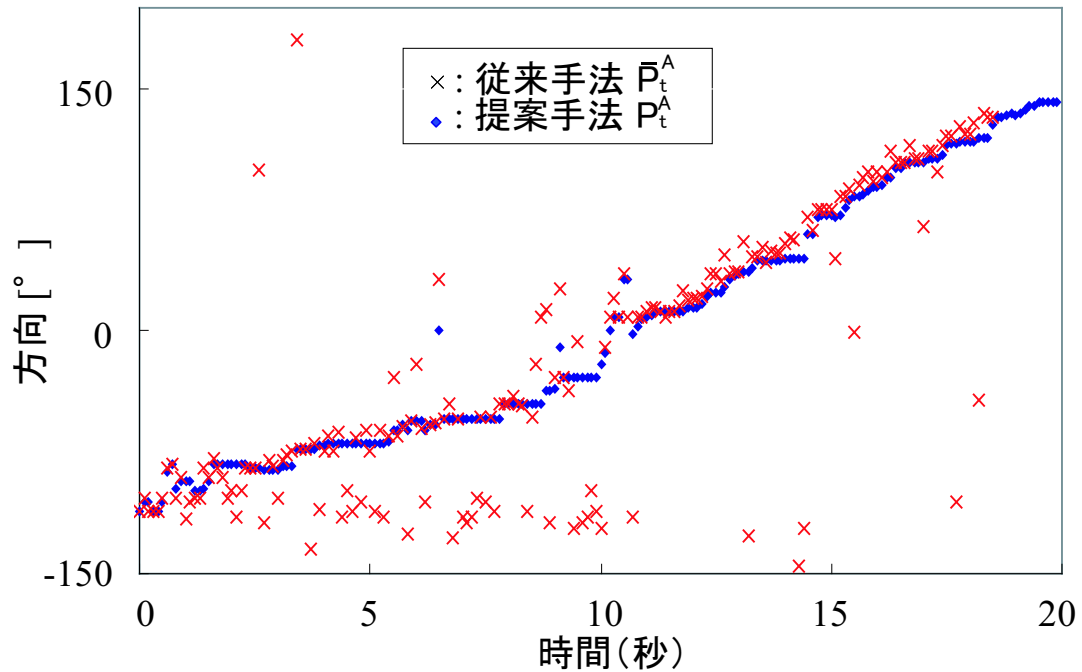


図 26 音源方向の追跡の経過

を  $a = 0.8$  として約 20 秒間処理を行った. 図 26 に  $\bar{P}_t^A(\phi)$  および  $P_t^A(\phi)$  から求めた物体位置の推定結果を示す.

$\bar{P}_t^A(\phi)$  による結果 (x) では, 語句や文の切れ目での音声の途切れにおいて雑音源を検出してしまい, 不安定な追跡結果となった. 一方で提案手法である  $P_t^A(\phi)$  による結果 (◆) では, わずかな区間を除き, 物体を正しく追跡することができた. 雑音源を検出してしまった区間を調べたところ, 文の終わりや, 発言者が切り替わる時の 1 秒前後の音の途切れにおいてであった. また, 推定された物体の方向を比較すると, 提案手法 (◆) の方が値の分散が少ないことが確認できる. これも, 確率密度分布を時間方向に加算平均したことの効果である. また前述の実験とは別に, 音源方向の推定結果をディスプレイに表示し, 音源の移動に対する即応性を主観的に確認したが, 動きに対する遅延は感じられなかった.

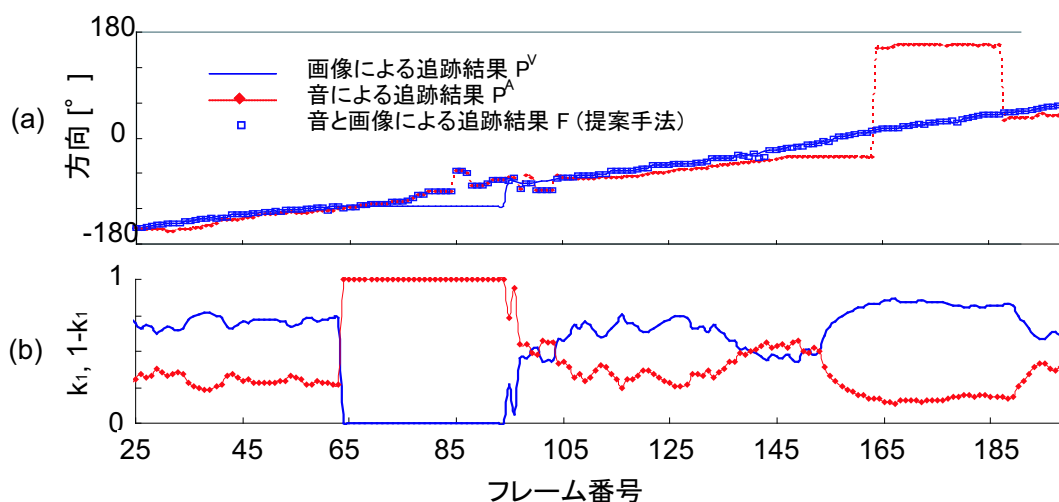


図 27 画像情報と音情報を統合した物体の追跡結果: (a) 推定した物体の方向, (b) 尤度  $k_1$  の変化の様子

### 2.3.3.3 画像情報と音情報を用いた物体追跡

2.3.2.2 a) 節に述べた画像情報と音情報の両方を用いた手法による物体の追跡を行った。図 23 に示す環境において、物体はマイクロホンアレイと全方位カメラを中心に、ほぼ等速度でその周りを移動した。その際、途中で隠蔽や比較的長時間（数秒）の声の途切れを発生させ、一方の情報が利用できない状況を作った。なお処理において、定数  $\nu, a, k_2$  はいずれも 0.8 とした。図 27 に物体の追跡結果を示す。

(a), (b) とともに横軸は画像のフレーム番号であり、時間軸に相当する。65～95 フレームにかけて約 3 秒間隠蔽が生じており、また 155～185 フレームにかけて約 3 秒間の声の途切れが生じている。(a) は各時刻における物体方向の推定結果で、画像情報と音情報を用いた追跡結果（提案手法）を □ で示している。また比較のため、画像情報のみを用いて物体を追跡した結果（背景差分法, 実線）と、音情報のみを用いて追跡した結果（CSP 法, -◆-）を同じグラフ上に示している。(b) は各時刻での尤度  $k_1$  の変化の様子を示している。

(b) の  $k_1$  に注目すると、隠蔽の発生（65 フレーム）と同時に値が小さくなり、終了（95 フレーム）と同時にほぼ元の値に戻っていることが確認できる。これは

画像情報の信頼性の低下, 上昇に伴い, 画像情報に関する確率密度分布の重みが自動的に変化したことを意味する. また声の途切れにより音情報の信頼性が低下した時 (155 フレーム) にも,  $k_1(t)$  が自動的に変化していることが分かる. 隠蔽発生時と声の途切れの発生時とで  $k_1(t)$  の変化の早さが異なる. これは, 隠蔽の発生中には画像情報での観測値が得られないのに対し, 音情報に関しては無音時でも周辺雑音が観測値として得られ, 確率密度分布の変化が緩やかになるためである. 以上の結果から, 提案手法を用いることで, 画像情報と音情報からなるモデル情報の状況に合わせた使い分けが実現し, その結果, 隠蔽や音の途切れが生じた場合でも, 安定した物体の追跡が実現できたことを確認した.

#### 2.3.4 考察

2.3 節では, 雑音や照明変動等の外乱や隠蔽が生じる状況において, 安定かつ継続的な物体の追跡を行うための, 画像情報と音情報を用いた追跡手法を提案した.

提案手法では, 画像情報と音情報を, それぞれから求めた確率密度分布の尤度による重み付き線形和として統合し, その尤度を周囲の状況に応じて変化させることで, 一方の情報の信頼性が著しく低い場合でも安定した精度での追跡を実現した.

本節では, 画像情報と音情報の統合に主眼を置いたため, 統合に関する尤度  $k_1(t)$  のみを変数とし, 時間情報に関する  $k_2, \nu, a$  を定数とした.  $k_2, \nu, a$  の値が大きいほど, 過去の情報が保持される時間が長くなり, 隠蔽や声の途切れが生じた場合にも対応できる. 一方で, それらの値は変化に対する時間的な即応性にも関係しており, 値が大きいほど変化に対して遅延が生じる. 追跡性能をさらに向上させるためには  $k_2, \nu, a$  を変数として扱い, 対象物体の動きや周囲の環境に合わせて, それぞれの値を適応的に変化させることが有効であると思われる. そこで以下では,  $k_2, \nu, a$  の決定法に関する考察を行う.

確率密度分布のピーク値が大きく分布形状が鋭いほど, そのデータは信頼性が高いと仮定できる. そこで, 時刻  $t$  と  $t-1$  での確率密度分布の最大値や分散に注目し, その比を用いる方法が考えられる. この方法は実装が容易という特長を持つ. しかし, 例えば観測信号の S/N が悪い状況下や, 元々鋭いピークをもった

確率密度分布が得られない情報に対しては、最大値や分散が大きくは変化しない。そのため、 $k_2$ ,  $\nu$ ,  $a$  の可変範囲が 0.5 近傍に限定されてしまい、十分な効果が現れない可能性も予想される。したがって今後、様々な環境下における実測データを用いて、その有効性を評価する必要がある。別のアプローチとしては、 $k_2$ ,  $\nu$ ,  $a$  および  $k_1(t)$  を状態変数とみなし、それらの最適な組み合わせを探すとといった方法も考えられる。その場合には EM 法のような再帰的手法や逐次モンテカルロ法の適用が有効と思われる。逐次モンテカルロ法では各状態変数の組み合わせについて複数のパターンを常に仮定し、それらを解の候補として保持し続ける。つまり  $k_1(t)$ ,  $k_2$ ,  $\nu$ ,  $a$  を分布として扱うため、外乱等に起因する観測信号の変動に対し、各値を一意に決める前述の手法と比較して、より頑健であると期待できる。また確率モデルに基づいた効率の良い解のサンプリングが行われるため、計算量の爆発を抑制することができる。計算量の観点から、今後用いるモデルの数が増加した場合に特に効果が現れると期待される。

## 2.4 結言

本章では、照明の変化や隠蔽に対しても頑健に、背景モデルに基づく方法を用いて移動物体領域を検出することを目的とした手法について述べた。

2.2 節では、従来手法の問題点であった背景領域の大部分が隠蔽されている場合（移動物体が画像中の大部分の領域に映る場合）にも正しく背景を推定し移動物体領域を検出することができる背景差分法を提案した。各照明条件下における背景画像は、あらかじめ用意した様々な照明条件下の背景画像の重み付き線形和として生成する。提案手法では、入力画像と背景画像の同じ座標の画素に注目し、その値が一致した画素の数を評価値として導入し画像中の背景の大部分が隠蔽されている場合にも背景の推定が可能になった。この評価値は、各蓄積画像の重み係数の推定に最小二乗法のような解析的な手法が使えないため、全ての重み係数の組み合わせについて調べる必要がある。全ての重み係数の組み合わせについて評価値を計算すると計算コストが膨大になるため、確率モデルに基づくサンプリングおよびその更新を行う逐次モンテカルロ法を導入し、各蓄積画像の重み係数の追跡問題として扱った。結果、緩やかな照明変化に対しては精度を維持しつつ計

算量の削減が実現できた。しかしながら、2.2節の手法には逐次モンテカルロ法を使って重み係数の追跡を行う中で急激な照明の変化が起こった場合、その変化を追跡できない場合がある。また全照明の消灯や物体の隠蔽などが起こり画像情報が使えない場合には、移動物体領域が検出できなくなる問題がある。

追跡する移動物体は、人物や車のように音を発することも多いため、2.3節では画像情報だけでなく、音情報も利用し、いずれかの情報が雑音や照明変動等の外乱により隠蔽された場合にも安定かつ継続的な移動物体の追跡を行う手法を提案した。提案手法では、画像情報と音情報を、それぞれから求めた確率密度分布の尤度による重み付き線形和として統合し、その尤度を周囲の状況に応じて変化させることで、一方の情報の信頼性が著しく低い場合でも安定した精度での追跡を実現した。さらに各情報の統合により得られた尤度分布に関し、過去の分布との重み付き線形和をとることで、S/Nの向上に加えて、尤度分布が連続的に変化するようになった。これにより、画像情報と音情報の両方で移動物体が検出できない状況があっても、それが短時間であれば継続的な追跡ができると考えられる。実験では、円形マイクロホンアレイと全方位カメラを用いて取得した情報を用いた移動物体の追跡を行い、周辺雑音がある環境において、隠蔽や声の途切れが生じた場合でも、継続的な追跡が実現できることを示した。

本章で示した2つの手法は、移動物体が何であるかを識別することができず、背景ではない移動物体領域であることしかわからない。物体が何であるかを識別して検出するためには検出対象モデルに基づく物体検出が必要である。本章の手法は、物体が何であるかを識別して検出する前処理として、照合すべき探索領域の絞り込み処理として用いることもできる。

次章では、本章の手法で解決できない物体が何であるかを識別する検出対象モデルに基づく物体検出手法について述べる。

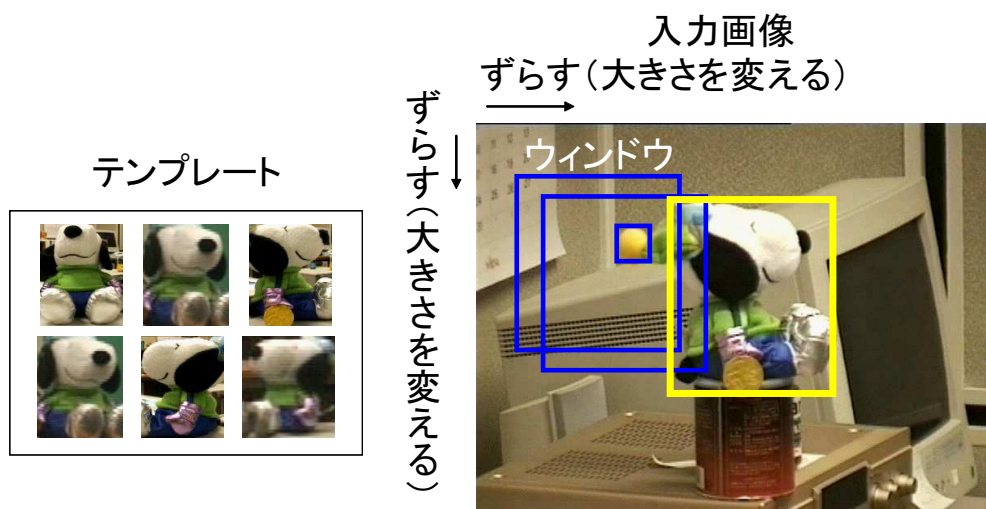


図 28 画像からの静止物体の検出における問題

### 3. 物体の位置・姿勢の変化に頑健な高速物体検出

#### 3.1 序言

2章では、照明の変化や隠蔽に対して頑健に移動物体領域を検出する手法について述べた。本章では、図28のように物体の位置・姿勢の変化に伴い物体の見え方が多様になった場合についてそれら全ての見え方を用意し、検出対象モデルに基づく手法により探索し、目的の物体を高速に検出することを目的とする。すなわち、本手法は、その物体が何であるかを識別し、物体の位置・姿勢を推定することもできる。

移動物体に対しては、2章で提案した手法のような背景モデルに基づく手法と併用し移動物体領域のみに対して本章の手法を適用し、更なる高速化を行うことも可能である。背景モデルに基づく手法との組み合わせを前提としていないため静止物体の検出も可能である。

室内から静止物体を検出する問題では、物体が3次元形状を持つ場合、カメラと物体の位置関係によって、見え方(物体の向きや大きさ)が異なるため、見え方に対応した多数のテンプレートを準備する必要がある。多数のテンプレートを用

いることによる計算量の増大を軽減するため、部分空間法を使ってテンプレートの次元数を圧縮する手法などが提案されている [MN94, TSOO00, MOKS00]. しかしながら、次元圧縮を行っても計算量が大きく、さらなる高速化が期待されている.

膨大な照合回数を大幅に削減する技術として、アクティブ探索法 [MV98, KSM99] が提案されている. アクティブ探索法は特徴照合の際にヒストグラムを用い、ヒストグラムの性質を用いて精度を落とさないことを保証しながら探索に不要な照合を削減する手法であり、これにより極めて高速な物体検出が可能となる. 更に、ヒストグラム照合を利用するため多少の形状変動に対しても頑健である. しかしながら、照明条件や物体の姿勢の変化により特徴が大きく変化する場合には、単一のテンプレートでは正しく探索できない.

まず、物体の多様な見え方に対応して作成した多数のテンプレートを高速に探索して室内環境から物体を検出する手法（動的アクティブ探索法）を提案する. 動的アクティブ探索法はヒストグラムの代数的性質を利用して未照合のテンプレートとウィンドウの組み合わせにおける類似度の上限値を使って照合の省略を行い探索の高速化を図る. さらに、画像上のウィンドウとテンプレートとの照合結果をカメラの制御回数の削減に利用したカメラパラメータの予測制御によって、カメラのパン・チルト・ズーム動作を含んだ探索時間を削減する. 本手法では、利用できる特徴が色ヒストグラム特徴に限定されているため、背景色が物体色に類似する場合には物体を見分けることができず、適用範囲が色が鮮やかな物体に限定されてしまう問題点がある.

そこで色ヒストグラム特徴に限定しない一般の距離特徴を用いた物体探索を行う場合の計算時間を大幅に削減する手法（適応的ウィンドウスキップ法）を提案する. 適応的ウィンドウスキップ法は、アクティブ探索 [MV98] で用いられている類似度上限値によるスキップの概念を一般的な距離によるテンプレートマッチングに適用し、距離下限値を用いた照合の枝刈りによって全探索と同一の精度を保証しながら、テンプレートマッチングを高速化している. この手法は、一般の距離を用いているため、距離が定義された特徴で識別できる任意のアルゴリズムを利用でき、適用範囲が広い.



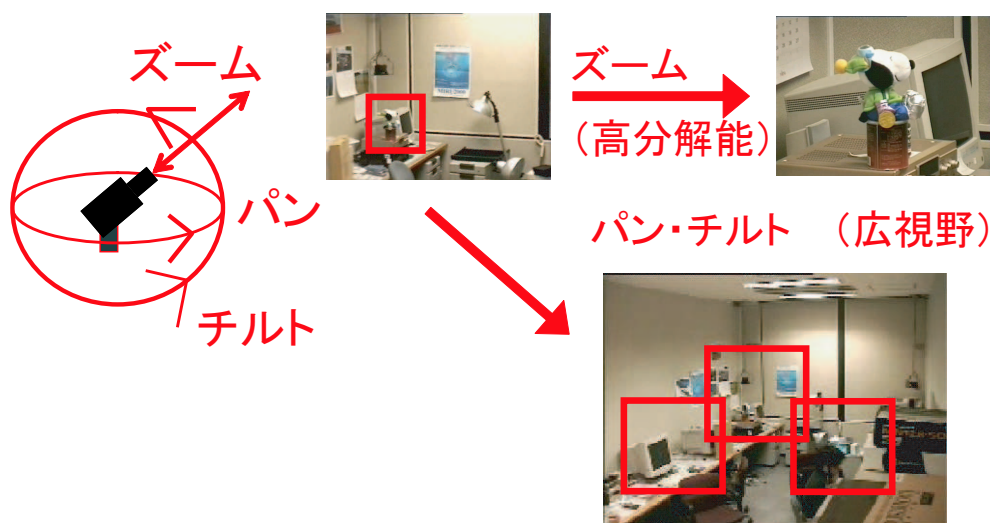


図 29 パン・チルト・ズームカメラを使った探索範囲の拡大

以下, 3.2 節において動的アクティブ探索法について述べ, 3.3 節において適応的ウィンドウスキップ法について述べる. 最後に 3.4 節で本章をまとめる.

### 3.2 色ヒストグラムを用いた物体検出の高速化

本節では, 室内環境において, 近くから, 遠くまで, 様々な位置にある物体を実時間で, 探索可能な手法について述べる. 広い視野を取得できるパン・チルト・ズームカメラを用いて室内環境から物体を検出する手法を提案する.

室内から静止物体を検出する問題では, 物体までの距離や方向が未知であるため, 大きさ・位置の異なる多数のウィンドウとの間で照合する必要がある上に, 物体の位置・姿勢の違いにより物体の見え方が大きく変化する. この変化を吸収するには, さまざまな条件で物体を撮影した複数のテンプレートを用意し照合する必要がある. 探索時間が指数関数的に増大する. 探索速度の改善のためにいくつかの手法 [MN94, TSOO00, MOKS00] が提案されているが, カメラのパン・チルト・ズームの制御までも考慮するような広大な探索空間 (図 29) に対しては, 照合コストが大きく十分な速度は得られていない.

また、照明の影響などにより特徴の変動がある特徴をアクティブ探索法で扱う場合、ベクトル量子化 (VQ) によるヒストグラム区画の構成が提案されている [KKM00a, KKM00b]。しかしこれらは VQ を探索の前処理として、全ての蓄積された入力情報に対してあらかじめベクトル量子化を行っている。本研究のようなリアルタイムで入力される大量の画像情報に対して行うことは想定されておらず、VQ に伴う距離計算に大きな計算量がかかり速度に問題がある。

また、ヒストグラム特徴には、物体領域が小さくなるほど識別性能が低下する性質がある [SC96] ため、ズームを細かく制御する必要がある。このため、数多くのカメラの制御回数が必要になる。

本節で提案する手法 [KM01] は、従来のアクティブ探索法 [MV98] の考え方を以下の 2 点において発展させたものである。

第 1 に、本手法ではウィンドウと複数のテンプレートとの照合を行うが、その際に、あるテンプレートの照合結果を利用して他のテンプレートの照合を適応的に削減する。ここでの照合の削減は、従来のアクティブ探索法と同様に原理的にヒストグラムによる総当り照合と同じ解を保証している。照合が省略可能かどうかは、照合結果の類似値とテンプレート間の類似値で決定され、テンプレート間で類似している場合は大きく削減し、類似していない場合にはあまり削減しないといった制御が自動的かつ厳密に行える。このため、精度を保つために対象に応じてパラメータを調整する必要はない。第 2 に、画像上のウィンドウとテンプレートとの照合結果を、カメラの制御回数の削減に利用する。照合には適さない小さいテンプレート（予測用テンプレート）に関して、提案する高速探索法を利用して物体の存在の可能性を予測し、その予測値を基に優先的にズームインするパン・チルトの方向を決定することでパン・チルト・ズームカメラの制御回数を削減する (図 30)。ズームして拡大した画像を得られた段階で、十分信頼できる大きなテンプレート（検証用テンプレート）を用いて物体の検出を試みる。

本節では、3.2.1 節で提案手法について述べ、3.2.2 節で提案手法を複数のパン・チルト・ズームカメラに適用し、位置を推定する方法について述べる。3.2.3 節で探索精度と探索速度の評価実験とその結果を、3.2.4 節で提案手法について考察する。

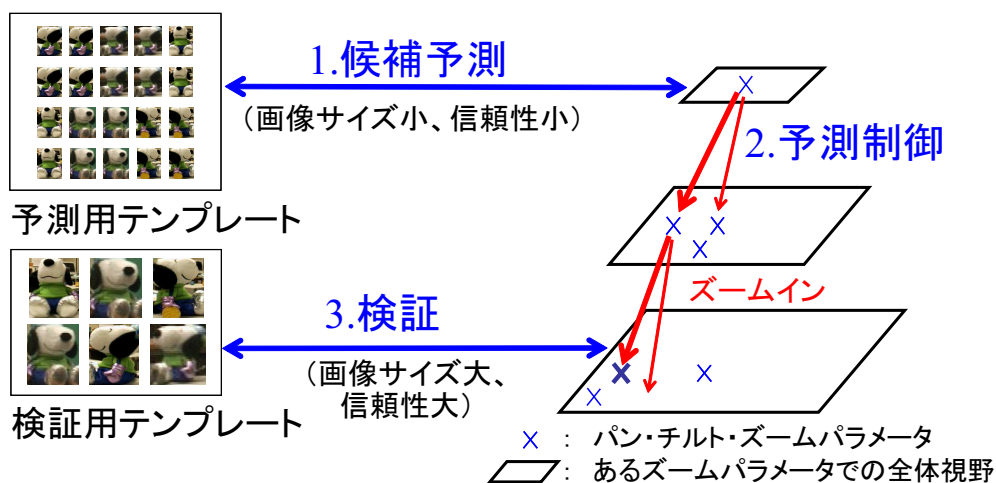


図 30 予測制御によるカメラ制御回数の削減

### 3.2.1 動的アクティブ探索法

本手法は、画像から特定の物体を高速に探索するアクティブ探索法 [MV98] を 3次元環境中の物体探索に適用できるように拡張したものである [KM00, KMTW01a, KM01]。具体的には室内環境などにパン・チルト・ズームカメラを設置し、パン・チルト・ズームのカメラパラメータを動的に制御しながら、あらかじめ登録された物体と類似した物体を高速に探し出す。

#### 3.2.1.1 手法の概要

図 31 に動的アクティブ探索法の処理の概要を示す。テンプレートデータベース (図 31 左) とアクティブ探索部 (図 31 中央下段) とカメラ制御部 (図 31 中央上段) からなる。以下、動的アクティブ探索法の処理の概要を説明する。まず準備としての学習段階では、様々な条件で撮影した物体の画像を用意し、テンプレートデータベースに蓄積する。次に、データベースに登録されたテンプレートからアクティブ探索に適したヒストグラム区画を生成する。更に、カメラ制御のために、3次元環境全体を見渡すパン・チルト・ズームのカメラパラメータ集合を求める。

探索段階では、まずパン・チルト・ズームカメラ制御部はカメラパラメータ集

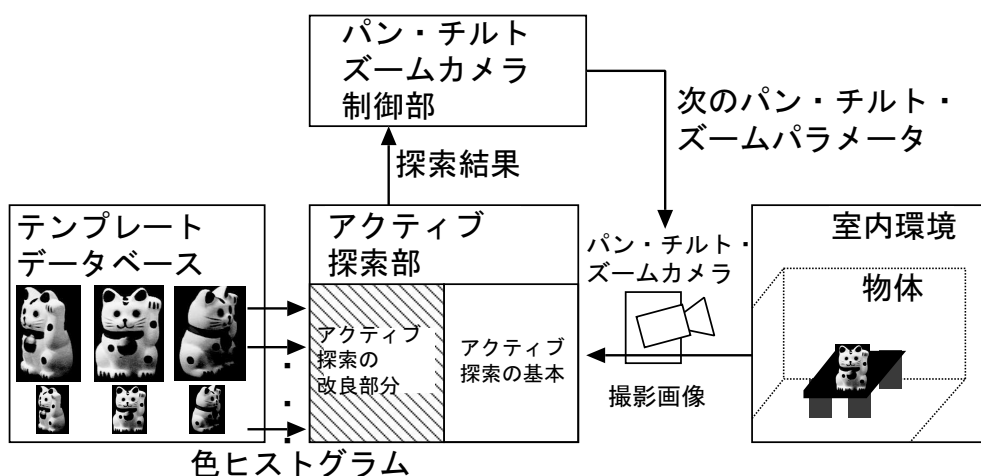


図 31 動的アクティブ探索法の処理の概要

合の最初のパン・チルト・ズームパラメータをカメラに通知する。カメラは受け取ったパン・チルト・ズームパラメータを用いて画像を撮影し、その画像をアクティブ探索部へ送る。アクティブ探索部では物体を画像中から高速に探索し、閾値を越えた領域を探索結果としてカメラ制御部へ通知する。カメラ制御部では物体を検出した場合には物体をユーザに表示して探索を終了し、検出しなかった場合には探索結果に基づいてカメラを制御する順序を動的に切替え最も優先度の高いパン・チルト・ズームパラメータで撮影する。

### 3.2.1.2 複数テンプレートの高速探索法

動的アクティブ探索法では、照明などの環境の変動に頑健にするために、探索処理自体も従来のアクティブ探索法に改良を行っている。具体的には、テンプレートの色分布を前もって学習し、アクティブ探索を効率化するヒストグラム区画を生成する点と、探索段階において複数のテンプレート間の関係を利用して照合の削減を行う点の改良である。

**効率的なヒストグラム区画の設計** 従来のアクティブ探索では, RGB の各軸を等間隔に分割することによって減色した各色の頻度を数えヒストグラムを生成した. しかし, このような減色は照明の変動や, 物体の姿勢の変化などに対して頑健でない. あらかじめ室内環境中の様々な位置, 姿勢で撮影された物体の画像が撮影できる場合にはその物体の色を学習サンプルとしてベクトル量子化 (VQ) することによって効率のよい減色を行うことができる. 提案手法では, VQ 処理による距離計算時間を省略するために符号化テーブルを用いる. 符号化テーブルは RGB の各軸を 32 分割 ( $32^3$  区画) した各色の小区画がどの VQ 符号に属するかを示すもので, 定数時間で色と VQ 符号とを対応づける. またテンプレートに含まれる色に対して VQ を行えるため, テンプレートに含まれない色を表す特別な符号 0 を導入することによって, テンプレートに含まれる色に効率良く符号を割り当てることができる. この方法では符号数が少なくなるため, ヒストグラム照合演算が高速になる. 更に, 物体の写らない領域を求める処理を高速に行うことができる. 符号化テーブルの作成手順の詳細を以下に述べる.

1. 全体の画素の平均色, 分散を求める. この平均色を代表色  $C_1$  とし, 全画素を  $C_1$  に帰属させる.
2. 分散最大の代表色  $C_i$  を選ぶ.  $C_i$  の分散  $\alpha_i < \sigma$  であれば, ステップ 4 へ飛ぶ.
3.  $C_i$  を 2 つに分割し, LBG アルゴリズム [NK88, HH93, McL93] により最適な代表色を決定する. 全画素と全代表色の距離を計算し各画素を最も近い代表色に帰属させる. 代表色に帰属された画素の色の平均色と分散  $\alpha$  を計算する. ステップ 2 へ戻る.
4. 色空間の全色  $P_i$  に対して最も距離の近い代表色  $C_j$  を求める.  $|P_i - C_j| < n\alpha_j$  であれば  $P_i$  に  $C_j$  の符号を割り当てる. さもない場合は,  $P_i$  に符号 0 を割り当てる.

$n$  は学習に用いた画素のうち符号 0 に含まれる画素の割合が一定以下になるように経験的に 2 程度に決めた.  $n$  が小さ過ぎる場合には, 符号 0 の割合が増加し, 多

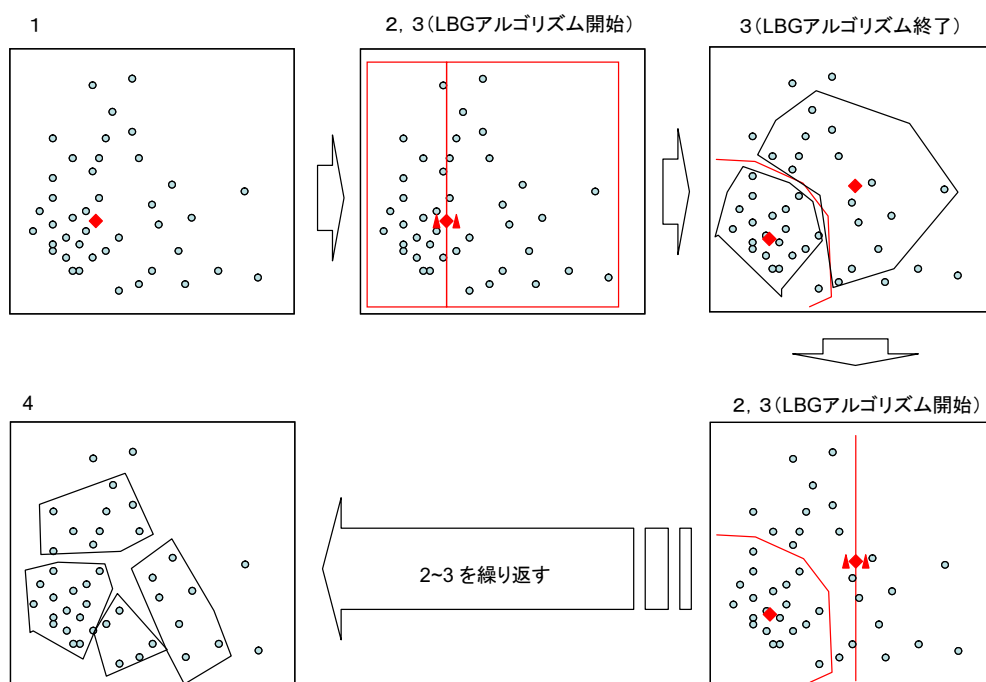


図 32 ヒストグラム区画の作成の例

過ぎる場合には 0 以外の符号に物体の色に類似しない色が含まれる割合が増加する。どちらの場合も誤検出が増加する。一方、 $\sigma$  は、小さい場合には照明や向きの変動に弱くなり、大きい場合には誤検出が多くなる。このため物体の色分布に応じて決定する。

図 32 に提案手法によるヒストグラム区画の作成の例を示す。

まず左上の○のような分布が与えられたときにヒストグラム区画を作成する。まず 1. のように全体  $C_1$  の画素の重心位置  $\diamond$  と分散  $\alpha_1$  を求める。分散が目標値  $\sigma$  に達しない場合、重心を重心付近の異なる 2 点に分割する。この 2 点を集合の基準点とする。分割した基準点のどちらか近い方に全ての点を属させる。作成した 2 集合の重心に基準点を更新する。基準点と重心が一致するまで、基準点の更新を繰り返す (LBG アルゴリズム)。基準点が収束すると、再度、各基準点の集合の分散と目標値  $\sigma$  を比較する。目標値  $\sigma$  に達しない集合があった場合、一番分散の大きい集合の基準点を分割し、同様の処理を行う。

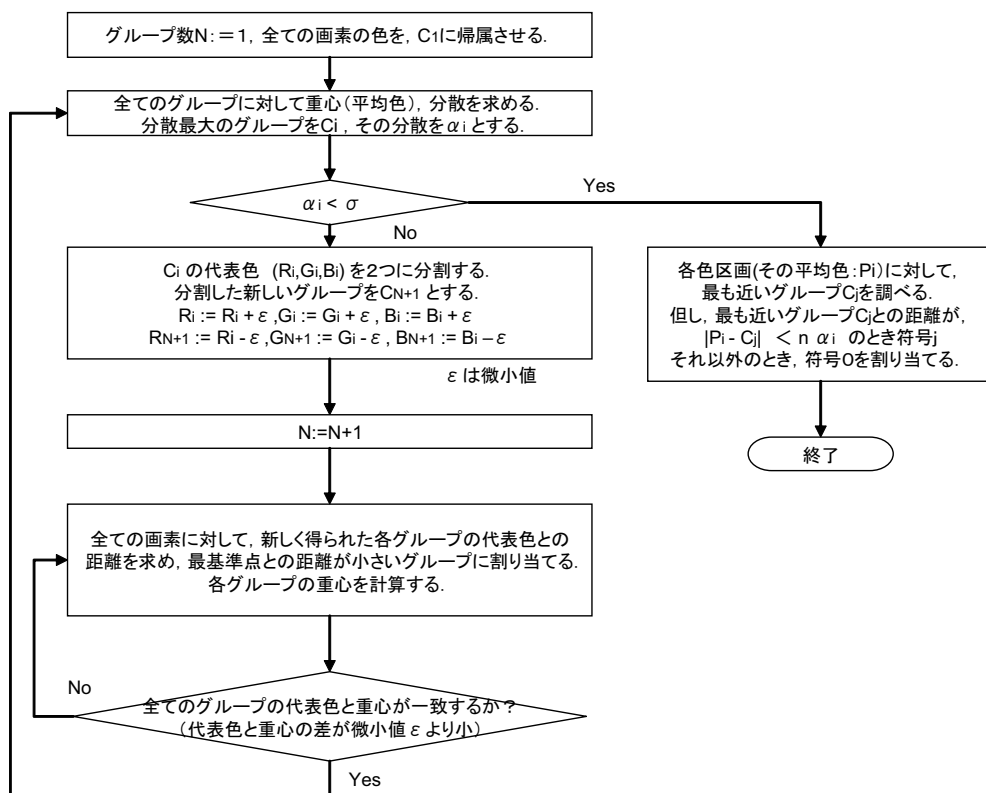


図 33 ヒストグラム区画の作成のフローチャット

上記処理を分散の最大が目標値  $\sigma$  を下回るまで繰り返すことによって, 適切なヒストグラム区画を得る. ヒストグラム区画の作成のフローチャートを図 33 に示す.

**アクティブ探索法の基本** アクティブ探索法 [MV98] は, 入力画像上での位置・大きさが未知の物体を検出できるように, テンプレートと, 入力画像のウィンドウとの照合を行う手法であり, 色ヒストグラム特徴の性質を利用したウィンドウの枝刈りにより精度を保証したまま照合回数を大幅に削減するという特徴がある.

アクティブ探索は任意の形状のテンプレートに適用できるが, 簡単のため, 矩形のテンプレートで基本アルゴリズム<sup>1</sup> を説明する. まず, 入力画像を符号化した

<sup>1</sup> ただしウィンドウの大きさによってヒストグラムの分布に変化が生じる場合を扱うので [MV98]

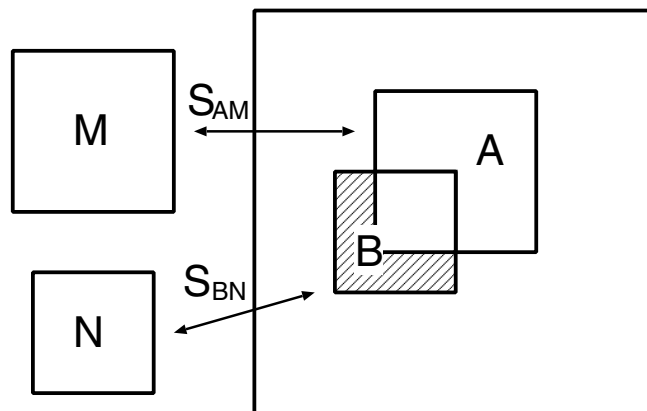


図 34 重なりを利用した照合のスキップ

画像に対し、テンプレートと同じ大きさのウィンドウをかけ、ヒストグラム特徴を作る。続いて、ヒストグラム同士の類似度を計算する。類似度にはヒストグラムの重なり率を用いる。ヒストグラム特徴  $H_M$  と  $H_A$  の類似度  $S_{AM}$  は、次のように定義される。

$$S_{AM} = \frac{1}{|M|} \sum_{i=1}^I \min(H_{Mi}, H_{Ai}) \quad (18)$$

ここで、 $H_M$  と  $H_A$  は、それぞれテンプレート  $M$  と入力画像中のウィンドウ  $A$  に対するヒストグラムであり、 $H_{Mi}$ 、 $H_{Ai}$  はそれぞれの  $i$  番目の符号を持つ画素数である。また  $|M|$  はテンプレートの画素数 (テンプレートの特徴の総数) であり、 $I$  は符号の種類である。 $S_{AM}$  が探索閾値  $\theta$  以上となる時、目的物体として検出する。

ウィンドウの枝刈り法を図 34 を用いて説明する。大きさの異なる 2 つのテンプレートを  $M, N$  とする ( $|M| > |N|$ )。  $M$  と  $N$  の色の変化はない、すなわち、 $N$  のヒストグラムは  $M$  のヒストグラムの定数倍であり、 $N$  の色は全て  $M$  に含まれるとする (異なる色が含まれる場合への一般化は後述する)。ここで、 $|A| = |M|$ 、 $|B| = |N|$  を満たし重複領域を持つウィンドウ  $A, B$  を考える。テンプレート  $M$  とウィンド

---

と異なりヒストグラムを大きさを正規化しない。このため類似度は可換でない。 $S_{AM} \neq S_{MA}$ 、 $S_{AM} \cdot |M| = S_{MA} \cdot |A|$



ウ  $A$  との類似度  $S_{AM}$  と、テンプレート  $N$  とウィンドウ  $B$  との類似度  $S_{BN}$  との間には (19) 式の不等式が成立する.

$$|N| \cdot S_{BN} < |M| \cdot S_{AM} + n \quad (19)$$

ここで、 $|M| \cdot S_{AM}$  は  $A$  と  $M$  との間で色の同じ画素の組が何組あるかを意味する. これを同色画素数と呼ぶことにする. この式は「物体と  $B$  間の同色画素数 ( $|N| \cdot S_{BN}$ ) は物体と  $A$  間の同色画素数 ( $|M| \cdot S_{AM}$ ) が  $A$  と  $B$  の共通領域に局在し、かつ、 $A$  に含まれない  $B$  の領域 (図 34 の斜線部分で画素数  $n$  とする) の画素すべてが物体と  $B$  の同色画素になった場合より多くなることはない」ことを意味する. すなわち、 $S_{AM}$  が既知となったとき、 $|N| \cdot S_{BN}$  の類似度上限値は (19) 式の右辺になる. この類似度上限値が  $|N| \cdot \theta$  より小さければ  $N$  と  $B$  との照合を省略することができ、探索の高速化が図れる. 照合を省略できる  $n$  の範囲は次の式で与えられる.

$$|M| \cdot S_{AM} + n < |N| \cdot \theta \quad \therefore n < |N| \cdot \theta - |M| \cdot S_{AM} \quad (20)$$

**複数テンプレートを考慮したアクティブ探索** 学習段階では、物体の姿勢の変化や照明の変化に対応するため、複数の位置・姿勢で物体を撮影し、探索段階においては、撮影した複数の画像をテンプレートに用いて高速に物体を探索する.

処理の流れを図 35 に示す. はじめに、それぞれのテンプレートに対して各画素の色を符号化テーブルに基づき符号化し各符号に属する画素の出現頻度を計算し、ヒストグラム特徴を作成する. 次にテンプレートに対する処理と同様に入力画像の各画素を符号化テーブルに基づいて符号化する. この画像上でヒストグラムが類似するウィンドウを探索する原理は基本的にはアクティブ探索法と同じである. すなわち、あるウィンドウに対して、符号を用いてヒストグラムを作成し、探索不要な領域を計算しながら探索を行う.

検証段階では、複数のテンプレートを同時に効率よく探索するために、予め計算したテンプレート間のヒストグラム間の差から、他のテンプレートとの類似度の類似度上限値を求め、他のテンプレートにおける照合をも省略する. これを並列探索と呼ぶ.

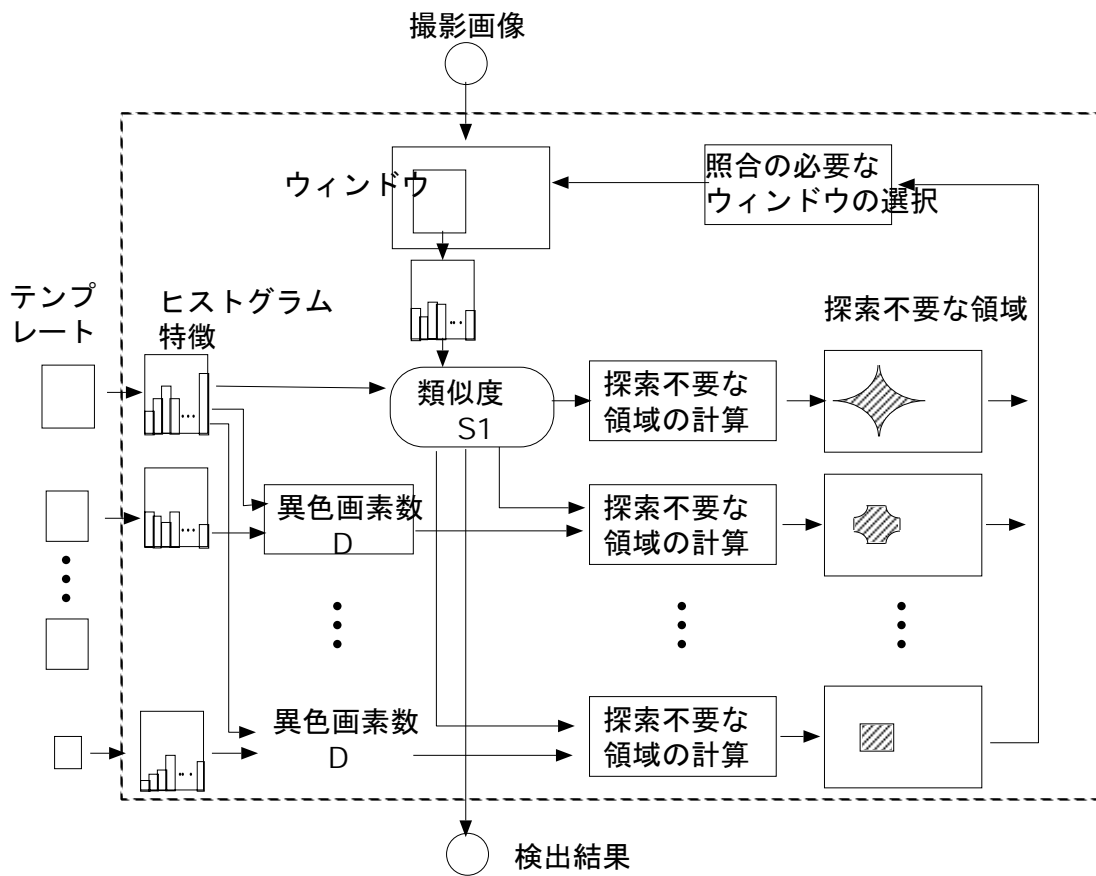


図 35 複数テンプレートを用いた物体探索の処理の流れ

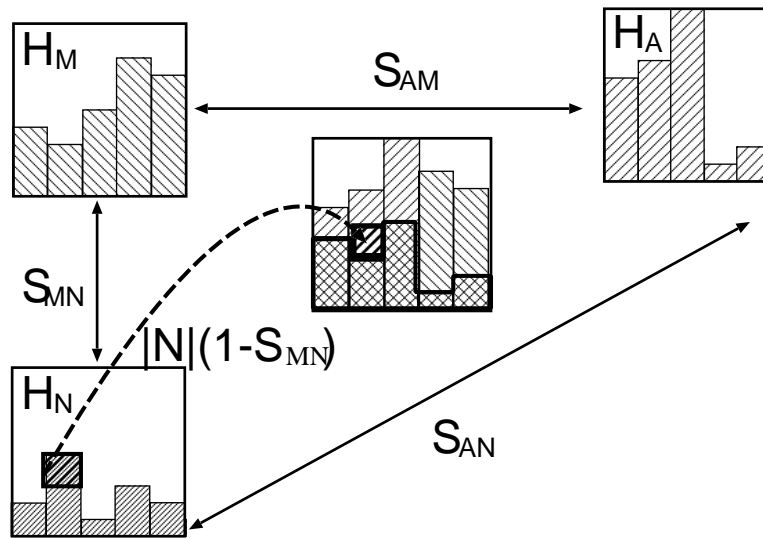


図 36 並列探索の仕組み

予測段階では取りこぼしのないことを保障して極めて高速に探索するために多数のヒストグラムを併合したヒストグラムを用いて類似度上限値を決定する手法も導入する. これを併合探索と呼ぶ. 以下, これらの手法について説明する.

**複数テンプレートの並列探索** テンプレート  $N$  に テンプレート  $M$  と色の異なる画素が含まれる場合を扱えるよう (19) 式, (20) 式を拡張する.  $N$  の画素のうち,  $M$  と異色な画素 ( $N$  の  $M$  に対する異色画素と呼ぶ) の数は  $|N|(1 - S_{MN})$  で表される (図 36).  $S_{AM}$  からは,  $N$  の  $M$  に対する異色画素が  $A$  中に存在するかの情報は得られないため,  $N$  の  $M$  に対する異色画素が  $A$  中に現れていると仮定して,  $|N| \cdot S_{BN}$  の類似度上限値を推定し,  $|N| \cdot \theta$  に達しない領域を求める必要がある. すなわち, 以下の不等式が導かれる.

$$\begin{aligned}
 |N| \cdot S_{BN} &< |M| \cdot S_{AM} + |N|(1 - S_{MN}) + n \\
 n &< |N| \cdot (\theta - 1 + S_{MN}) - |M| \cdot S_{AM}
 \end{aligned}
 \tag{21}$$

並列探索のフローチャートを図 37 に示す.

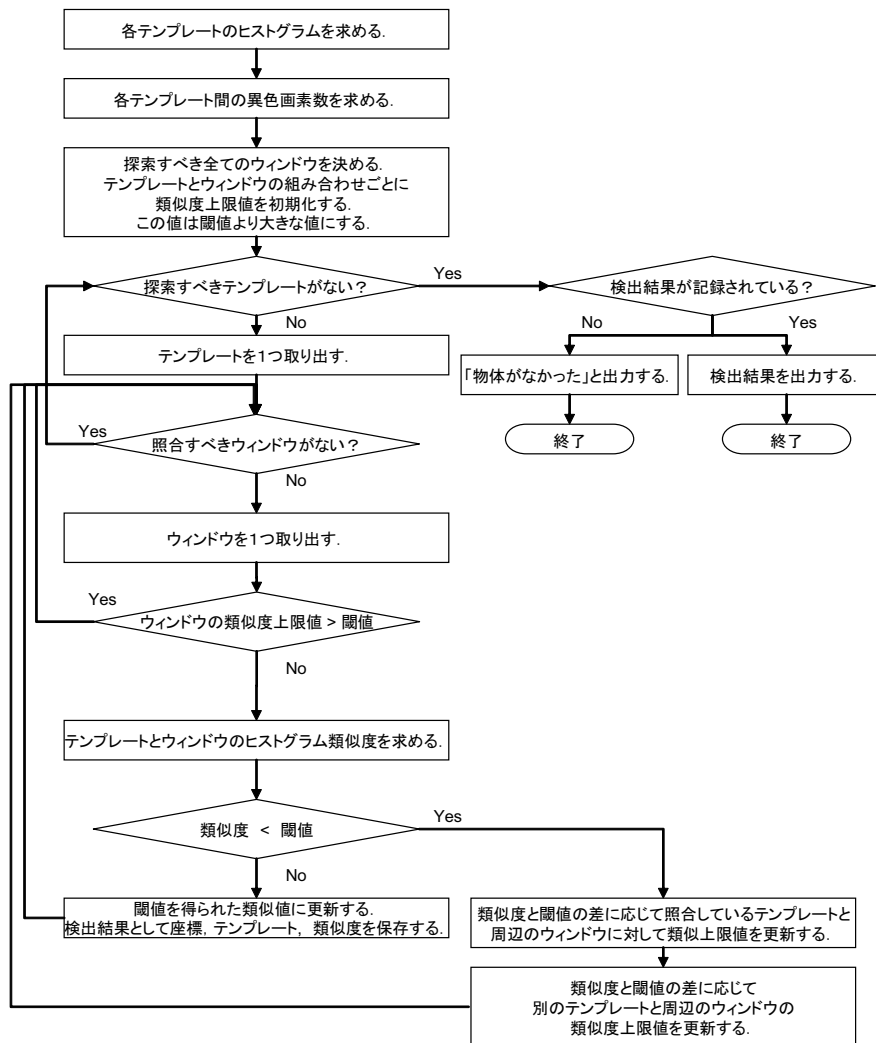


図 37 並列探索のフローチャート

**複数テンプレートの併合探索** 探索するテンプレートの間でヒストグラムの違いが大きい場合には, (21) 式に含まれる加算部分  $|N|(1 - S_{MN}) \cdot |M|$  が大きくなり照合を省略する効果が少ない. そこで, 併合探索では並列探索のように類似度にヒストグラムの違いを反映させるのではなくヒストグラム自身に反映させることによって不要な照合を省略させる. テンプレートの全てのヒストグラムを  $H_{M^0}, H_{M^1}, \dots, H_{M^k}$  とする. この時, 併合ヒストグラム  $H_U$  を

$$H_{U_i} = \max \left( H_{M_i^0}, H_{M_i^1}, \dots, H_{M_i^k} \right) \quad (22)$$

と定義する.  $H_{U_i}$  は  $U$  の各符号の数である.  $H_{U_i}$  の各要素は, 各テンプレートの最大値から構成されるため, ある探索窓  $A$  と  $U$  との類似度は  $A$  と  $U$  に属する  $M^0, M^1, \dots, M^k$  の類似度よりも必ず大きい. 従って  $|U| \cdot S_{AU}$  は  $|M| \cdot S_{AM}^j (\forall j \in 1, \dots, k)$  の類似度上限値となる.  $U$  を探索することによって  $M^0, M^1, \dots, M^k$  の照合を省略できる. 但し, 類似度自体を求める必要がある場合には,  $U$  により検出した領域に関して, 更に  $M^0, M^1, \dots, M^k$  を用いて探索する必要がある.

併合探索のフローチャートを図 38 に示す.

### 3.2.1.3 高速探索法のカメラ制御への適用

本節では, 3.2.1.2 節で提案した高速探索の結果を利用して動的なパン・チルト・ズームパラメータの予測制御を行う手法を説明する. この結果, カメラのパン・チルト・ズームの制御回数を削減することができる.

ズームの総段数を  $d$ , ズーム倍率を  $\delta$ , 各ズームの段階を  $i$  とすれば, パン・チルト・ズームで行って得られる総探索空間は, 以下の式で表せる.

$$\text{総探索空間} = S \sum_{i=0}^d \delta^{2i} \quad (23)$$

ここで,  $S$  はズーム倍率 1 で撮影した 1 枚の撮影画像を探索するための探索空間である. このとき倍率  $\delta$  を小さくすれば, 段数  $d$  が非常に大きくなる. このため, 物体検出には十分な精度が保証できる範囲で総探索空間を最小にする  $\delta$  を選ぶ. 例えば実験で用いた物体の場合には  $\delta = 2 \sim 3$  となる. また, 探索精度はテンプレートが大きいほど良くなり, 例えば一辺の画素数が 30 以上の物体の場合には十

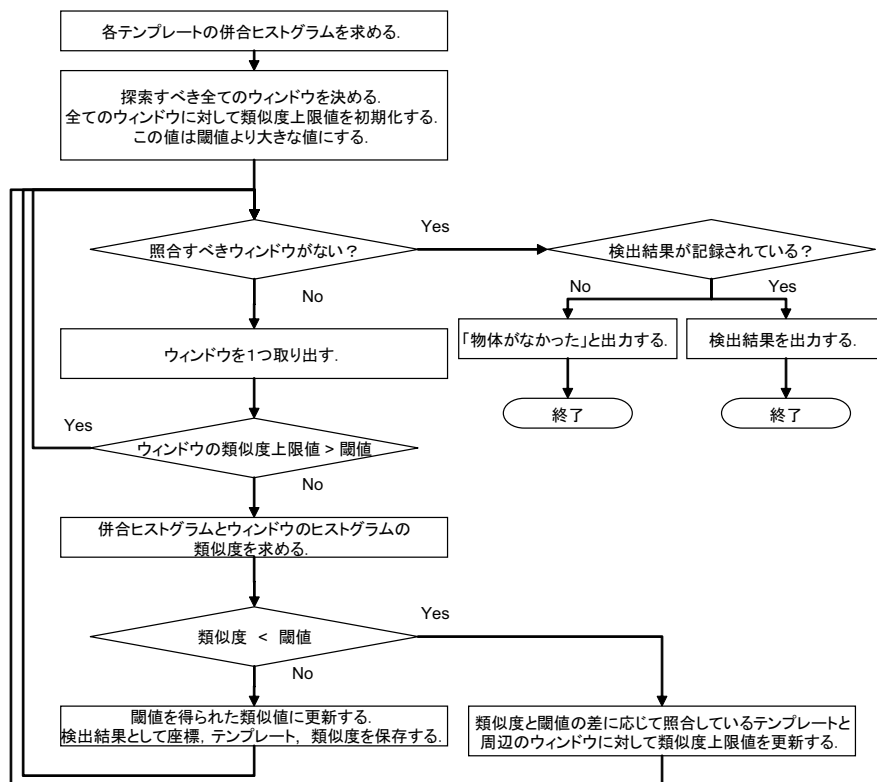


図 38 併合探索のフローチャート

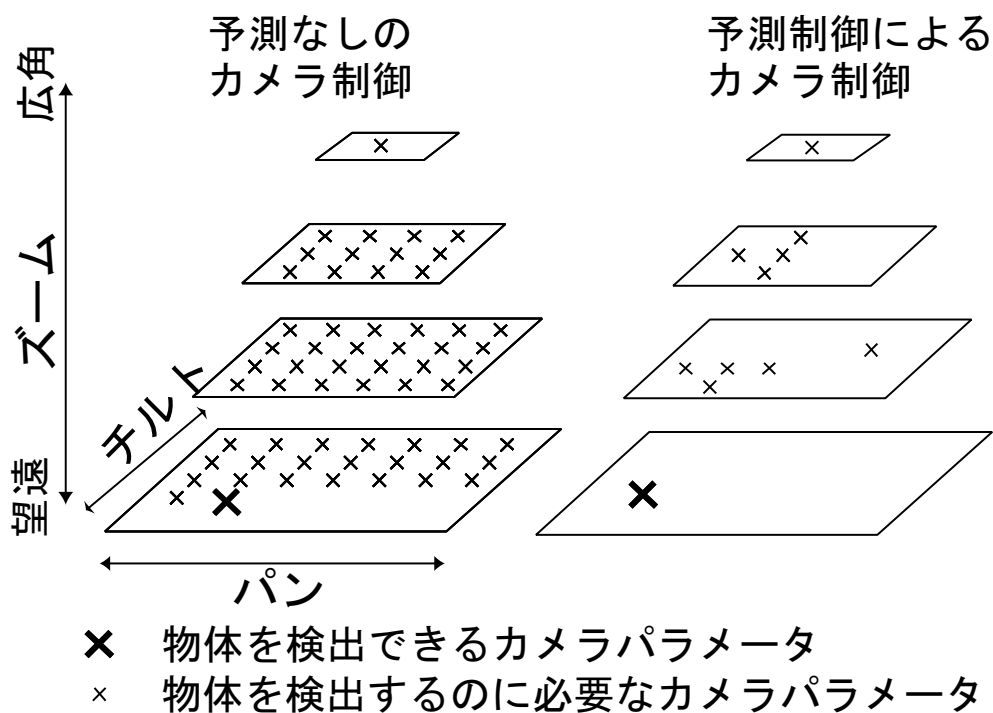


図 39 カメラ制御回数の削減の必要性

分な精度があることがわかっている。つまり、ここでは一辺の画素数が 30 以上の物体を探索し、検出した場合には、探索を終了し、検出できなかった場合にはもう一段階ズームインし入力画像をより拡大する。更にパン・チルトを順に動かして視野内をもれなく探索している。これにより大きく写る近い物体から小さく写る遠い物体までを探索することができる（図 39 左）。しかしながら、ズームインを行うと遠い物体が大きく見える反面、一度に撮影できる視野が狭くなるため数多くのパン・チルト・ズーム回数を必要とし、膨大な探索回数になるため、単純な順序で探索する手法では計算時間が膨大になる。そこであらかじめ物体がありそうな領域を優先してカメラを向ける。つまり、小さいテンプレートを探索した結果を用いて、探索する順序を物体がありそうな領域が最初になるように動的に切替えること（予測制御）によって物体を検出するまでの探索回数を小さくする（図 39 右）。

以下詳細な予測制御のアルゴリズムを述べる。なお、以下のアルゴリズム中、一辺 30 画素以上のテンプレートのサイズを「検出サイズ」と呼ぶ。一方、一辺 10 ~ 30 画素の領域はズームしないと検出物体として確定できないため、このサイズを「候補サイズ」と呼ぶことにする。

#### 初期化

1. 最適な総探索空間を実現するパン・チルト・ズーム位置をズーム倍率の昇順に並べる。これをカメラ位置リストと呼ぶ。
2. 手続き：探索を実行する。

#### 手続き：探索

1. カメラ位置リストの先頭の値を、パン・チルト・ズームのパラメータにセットする。カメラ位置リストの先頭の値を削除する。画像を取得する。複数テンプレートを用いたアクティブ探索を実行する。
2. 検出サイズで検出した場合、確かに物体を検出したことを意味するので、探索を終了する。
3. 候補サイズで検出した場合、検出した領域を含むパン・チルト・ズームパラメータをカメラ位置リストの先頭に並べ換える。複数検出した場合は、候補サイズの降順に並べる。1. へ戻る。
4. 物体を検出しない場合、カメラ位置リストが空ならば 5. へ。そうでなければ、1. へ戻る。
5. 物体が存在しなかったとして、終了。

カメラ制御のフローチャートを図 40 に示す。



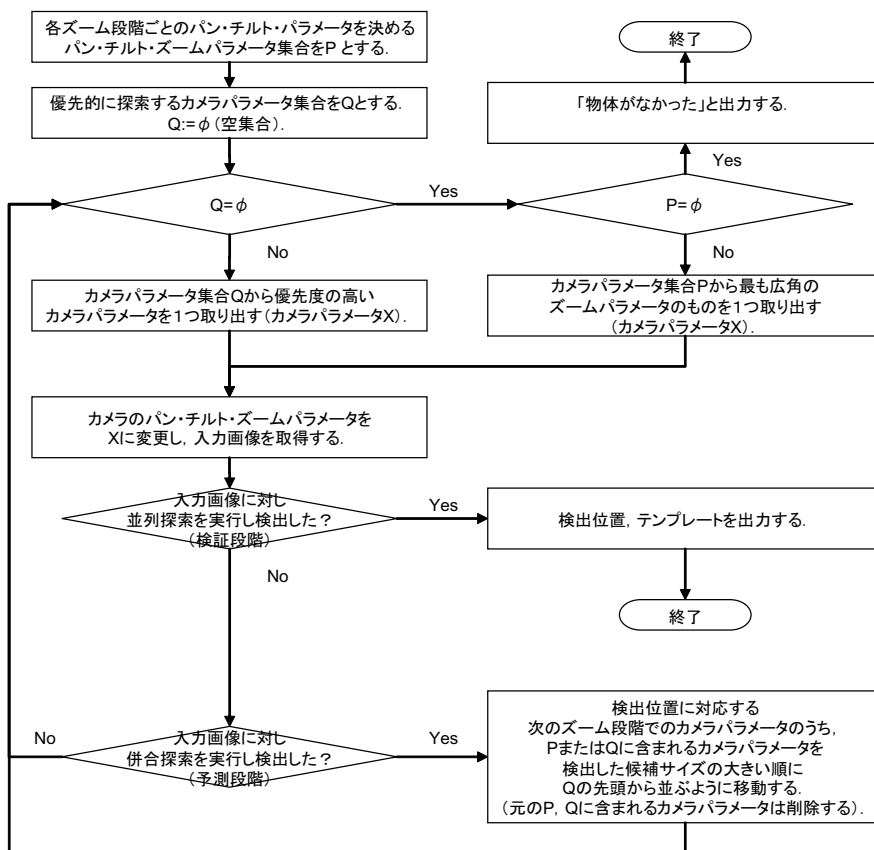


図 40 カメラ制御のフローチャート

### 3.2.2 複数のパン・チルト・ズームカメラによる物体の位置推定

3.2.1節で提案した動的アクティブ探索法を物体の位置推定に用いる方法を述べる。ある1つのパン・チルト・ズームカメラで物体を検出した場合、そのパン・チルト・ズームカメラで物体のおおよその位置を推定する。次に推定した位置に他のパン・チルト・ズームカメラを誘導する。複数のパン・チルト・ズームカメラで同一物体を検出し検証することにより、位置推定の精度を増すことが可能になる。

動的アクティブ探索法は、物体の色ヒストグラム特徴を用いるため、物体検出の位置推定の精度が低い。複数のパン・チルト・ズームカメラを用いて三角測量と自乗誤差最小の原理 [Hor93] を用いることによって、位置推定精度の大きな向上が望める。

以下、一台のパン・チルト・ズームカメラによる位置推定法を 3.2.2.1 節で述べ、複数台のパン・チルト・ズームカメラによる位置推定法について 3.2.2.2 節で述べる。

#### 3.2.2.1 一台のパン・チルト・ズームカメラによる物体の位置推定

3次元位置情報が既知のパン・チルト・ズームカメラ上で動的アクティブ探索法によって得られた物体の方向と大きさから、おおよその物体の3次元位置を求めることができる [KMTW01b, KMT02]。

物体が検出されたとき、画像上の物体の座標  $\mathbf{T}^i(t_x^i, t_y^i)$  および大きさ  $\mathbf{W}^i(w_x^i, w_y^i)$  が得られる (図 41)。ただし  $i$  はパン・チルト・ズームカメラ番号である。

物体を画像の中心になるようにカメラのパン・チルトを移動させると、カメラ座標系 (カメラの光軸を中心としたパン角, チルト角による球面座標系) における物体の方向は以下のようなになる (図 42)。

$$\begin{aligned}\theta_x^i &= \arctan \left( \left( \frac{r_x^i - 2t_x^i}{r_x^i} \right) \tan \frac{f_x^i}{2} \right) \\ \theta_y^i &= \arctan \left( \left( \frac{r_y^i - 2t_y^i}{r_y^i} \right) \tan \frac{f_y^i}{2} \right)\end{aligned}\quad (24)$$

ただし、視野  $\mathbf{F}(f_x^i, f_y^i)$ , 入力画像サイズ  $\mathbf{R}^i(r_x^i, r_y^i)$  とする。

パン・チルト・ズームカメラの3次元上の外部パラメータ (位置  $\mathbf{P}^i(p_x^i, p_y^i, p_z^i)$ , 初期方向  $\mathbf{D}^i(d_x^i, d_y^i, d_z^i)$ ), 初期方向からのパン・チルト角  $\mathbf{C}^i(c_x^i, c_y^i)$ , 物体の大きさ

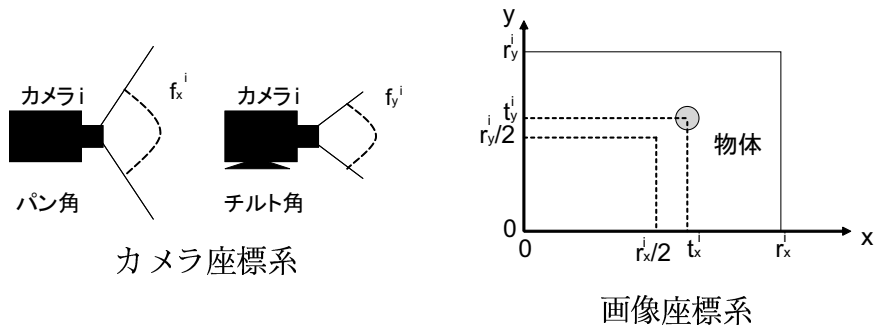


図 41 カメラ座標系と画像座標系

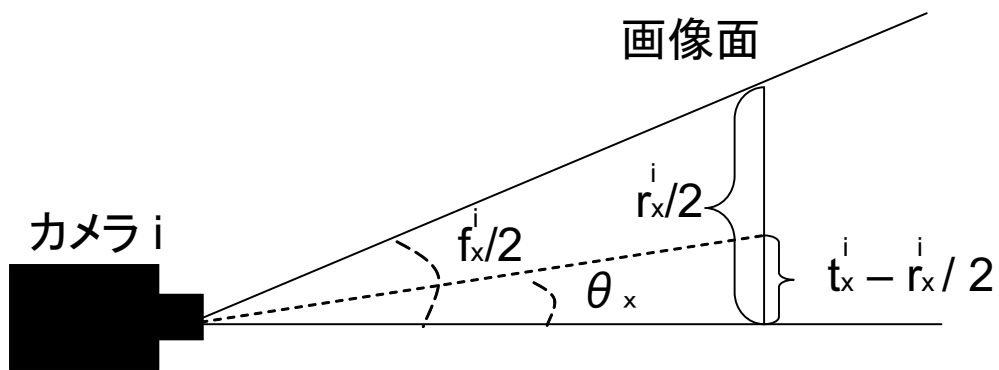


図 42 物体検出結果からの物体方向の導出

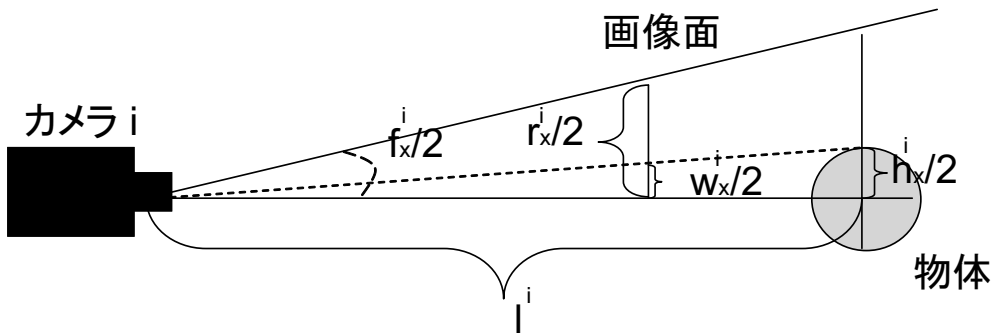


図 43 一台のカメラでの物体距離の導出

$H(h_x, h_y)$  が既知のとき, 物体までの距離  $l^i$  は以下のように計算できる (図 43) .

$$l^i = \frac{h_x \cdot r_x^i}{2w_x^i \cdot \tan \frac{f_x^i}{2}} = \frac{h_y \cdot r_y^i}{2w_y^i \cdot \tan \frac{f_y^i}{2}} \quad (25)$$

このとき物体の座標  $\mathbf{Q}(q_x, q_y, q_z)$  は,

$$\mathbf{Q} = \mathbf{P}^i + l^i \cdot Z^i(\alpha^i) \cdot Y(\beta^i) D^i \quad (26)$$

として得られる. ここで  $Z$  はパン方向の回転,  $Y$  はチルト方向の回転, ( $\alpha^i = c_x^i + \theta_x^i, \beta^i = c_y^i + \theta_y^i$ ) である.

### 3.2.2.2 複数台のパン・チルト・ズームカメラによる物体の位置推定

各パン・チルト・ズームカメラからみた物体の座標は, 物体までの距離  $k^i$  を未知数として, (26) 式より

$$\mathbf{Q} = \mathbf{P}^i + k^i \cdot Z(\alpha^i) Y(\beta^i) D^i \quad (27)$$

として求めることができる. すなわち, (27) 式から  $k^i$  を消去して複数のパン・チルト・ズームカメラについて解けば物体の位置が求まる. 2 つ以上のパン・チルト・ズームカメラで検出した場合には, (27) 式に対して自乗誤差が最小となる解を物体の位置とする. これは特異値分解 [WTVF93] により計算可能である.



図 44 実験に使用したパン・チルト・ズームカメラ

### 3.2.3 実験

#### 3.2.3.1 1台のパン・チルト・ズームカメラを用いた場合の従来手法との比較

3.2節で提案した探索法を用いて、室内環境から小物体を探索する実験を行い、探索速度と探索精度を評価した。

**実験条件** 実験に用いたワークステーション、カメラの仕様を表1に示す。探索対象には、図45に示すような3種類の物体を選んだ。物体の大きさは、高さ約5cmから10cm程度である。

まず、学習段階において、対象とする物体を各100枚程度撮影し、テンプレート

表1 動的アクティブ探索の実験に用いた計算機およびカメラの仕様

計算機	SGI社 O <sub>2</sub>
CPU	R10000(250MHz)
カメラ	Sony社 EVI-D30
解像度	320 × 240

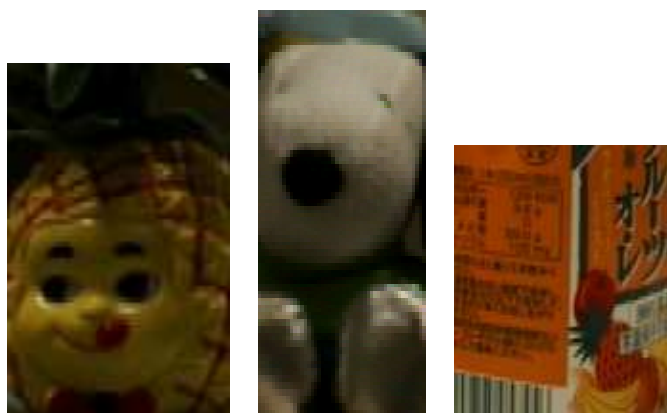


図 45 1 台のパン・チルト・ズームカメラによる物体検出実験で用いたテンプレートの例

とした。具体的には、照明条件を変えるために図 46 の A ~ E の 5 箇所に配置し、物体の正面、左右の計 3 方向を、1.2 倍ずつカメラズームを変え撮影した。

探索段階では、物体をカメラから 5m 以内の近くから遠くまで、図 46 の a ~ e の 5 箇所に配置した。更に、物体の姿勢を左右 2 種類の向きに変化させた。このとき物体の輪郭形状が異なる場合がある。この場合にもアクティブ探索法は適用可能であるが、より高速に探索を行うため、輪郭形状が相似な物体内部の矩形をテンプレートに利用した。また、ヒストグラム区画生成に使用した  $\sigma$  は、色の変動が大きいものに対しては、大きい  $\sigma$  を、色の変動が小さいものに対しては小さい  $\sigma$  を物体毎に選択した ( $\sigma = 1.5 \sim 3$ )。ここで  $\sigma$  は計算速度にはあまり影響を与えなかった。その理由は符号数の減少による計算速度の向上の効果と、背景類似度の上昇による枝刈り効率の低下の効果が相殺しているためと推察される。閾値は、予備実験により物体毎に設定した。

まず、カメラのパン・チルト・ズーム制御を行わずに、提案した画像探索手法だけについて、精度・速度を評価する。次にカメラのパン・チルト制御を加えた評価を行う。

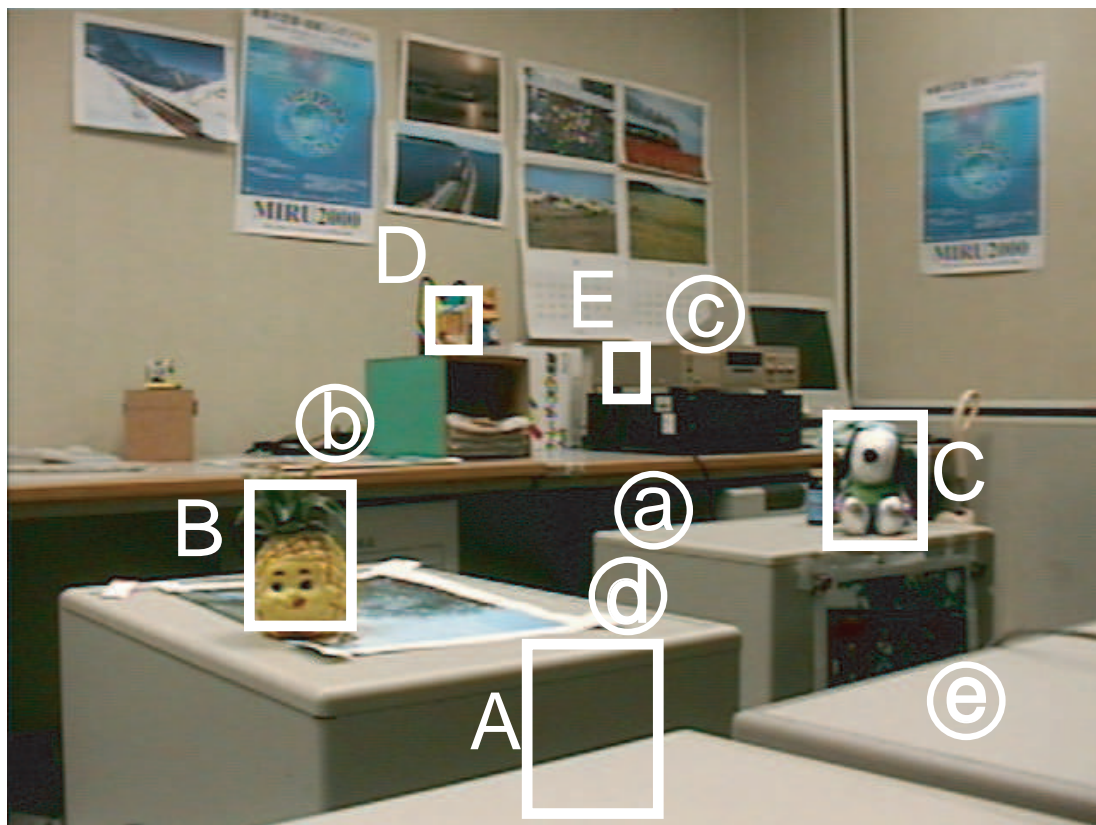


図 46 1 台のパン・チルト・ズームカメラによる物体検出実験で用いた物体の学習位置と探索位置

a) **画像探索の精度評価** 3.2.1.2 節に述べた特徴による探索精度を評価するため、等分割のヒストグラム区画を用いた従来のアクティブ探索法 [MV98] との探索精度の違いのみを調べた。入力画像としては広角段階で撮影した画像 15 枚を用いた。

精度は適合率 (precision rate) と再現率 (recall rate) の平均値で評価した。ここで適合率とは、探索結果として出力されたもののうち正しいものの割合であり、再現率とは、探索されるべきもののうち探索結果として出力されたものの割合である。適合率と再現率がともに 100% であれば、検出もれや余分な検出がなかったことを意味する。

実験結果を表 2 に示す。この実験では、ウィンドウの大きさをパラメータとして精度を評価した。従来手法では、テンプレートを多数撮影したにも関わらず一辺の長さが 10 ~ 30 画素の範囲では、60% 程度の精度しか達成できなかった。一方、VQ の手法を用いた提案手法は、一辺 10 ~ 30 画素の物体に対しても、90% 近い精度を達成した。

b) **画像探索の速度評価** 3.2.1.2 節に述べた探索法の高速化の効果を評価するため、3.2.3.1 節 a) と同じ条件で入力画像 1 枚あたりの探索時間を測定した (表 3)。比較した手法は、従来手法 [MV98]、VQ による符号化を用いた VQ 法、VQ 法に加え並列探索を行った並列探索法、同様に併合探索を行った併合探索法の計 4 種類である。VQ 法、並列探索法、併合探索法は提案手法である。まず、VQ により約 2 倍程度、探索時間が向上していることがわかる。並列探索法を用いることによって精度を保証しつつ速度は更に 10% 程度改善している。一方併合探索法では、速度は更に 80% 程度向上する。併合探索により速度が向上している理由は、5 ~ 10

表 2 ベクトル量子化を行った色ヒストグラムによる物体検出精度

	一辺 30 画素以上	一辺 10 ~ 30 画素
従来手法	100%	57%
提案手法	100%	89%



枚の複数テンプレートの照合をまとめて1つで探索しているためである。併合探索では、物体検出の取りこぼしはないが別の領域を誤検出する危険はある。しかし、今回の場合、完全には精度を保証しないもののパン・チルト・ズームの制御に利用するためには十分な精度が得られている。

c) **パン・チルト・ズームカメラ制御への適用** 高速探索手法をパン・チルト・ズームカメラの予測制御に適用した場合について評価する。一辺 10 ~ 30 画素のテンプレートを予測用テンプレートに用い、一辺 10 ~ 30 画素以上のテンプレートを検証用テンプレートに用いた。今回の実験では、検出するために少なくとも1度のズームが必要な a ~ c にある物体を探索対象とした。

比較する手法は、予測なしの手法（従来手法）、従来のアクティブ探索に VQ を導入して物体の検出とカメラの動的制御を行う VQ 予測法、並列探索で物体の検出とパン・チルト・ズームの動的制御を行う並列探索予測法、並列探索で物体の検出を行いパン・チルト・ズームの動的制御を併合探索で行う併合探索予測法の4通りである。

実験結果を表 4 に示す。探索精度はいずれも 100% であった。動的制御により物体検出までの時間を約 50% 削減できたことがわかる。更に並列探索で 10%、併合探索で 50% の探索時間を削減できた。

表 3 アクティブ探索を拡張した高速探索手法の速度評価

	一辺 30 画素以上	一辺 10 ~ 30 画素
従来手法	2.7s	9.1s
VQ 法	1.7s	3.9s
並列探索法	1.5s	3.5s
併合探索法	0.33s	0.35s

### 3.2.3.2 パン・チルト・ズームカメラの台数増加による効果の評価

本実験ではまず、動的アクティブ探索法による探索速度向上の効果を確認し、次に複数台のパン・チルト・ズームカメラを利用することによる物体の位置推定精度を評価する。実験環境を図 47 で示す。図 47 の  $a \sim g$  に置かれた物体を、カメラ 1 ~ 4 で探索した。

テンプレートとして、あらかじめ異なる 5 箇所、3 方向に物体を配置し、ズームを変えて撮影した約 100 枚を用いた。テンプレートのうち、100 ~ 400 画素のテンプレート約 50 枚を予測用テンプレートとして使い、400 ~ 2000 画素のテンプレート約 50 枚を検証用テンプレートとして用いた。閾値は予測用に .85 を使い、検証用には .65 を用いた。実験に用いた計算機およびカメラの仕様は表 5 の通りである。

**探索時間** 予測制御なし、予測制御ありの 2 つの場合について探索時間を比較する。4 台のパン・チルト・ズームカメラのうち最も早く検出したカメラの探索時間、4 台のカメラの平均探索時間を表 6 に示す。予測制御の導入により速度は、2 倍程度高速になる。また複数のパン・チルト・ズームカメラを利用した探索を行えば、1 台のカメラで探索する時間より、3 倍以上高速に探索できることがわかる。これは物体を探索しやすい位置、方向にあるカメラが高速に探索できるからである。予測制御と複数台のカメラによる探索により合計で 6 倍以上高速化できることになる。

表 4 1 台のパン・チルト・ズームカメラによる探索時間の比較

	時間
従来手法	44.6s
VQ 予測法	22.6s
並列探索予測法	20.5s
併合探索予測法	9.6s

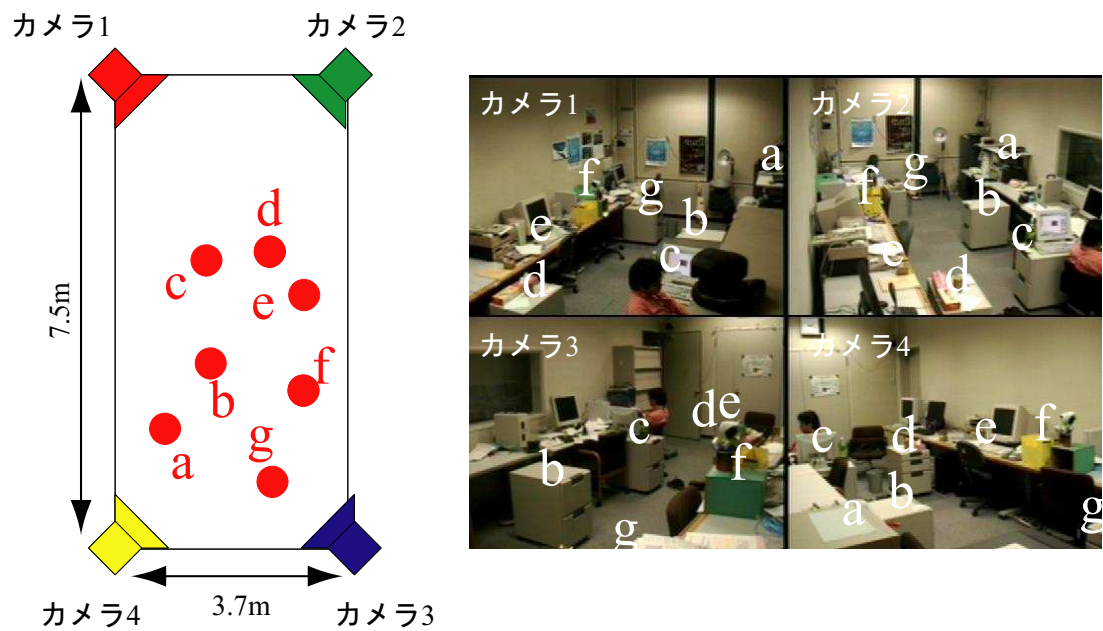


図 47 複数台のパン・チルト・ズームカメラを使った実験環境

表 5 複数台のパン・チルト・ズームカメラを使った実験の計算機およびカメラの仕様

計算機	SGI 社 O <sub>2</sub>
CPU	R12000(400MHz)
カメラ	Sony 社 EVI-D30
解像度	320 × 240

表 6 複数台のパン・チルト・ズームカメラを使った探索時間の比較

	最速	平均
予測なし	31.9s (16.9s,15.0s)	89.1s (49.8s,39.3s)
予測あり	15.3s (7.2s,8.2s)	40.8s (18.6s,22.2s)

ただし（）内は画像探索時間, パン・チルト・ズームカメラ制御時間

表 7 複数台のパン・チルト・ズームカメラを使った位置推定精度

	1台（平均）	1台（最良）	4台
平均誤差	124.8cm	47.6cm	18.9cm
標準偏差	134.4cm	18.5cm	10.5cm

**物体の位置推定精度** 本実験では、動的アクティブ探索法による3次元物体の3次元位置の推定精度を調べた。

1台のパン・チルト・ズームカメラによる推定結果の平均誤差、4台のカメラの推定結果の中の最小誤差、カメラ4台の結果を統合した推定誤差を表7に示す。

ヒストグラム特徴による照合の場合、物体の一部と一致することがあるので、方向や大きさの推定精度はあまり高くない。特に大きさに関しては、大きさの異なるテンプレートと一致した場合には大きな誤差となる。1台のカメラの位置推定結果の中で最も精度が良いものは約50cm程度の誤差で済んでいるものの、1台のカメラからの結果の平均値はあまり精度は高くない。しかし4台のカメラの計測結果を統合すれば、精度は大きく向上することがわかる。動的アクティブ探索法では、物体までの距離の推定精度に比べ方向推定の精度が良好に得られるため、複数カメラによる測量が効果的に行えたと推察される。

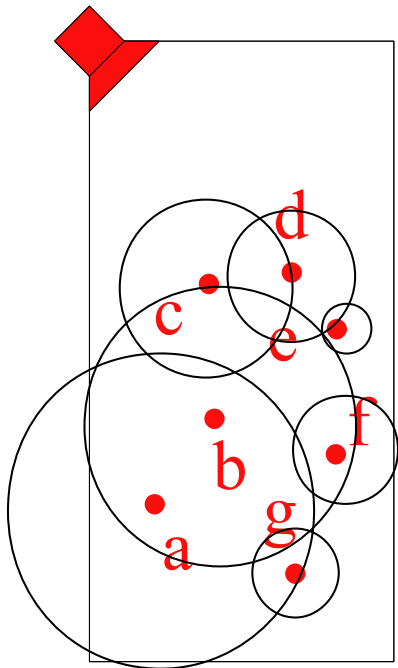
**物体の位置による探索時間・位置推定精度に関する考察** 物体の位置による探索時間の違いを図48に示す。図48の円の半径は探索時間を表している。左が予測ありの場合のカメラ1だけを用いた場合の探索時間であり、右が予測ありの場合の4台のカメラを用いた場合の探索時間である。カメラ1から遠いほど、時間がかかる傾向がある。遠い物体でも近い物体より速く検出する場合があるのは、探索時間が、予測制御の性能にも依存するからである。予測段階での誤った予測に対して、検証段階で検証した結果、バックトラックを繰り返す場合などに速度が低下する。このように距離や物体の方向により探索時間に大きな違いがでるため図48右のように複数のカメラを用いると検出しやすいカメラが存在する確率が高くなり、速度向上が実現できる。

同様に物体の位置による精度の違いを図49に示す。図49中の円の半径は、推定誤差である。図49左にカメラ1により推定した場合の誤差を示し、図49右に、複数のカメラを用いた三角測量により推定した場合の誤差を示す。

1台のカメラによる誤差は、場所により大きく異なる。これは、色ヒストグラム特徴による検出の場合、物体の見え方によって、物体の大きさを正しく検出することが難しいため、カメラからの距離推定が困難であることによる。

また、物体画像の一部が類似度の最大値になってしまう場合がありこの場合に

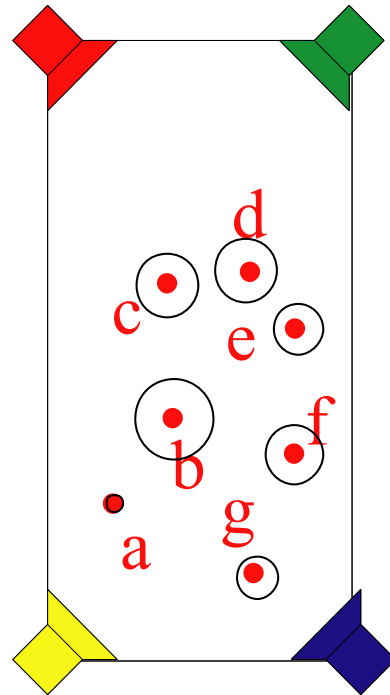
カメラ1



1台のカメラ

カメラ1

カメラ2

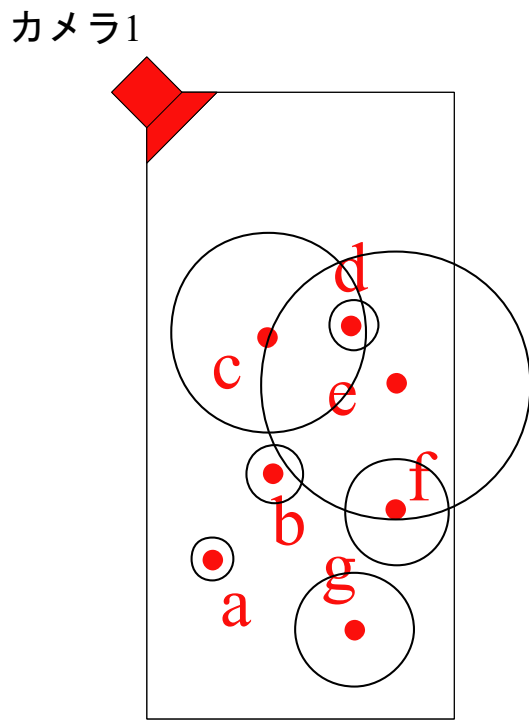


カメラ4

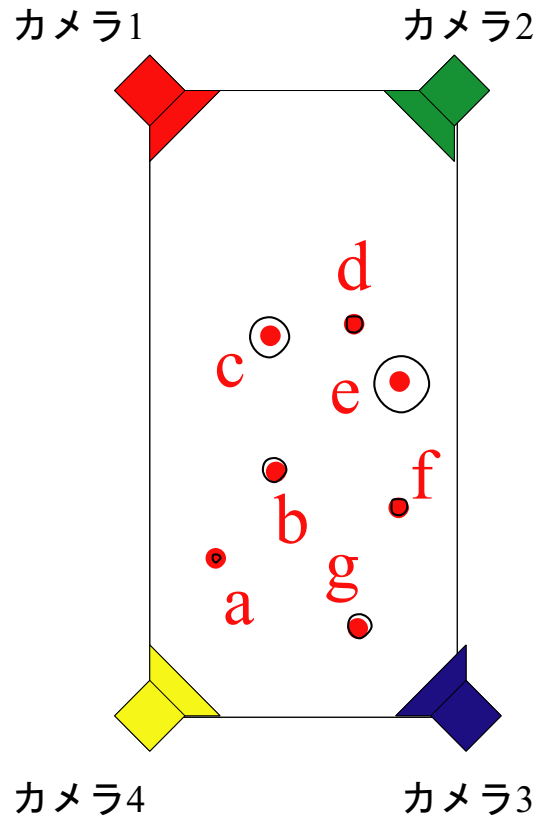
カメラ3

複数のカメラ

図 48 複数台のパン・チルト・ズームカメラを使った場合の物体位置に対する探索時間の比較



1台のカメラ



複数のカメラ

図 49 複数台のパン・チルト・ズームカメラを使った場合の物体位置に対する位置推定精度の比較

は、物体の一部を物体全体として認識してしまうため、方向推定・距離推定の誤差が生じる。カメラに近いところほど分解能が粗いため、方向推定誤差・距離推定誤差が大きくなる。

一方、動的アクティブ探索法から得られる方向情報のみを扱う複数のカメラの測定結果を統合した手法では、良好な精度が得られた。動的アクティブ探索法では、ズームを行うことによって方向を表す角度の分解能が高いために物体の部分的な特徴を検出した場合に生じる多少の方向ずれはほとんど影響を与えないからである。

#### 3.2.4 考察

3.2節では、パン・チルト・ズームカメラを用いて実環境からの物体探索を可能とする動的アクティブ探索法を提案した。約5mの範囲内の最小5cm程度の大きさの物体をカメラを制御して探索した場合、平均約10秒で検出可能であった。これは従来の探索の手法にアクティブ探索法を用い、単純なカメラ制御を行った場合に比較し4倍以上の速度向上に相当する。また複数のカメラを利用することで、1台のカメラの場合よりも探索時間を大幅に短くすることが可能となる。室内環境に4台のカメラを設置し、30m<sup>2</sup>ほどの室内環境の様々な位置に物体を配置したところ、平均15秒で探索でき、誤差20cmで物体位置が得られた。これはカメラ1台で予測制御なく探索した場合の6倍の速度に相当する。しかしながら、3.2節で提案した手法は、物体を色ヒストグラム特徴で識別できる場合にしか適用できず、背景色が物体色に類似する場合には物体を見分けることができず、適用範囲が色が鮮やかな物体に限定される。このため、色ヒストグラムに限定しない一般の距離特徴での高速化手法も必要になる。

本手法では、幾何学的構造が異なるが類似した色ヒストグラムを持っている物体や、ある程度大きな隠蔽などがある場合にも精度が低下することがある。これらの問題に対してはテンプレートを複数領域に分割してサブヒストグラムを構成し、サブヒストグラム毎に照合する手法などが考えられる。これらテンプレートの構成法については今後さらに検討する必要がある。



### 3.3 一般距離特徴に拡張した物体検出の高速化

前節では、室内環境に配置した静止物体を見つけ出すため、色ヒストグラム特徴を使って探索を高速化する技術を提案した。しかしながら、色ヒストグラム特徴を利用した場合には、色が類似した物体を見極めることが難しく、また色が背景色に類似した場合には、検出できないといった問題がある。

本節ではアクティブ探索 [MV98] で用いられている類似度上限値によるスキップの概念を一般的な距離によるテンプレートマッチングに適用し、距離下限値を用いた照合の枝刈りによって全探索と同一の精度を保証しながら、テンプレートマッチングを高速化する手法を提案する [KKK+05a]。従来の距離下限値を使う手法 [VR77, LS95, LC97, GA01] に比べ、提案手法は、テンプレート内部の特徴の類似性を利用するため、図 50 の例のようにテンプレート内部の色の違いが小さい場合には効果が大きい。また、テンプレート内部の類似性という従来手法とは異なる性質を用いるため、従来手法と組み合わせた時の相乗効果も期待できる。

以下、3.3.1 節でテンプレートマッチングの枠組み、その高速化手法である SSDA 法 [BS72] について簡単に述べる。3.3.2 節では提案手法である適応的なウィンドウスキップによる高速テンプレートマッチング法について述べ、SSDA 法と組み合わせるための実装上の工夫について述べる。3.3.3 節で複数テンプレートを探査する場合の方法について述べ、3.3.4 節では実験により提案手法をいくつかの従来手法と比較し、提案手法の有効性を検証し、3.3.5 節で本節の手法を考察する。

#### 3.3.1 テンプレートマッチング

テンプレートマッチングは、あらかじめ与えられたテンプレートに最も類似する領域を検出する手法である。テンプレートを  $\mathbf{T}$ 、入力画像上のある場所  $(u, v)$  におけるウィンドウを  $\mathbf{W}^{(u,v)}$  とし、テンプレート  $\mathbf{T}$  とウィンドウ  $\mathbf{W}^{(u,v)}$  における距離を  $d(\mathbf{T}, \mathbf{W}^{(u,v)})$  と記述する。この距離には、絶対値距離 (Sum of Absolute Difference : SAD) が計算量が小さいためよく用いられる。

$$SAD : d(\mathbf{T}, \mathbf{W}^{(u,v)}) = \sum_{p=0}^{N-1} |T^p - W^p| \quad (28)$$

検出したウィンドウ

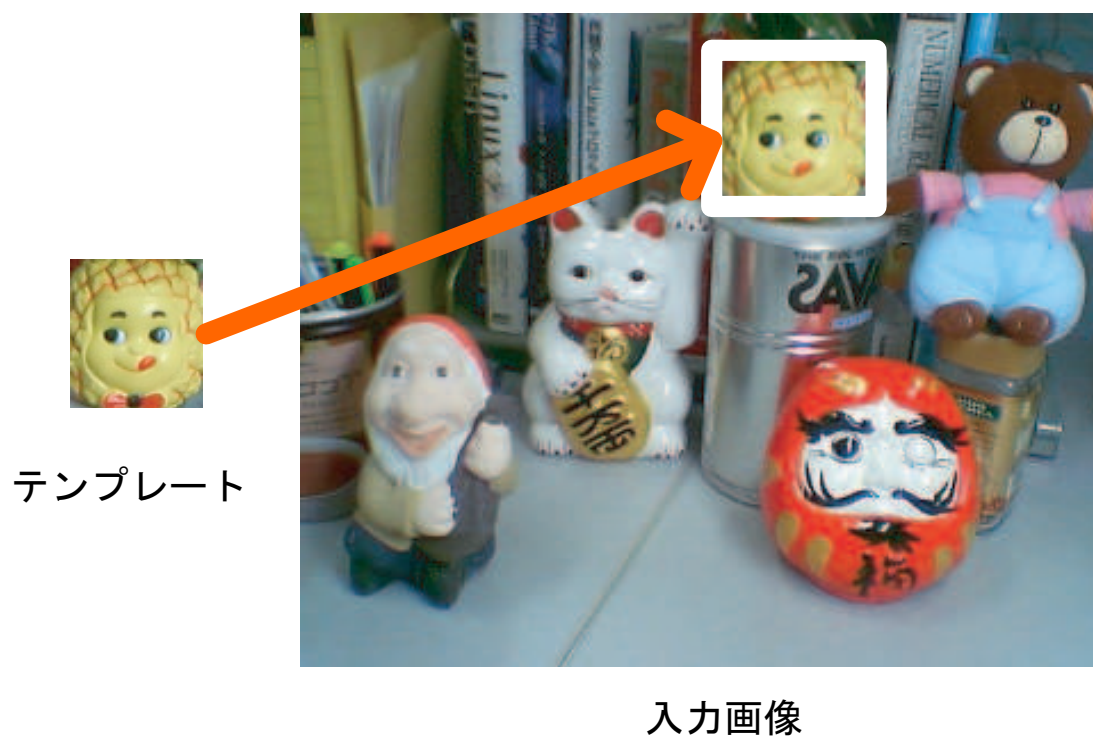


図 50 テンプレートマッチングによる物体検出の例

ここで  $T^p, W^p$  は,  $\mathbf{T}, \mathbf{W}^{(u,v)}$  のそれぞれ  $p$  番目の画素の色を表す値であり,  $N$  は,  $\mathbf{T}$  と  $\mathbf{W}^{(u,v)}$  の画素数である. SAD 以外にも二乗誤差 (Sum of Squared Differences : SSD) などにも利用できる.

テンプレートのサイズを  $(M_x \times M_y)$ , 入力画像のサイズを  $(X \times Y)$  としたとき,  $(u, v)$  はそれぞれ,  $0 \leq u < X - M_x, 0 \leq v < Y - M_y$  の範囲を動く. このため, ウィンドウの総数は  $(X - M_x) \times (Y - M_y)$  になる. これら全てのウィンドウで距離  $d(\mathbf{T}, \mathbf{W}^{(u,v)})$  を計算する必要がある. ウィンドウをずらしていく過程で,  $d(\mathbf{T}, \mathbf{W}^{(u,v)})$  が閾値  $\theta$  を下回る場合,  $\theta$  を  $d(\mathbf{T}, \mathbf{W}^{(u,v)})$  に更新し,  $(u, v)$  を記録する.

### 3.3.1.1 テンプレート マッチングの高速化

SSDA は, (28) 式の計算が画素の評価順序に従って単調に増加することを利用して, ある部分画像との照合時に, 全ての画素を比較し終える前に不一致であることを検定し, 距離計算を打ち切る手法である.  $n$  番目の画素まで評価したときの距離  $D(n < N)$  を次のように定義する.

$$D(n) = \sum_{p=0}^n |T^p - W^p| \quad (29)$$

ここで,  $D(n)$  が単調増加であるから,  $D(n)$  が次の条件を満たすとき, 計算を打ち切ることができる.

$$D(n) \geq \theta \quad (30)$$

ここで, 毎回閾値比較をすると分岐処理による速度の低下を生じるため, 閾値比較の間隔を適切に開ける必要がある.

### 3.3.2 適応的ウィンドウスキップによる高速テンプレート マッチング法: Adaptive Window-Skipping Method (AWS)

3.3節で提案する適応的ウィンドウスキップ法 (AWS) は, ウィンドウの一部 (サブウィンドウ) を多数のサブテンプレートと比べることによって, 周囲にある多数のウィンドウの距離下限値を求め, その距離下限値が閾値を上回る場合に, 照合

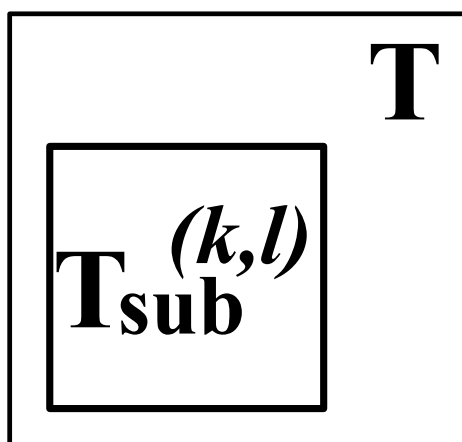


図 51 サブテンプレート

の省略を行う。このため、照合の不要なウィンドウを適応的にスキップしながら探索を行うことができる。

サブテンプレートの概念を、図 51 に示す。サブテンプレートは、テンプレートのサイズ  $(M_x, M_y)$  より  $(m_x, m_y)$  小さい領域を切り出したものである。この  $(m_x, m_y)$  をマージンと呼び、テンプレート  $\mathbf{T}$  における位置  $(k, l)$  におけるサブテンプレートを  $\mathbf{T}_{\text{sub}}^{(k,l)}$  と表記する。

### 3.3.2.1 AWS の概要

AWS は、あるテンプレートとあるウィンドウを照合した結果、その距離が十分大きいときに、近傍のウィンドウの距離下限値を推定する。ウィンドウを照合する前に、そのウィンドウにおける距離下限値が閾値を超えるかどうか判定し、距離下限値が閾値を上回る場合には、そのウィンドウでの照合を省略し、計算量を削減する。すなわち、図 52 のように、いったん  $\mathbf{T}$ ,  $\mathbf{W}_A$  の距離  $d(\mathbf{T}, \mathbf{W}_A)$  を計算し、 $d(\mathbf{T}, \mathbf{W}_A)$  が大きい場合には  $\mathbf{W}_A$  の近傍のウィンドウ  $\mathbf{W}_B$  の距離下限値を計算する。しかし、SAD や SSD を距離尺度に用いる場合、ウィンドウの走査により、 $\mathbf{W}$  の各要素がシフトするため、 $\mathbf{T}$  と  $\mathbf{W}_A$  との距離から、 $\mathbf{T}$  と  $\mathbf{W}_B$  の距離を推定することは困難である。

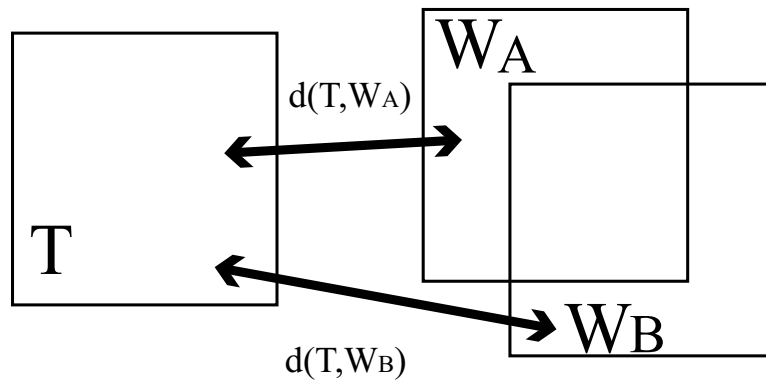


図 52 テンプレート  $T$  と ウィンドウ  $W_A$  の距離を用いて, テンプレート  $T$  と ウィンドウ  $W_B$  の距離を推定

そこで, AWSでは, 図 53のように, テンプレート内のサブテンプレートと, ウィンドウ内のサブウィンドウを使って距離下限値を求める.

まず  $T, W_A$  を照合する代わりに,  $T_{\text{sub}A}, W_{\text{sub}}$  の距離を計算する. 次にその近傍のウィンドウの距離下限値を推定することになるが,  $T_{\text{sub}A}$  と,  $W_{\text{sub}}$  の近傍のサブウィンドウとの距離下限値を求めるのではなく,  $W_{\text{sub}}$  と,  $T_{\text{sub}A}$  の近傍のサブテンプレート  $T_{\text{sub}B}$  との距離下限値を求める.

サブテンプレート間の距離  $d(T_{\text{sub}A}, T_{\text{sub}B})$  はあらかじめ求めることができるので, このサブテンプレート間の距離と,  $T_{\text{sub}A}, W_{\text{sub}}$  の距離から, 3.3.2.3 節に述べる三角不等式を使って,  $W_{\text{sub}}$  と,  $T_{\text{sub}B}$  の距離下限値を求める. SADやSSDは画素の増加に対して単調増加であるから, サブテンプレートとサブウィンドウ間の距離はテンプレートとウィンドウ間の距離の距離下限値になる. つまり求めた  $W_{\text{sub}}$  と,  $T_{\text{sub}B}$  の距離下限値が閾値を上回れば, テンプレート  $T$  とウィンドウ  $W_B$  間の距離計算を省略できる. テンプレート内のサブテンプレート間の距離をあらかじめ求めておき, 上記の処理により, 1度のサブテンプレートとサブウィンドウの照合の結果, その距離が大きい場合には, 近傍の多くのウィンドウにおける距離下限値が計算でき, その値が閾値を上回れば, 距離計算が省略でき, 計算量の大幅な削減につながる.

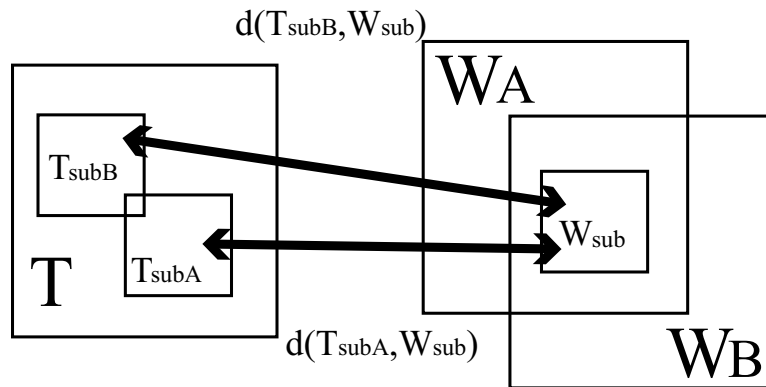


図 53  $T$  と  $W_A$  の距離を用いて、 $T$  と  $W_B$  の距離を推定する代わりに、 $T_{subA}$  と  $W_{sub}$  の距離を用いて、 $T_{subB}$  と  $W_{sub}$  の距離を推定

### 3.3.2.2 AWS の処理の流れ

AWS は次の 2 つの処理からなる。

#### 前処理

1. 探索に入る前に、テンプレートからサブテンプレートを全て切り出す。
2. サブテンプレートからリファレンスサブテンプレート  $\mathbf{R}_{sub}^{(i,j)}$  を 1 つ選ぶ。  
(以降、 $\mathbf{R}_{sub}$  と記す。)
3.  $\mathbf{R}_{sub}$  とすべての  $\mathbf{T}_{sub}^{(k,l)}$  の距離  $d(\mathbf{R}_{sub}, \mathbf{T}_{sub}^{(k,l)})$  を求める。
4. 全ウィンドウの距離下限値を 0 に初期化する。

#### 探索時の処理

1. 残るウィンドウがなければ終了。ウィンドウがあれば、1 つ取り出す。
2. このウィンドウの距離下限値が閾値より大きければ、照合を省略。1. へ。

3. リファレンスサブテンプレートに対応するサブウィンドウを切り出す.
4. リファレンスサブテンプレートとサブウィンドウの距離 (部分距離) を計算する.
5. 部分距離が閾値より大きいとき, 周囲のウィンドウの距離下限値を更新し, 1. へ.
6. 部分距離が閾値より小さいとき, テンプレートとウィンドウの残りの部分の距離も計算し, テンプレートとウィンドウの距離 (全体距離) を求める.
7. 全体距離が閾値より小さいとき, 閾値をこの全体距離の値に更新し, この位置を記録する.
8. 1. へ.

### 3.3.2.3 距離下限値の計算

一度, リファレンスサブテンプレート  $\mathbf{R}_{\text{sub}}$  と, サブウィンドウ  $\mathbf{W}_{\text{sub}}^{(u,v)}$  の照合が行なわれたとき, テンプレート  $\mathbf{T}$  と,  $\mathbf{W}_{\text{sub}}^{(u,v)}$  を含むウィンドウ  $\mathbf{W}^{(u-k,v-l)}$  (ただし,  $0 \leq k < m_x, 0 \leq l < m_y$ ) は, 次のように求められる.

すでに, サブテンプレート間の距離  $d(\mathbf{R}_{\text{sub}}, \mathbf{T}_{\text{sub}}^{(k,l)})$  と, リファレンスサブテンプレート  $\mathbf{R}_{\text{sub}}$  とサブウィンドウ  $\mathbf{W}_{\text{sub}}^{(u,v)}$  の距離が既知であるので, 図 54 に示す三角不等式の関係から, サブテンプレート  $\mathbf{T}_{\text{sub}}^{(k,l)}$  とサブウィンドウ  $\mathbf{W}_{\text{sub}}^{(u,v)}$  の距離  $d(\mathbf{T}_{\text{sub}}^{(k,l)}, \mathbf{W}_{\text{sub}}^{(u,v)})$  に関して次式が導かれる.

$$\begin{aligned} d\left(\mathbf{T}_{\text{sub}}^{(k,l)}, \mathbf{W}_{\text{sub}}^{(u,v)}\right) & \\ & \geq d\left(\mathbf{R}_{\text{sub}}, \mathbf{W}_{\text{sub}}^{(u,v)}\right) - d\left(\mathbf{R}_{\text{sub}}, \mathbf{T}_{\text{sub}}^{(k,l)}\right) \end{aligned} \quad (31)$$

次に, SAD や, SSD の距離が, 単調増加であることから, テンプレートとウィンドウの距離は, サブテンプレートとサブウィンドウの距離より大きい. すなわち,

$$\begin{aligned} d\left(\mathbf{T}, \mathbf{W}^{(u-k,v-l)}\right) & \\ & \geq d\left(\mathbf{T}_{\text{sub}}^{(k,l)}, \mathbf{W}_{\text{sub}}^{(u,v)}\right) & (32) \\ & \geq d\left(\mathbf{R}_{\text{sub}}, \mathbf{W}_{\text{sub}}^{(u,v)}\right) - d\left(\mathbf{R}_{\text{sub}}, \mathbf{T}_{\text{sub}}^{(k,l)}\right) \end{aligned}$$

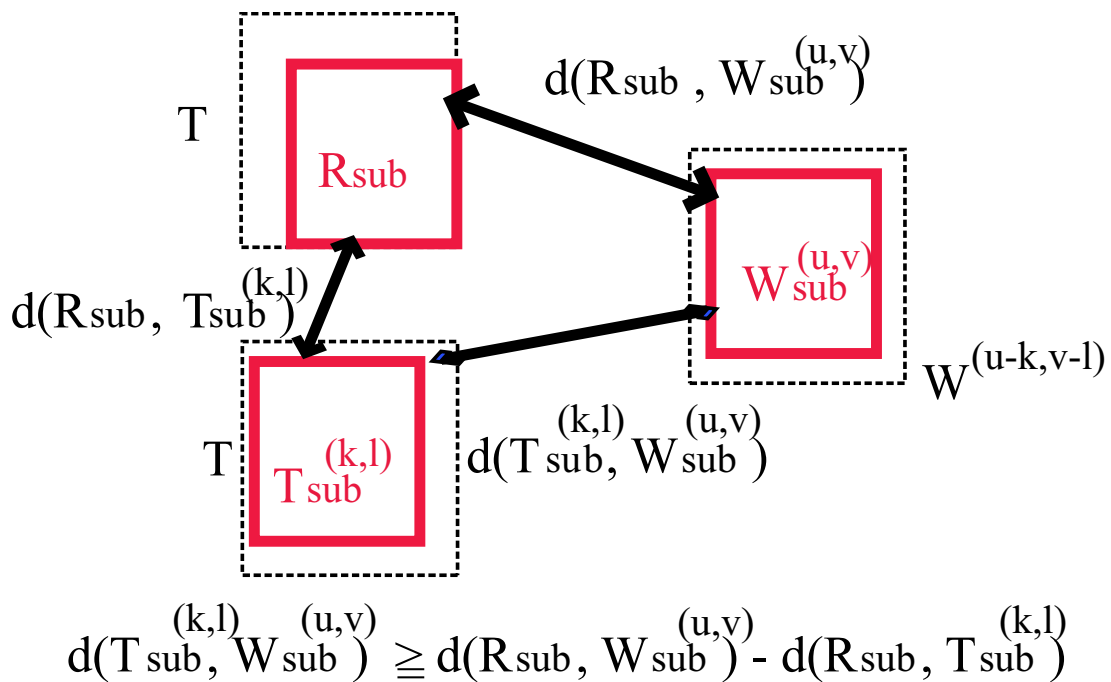


図 54 三角不等式による距離下限値の導出

が成立する. (32) 式は,  $d(R_{\text{sub}}, W_{\text{sub}}^{(u,v)}) - d(R_{\text{sub}}, T_{\text{sub}}^{(k,l)})$  がテンプレート  $\mathbf{T}$  とウィンドウ  $\mathbf{W}^{(u-k,v-l)}$  の距離  $d(\mathbf{T}, \mathbf{W}^{(u-k,v-l)})$  の距離下限値であることを意味する. この距離下限値が, 閾値を越えた場合, ウィンドウ  $\mathbf{W}^{(u-k,v-l)}$  における照合が省略できる.

### 3.3.2.4 実装上の工夫

**リファレンスサブテンプレートの選択とウィンドウの走査順序** リファレンスサブテンプレートの選択と, ウィンドウを走査する順序は AWS の効率を考える上で非常に重要である. たとえば, 左上のリファレンスサブテンプレート  $\mathbf{R}_{\text{sub}}^{(0,0)}$  を使用して AWS を実行した場合, 距離下限値を計算するウィンドウは, 照合したウィンドウの左上の方向になる. すなわち, 左上から右下へ走査した場合には, 距離下限値を計算するウィンドウは, すでに走査し照合の終わったウィンドウになっ



てしまい、照合省略の効果が見られなくなる。

距離下限値計算の効果を生かすためには、左上のサブテンプレートをリファレンスサブテンプレートに選んだ場合には、ウィンドウは、右下  $\mathbf{W}^{(X-M_x, Y-M_y)}$  から左上  $\mathbf{W}^{(0,0)}$  へ走査し、右下のサブテンプレートをリファレンスサブテンプレートに選んだ場合には、ウィンドウは、左上  $\mathbf{W}^{(0,0)}$  から右下  $\mathbf{W}^{(X-M_x, Y-M_y)}$  へ走査する必要がある。

**AWS と SSDA との組み合わせ** SSDA 法は、リファレンスサブテンプレートとサブウィンドウの照合にも利用できる。

リファレンスサブテンプレートとサブテンプレートの距離の最大値、 $D_{\max}$  は、次の式で表現される。

$$D_{\max} = \max \left\{ |\mathbf{R}_{\text{sub}} - \mathbf{T}_{\text{sub}}^{(k,l)}| \right\} \quad (33)$$

for any  $k$  and  $l(0 \leq k < m_x, 0 \leq l < m_y)$

もし、 $D(n)$  が、 $\theta + D_{\max}$  より大きくなれば、そのサブウィンドウを含む、全てのウィンドウにおいて (32) 式が  $\theta$  より大きくなり、これら全てのウィンドウの照合を省略することができる。 $D(n)$  が、 $\theta + D_{\max}$  を超えた場合、さらなる距離計算は必要ない。

AWS のフローチャートを図 55 に示す。

### 3.3.3 複数テンプレートの高速探索のための拡張

リファレンスサブテンプレート  $\mathbf{R}_{\text{sub}}$  を切り出したテンプレートを  $\mathbf{T}$ 、 $\mathbf{T}$  とは異なるもう一つのテンプレートを  $\mathbf{T}'$  とする。予め、リファレンスサブテンプレート  $\mathbf{R}_{\text{sub}}$  と、 $\mathbf{T}'$  上のサブテンプレート  $\mathbf{T}'_{\text{sub}}^{(k,l)}$  との距離が計算してあれば、 $\mathbf{R}_{\text{sub}}$  リファレンスサブテンプレートとあるサブウィンドウとの照合結果を用いて、他のテンプレートで照合するときの距離下限値も (32) 式の導出と同様に計算でき、以下のようなになる。

$$\begin{aligned} d(\mathbf{T}', \mathbf{W}^{(u-k, v-l)}) &\geq d\left(\mathbf{T}'_{\text{sub}}^{(k,l)}, \mathbf{W}_{\text{sub}}^{(u,v)}\right) \\ &\geq d\left(\mathbf{R}_{\text{sub}}, \mathbf{W}_{\text{sub}}^{(u,v)}\right) - d\left(\mathbf{R}_{\text{sub}}, \mathbf{T}'_{\text{sub}}^{(k,l)}\right) \end{aligned} \quad (34)$$

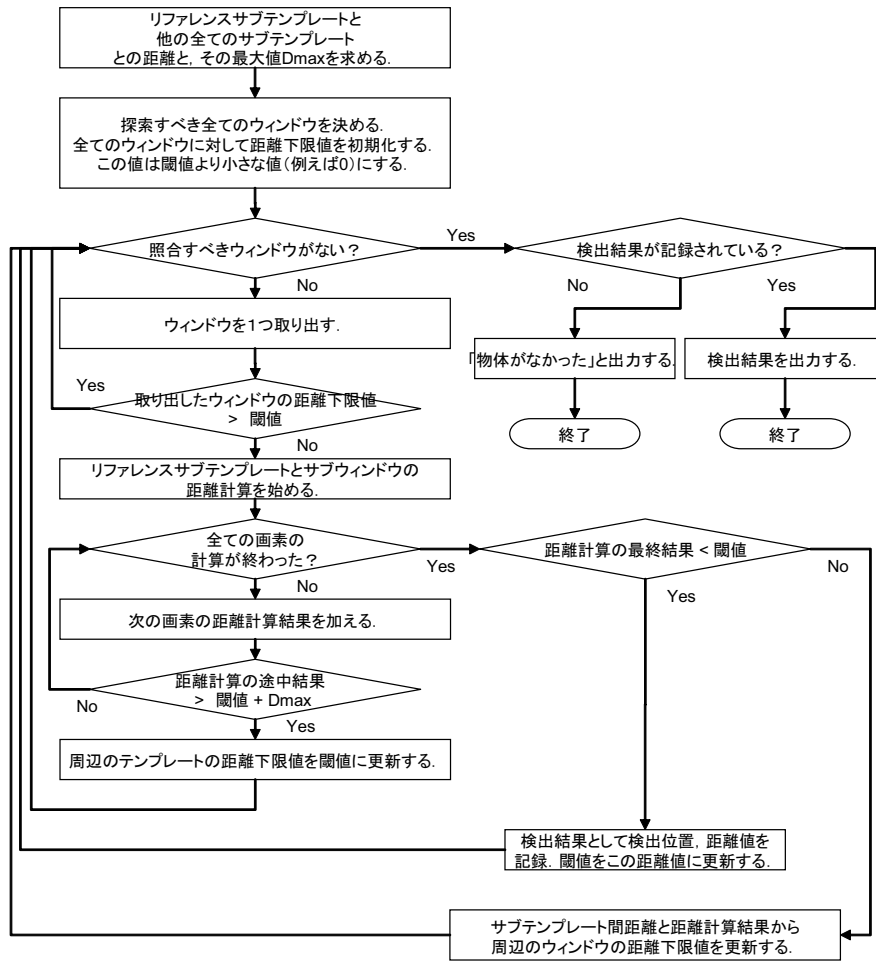


図 55 AWS のフローチャート

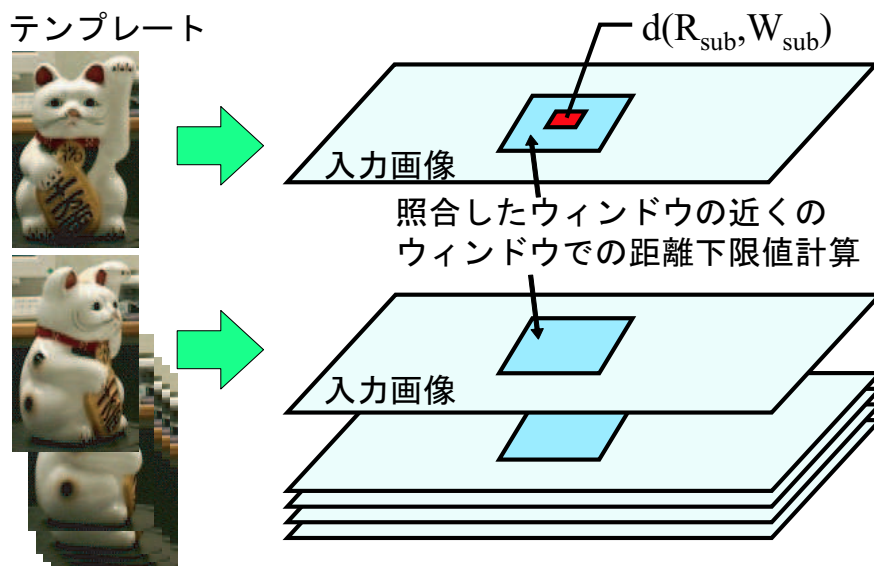


図 56 異なるテンプレート間での距離下限値計算

すなわち、異なるテンプレート間でもこれまでと全く同様に距離下限値を計算することができる [KKT04] .

そこで、前処理段階で、リファレンスサブテンプレートと他のテンプレート上に作成したサブテンプレートとの距離を計算する。部分窓と、リファレンスサブテンプレートとの距離計算を行ったとき、類似する複数のテンプレートに対して距離下限値の計算を行う (図 56) .

複数テンプレートに対する AWS のフローチャートを図 57 に示す。

### 3.3.3.1 距離下限値を求める範囲

角度や距離などの撮影パラメータが大きく異なるテンプレートでは、一般にサブテンプレート間の距離が大きくなる。このようなテンプレートに対して距離下限値計算を行ったとしても、照合を省略できる可能性は少ない。一方で距離下限値計算を行う領域が増加するというデメリットが生じる。そこでテンプレートのグループ化を行い、距離下限値計算を行う領域を限定する。

リファレンスサブテンプレートを選んだ後、それを含むテンプレートと撮影パ

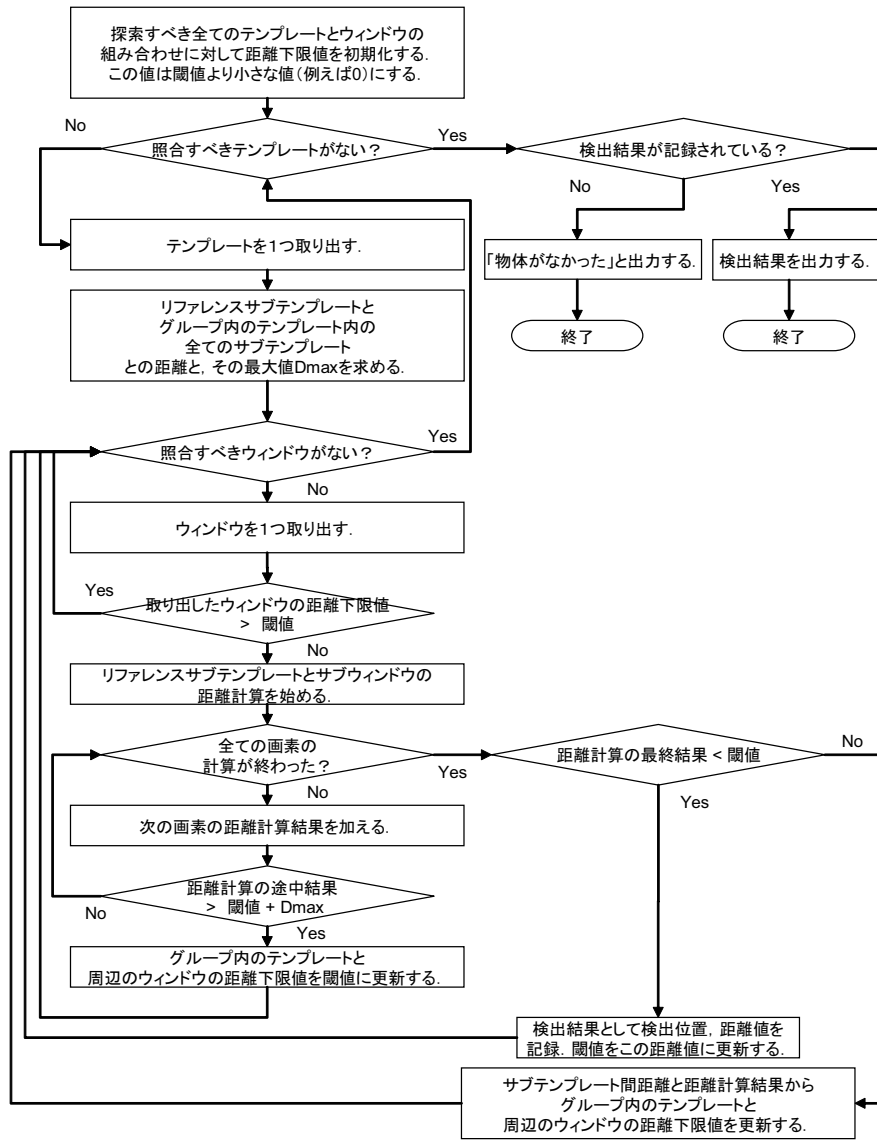


図 57 複数テンプレートに対する AWS のフローチャート

ラメータの近いテンプレートを  $N$  枚選ぶ。これをグループとよぶ。このグループ内で、リファレンスサブテンプレートと他の全てのサブテンプレートとの距離を求め、距離下限値計算に利用する。全てのグループリファレンスサブテンプレートと内のサブテンプレートとの距離を求める。

### 3.3.3.2 複数のスケールのテンプレートの探索

テンプレートの大きさが変わるため、閾値をテンプレートの大きさに合わせて変える。閾値には  $th \times P$  を用いる。  $th$  は 1 画素あたりの許容誤差、  $P$  は画素数である。テンプレートの大きさが異なると、テンプレートごとのマージンが変化する。マージンを大きくとっても、サブテンプレート間の距離が大きくなり、探索を省略できる可能性は低い。一方、マージンの変化により、テンプレートごとに探索領域が変化し、処理が複雑になる。このため、グループ化したテンプレートの中で一番小さいテンプレートサイズに合わせ、その他のテンプレート領域を切り取って処理を行った。

## 3.3.4 実験

### 3.3.4.1 1枚のテンプレートの探索実験

AWS の効果を確かめるために、テンプレートマッチング法 (TM), SSDA[BS72], SEA[LS95], BSPA[LC97] の各手法との探索時間の比較を行った。

TM は、すべての SAD を計算する基本的なテンプレートマッチング法である。SSDA は 高速化のためウィンドウの走査列が変わるときのみ、打ち切り判定を行った。BSPA は  $2^n \times 2^n$  のテンプレートサイズにのみ対応しているので、  $2^n \times 2^n$  のサイズで、探索した後、閾値を超えない領域に対して、残りの部分の距離計算を SSDA を用いて行った。同様に、AWS のサブテンプレートの残りの部分の距離計算にも SSDA を用いた。これらの全ての手法は厳密に同じ解を保障しているので、精度は全く同じである。

本実験では、Pennsylvania State University から配布された実画像データセット [WLW01, IMG] を用いた。このデータセットは 10 カテゴリー各 100 枚の JPEG



図 58 AWS の評価実験に用いた入力画像の例



図 59 AWS の評価実験に用いたテンプレートの例

画像からなる。実験には各カテゴリーからランダムに選んだ各 1 枚を用いた。図 58 に実験に使用した画像の例を示し、実験に使用したテンプレートの例を図 59 に示す。図中の画像の下の番号はデータセットにおける画像番号である。

入力画像には、テンプレートが含まれる各画像に対して 30db のガウスノイズを加えたもの 10 枚を使用した。各入力画像のサイズは  $384 \times 256$  である。画像の色表現には、RGB それぞれ 8bit の値を使用した。距離尺度には、絶対値距離を用いた。リファレンスサブテンプレートには、 $\mathbf{R}_{\text{sub}}^{(0,0)}$  を用い、ウィンドウは、右下から左上方向へ、走査した。本実験に使用した計算機の仕様を表 8 に示す。本実験のコードの重要な部分である絶対値距離計算には、Intel Pentium SSE 命令 (psadbw など) を使用した。

**探索時間とマージンサイズの関係** AWS は、マージンサイズによって性能が変わる。マージンを大きく取った場合、距離下限値を計算できるウィンドウの数は増えるものの、サブテンプレート間の類似性は低下し、 $D_{\text{max}}$  が大きくなり、SSDA の

効果が低下する。このため、マージンサイズの増加は必ずしも探索時間の短縮につながらない。マージンサイズと探索時間の関係を調べるため、マージンサイズを変えて、探索時間を測定した。使用したテンプレートの大きさは  $80 \times 80$ 、マージンの大きさは、 $(m_x = m_y = 0, 4, 8, 12, \dots, 40)$  である。ここで、 $m_x = m_y = 0$  のときの AWS は、SSDA と等価のアルゴリズムになる。 $\theta$  には、十分正解が見つかるのを保証できる  $\theta = 300000$  を用いた。図 60 に実験結果を示す。図 60 の右の番号は、使用したテンプレート番号であり、それぞれのテンプレートに対して、マージンの値を変えて探索時間を調べた結果を示している。

どのテンプレートに関してもマージンの値が 4 ~ 16 のあたりで、探索時間が最小になった。多くの物体では、マージン増加による照合回数省略と、 $D_{\max}$  増加による照合の計算量の増加が、打ち消しあう関係になり、マージンを大きくとってもほとんど計算速度に影響を与えない。しかしながら、846 のテンプレートのように、テンプレート内の色の違いが大きい場合には、マージンを大きく取ることによってサブテンプレート間の距離が大きくなるため、探索時間が著しく増加している。どの物体でもマージンの値が 4 ~ 16 辺りで探索時間は最小に近い値になる。この前後でも大きな差がないため、マージンの決定方法はテンプレートの大きさの 1 ~ 2 割あたりを選べばよいことがわかる。

**探索時間と初期閾値の関係** テンプレートマッチングの高速化手法は、すべて閾値により性能が変わる。閾値が低いほど、処理が早い段階で打ち切れるため、高い計算量削減効果が期待できるからである。本実験では閾値を 100000, 200000, 300000, ..., 1000000 に変えて実験を行い、各手法の性能を評価した。テンプレ

表 8 AWS の実験に使用した計算機とソフトウェアの仕様

CPU	Intel XEON CPU (3.2GHz) with EM64T Dual
Memory	4GB DDR2
OS	Fedora Core2 x86_64
Compiler	GNU C++ (-O3 オプション使用))

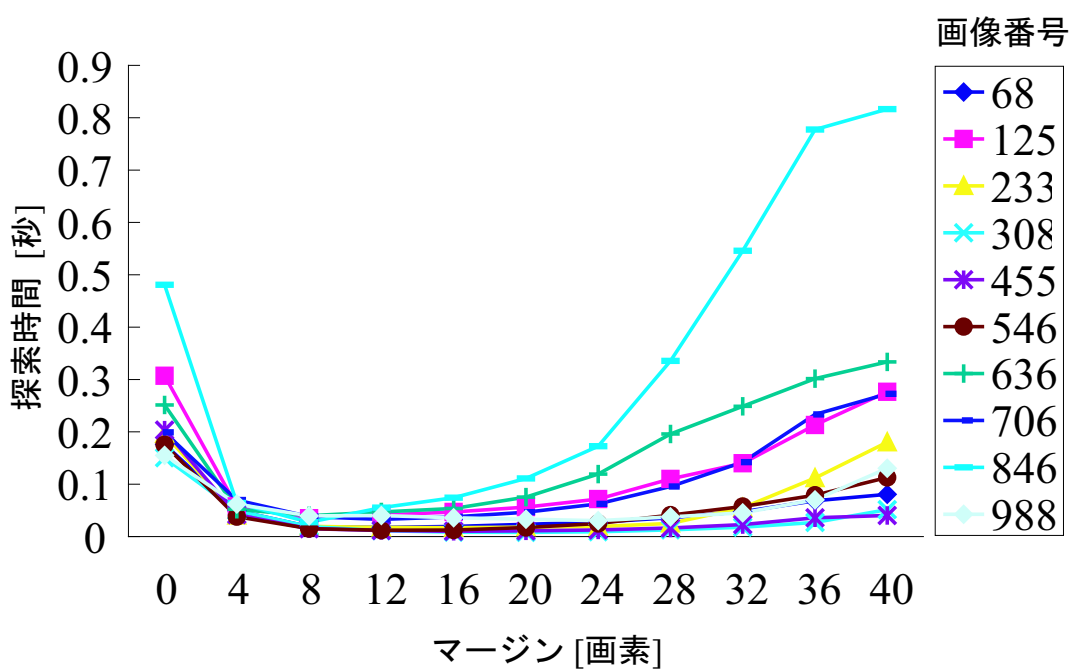


図 60 マージンサイズを変えて AWS の探索時間を測定した結果



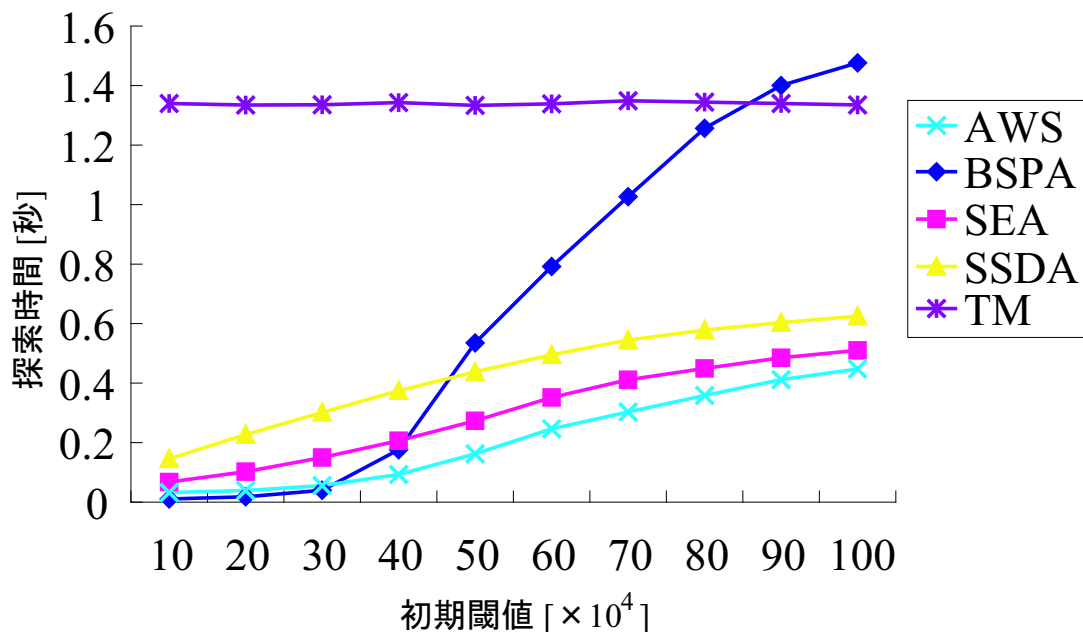


図 61 初期閾値を変えて AWS と他の手法の探索時間を比較した結果

トの大きさは  $80 \times 80$  とした。AWS のマージンサイズは  $m_x = m_y = 5$  を用いた。実験の結果を 図 61 に示す。

TM は閾値による高速化は行われなため、探索時間は一定である。計測結果が微妙に異なるのは約 0.2% の計測誤差のためである。

SEA,SSDA,AWS の各手法は、初期閾値の増加に伴い、なだらかに探索時間が増大する。一方で、BSPA は、初期閾値の増加により、計算時間が著しく増大する。初期閾値が大きい場合、平均画像による距離では打ち切れないため、多重に生成したすべての平均画像を探索する必要があるからである。閾値 100000 を用いたとき、BSPA, SEA, SSDA, AWS は、TM と比較して、.7%, 5%, 10%, 2.4% の探索時間であった。

**探索時間とテンプレートサイズの関係** BSPA は  $2^n \times 2^n$  のテンプレートサイズの時、最も効果を発揮するが、それ以外のテンプレートサイズの場合、SSDA の効果やその他の要因に依存する。そこで、さまざまなテンプレートサイズで探索時間

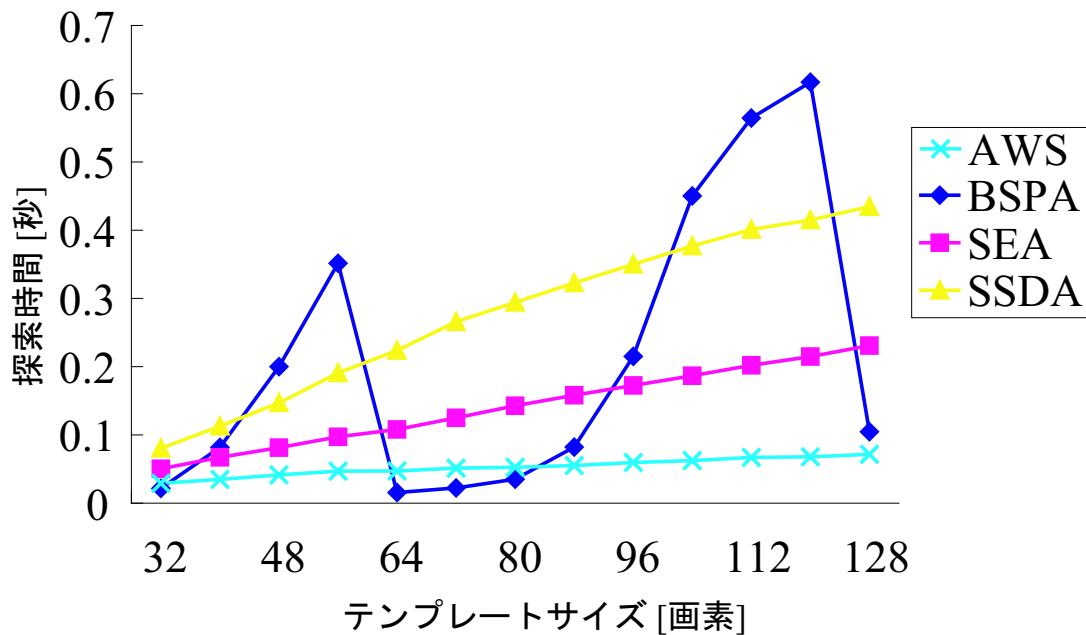


図 62 テンプレートサイズを変えて AWS と他の手法の探索時間を比較した結果を測定した。用いたテンプレートサイズは、 $32 \times 32, 48 \times 48, 64 \times 64, \dots, 128 \times 128$  である。閾値には、テンプレートサイズで正規化した値  $th \times P$  を用いた。  $th$  は 1 画素あたりの許容誤差、  $P$  は画素数である。本実験では  $th = 45$  とした。実験結果を図 62 に示す。ここでは、図の見易さのため、TM の表示を省いた。TM の探索時間は同じテンプレートサイズの SSDA と比べ約 4.4 倍の探索時間であった。

BSPA は、 $2^n$  に近い  $32, 64, 72, 80$  のとき最速であった。一方で、それ以外の場合には AWS が最速であり、BSPA の探索時間が SEA よりも大きいことが多かった。AWS は比較的全てのテンプレートサイズで良好な探索時間を示し、 $128 \times 128$  でも BSPA より優れた結果を残した。BSPA は閾値の増加に対して計算量の増加が著しいため、テンプレートが大きい場合に閾値が増加し、性能が低下したと思われる。一方で、AWS は閾値やテンプレートサイズの増加による探索時間の増加が小さく、適切な閾値が容易に決定できない場合や、テンプレートのサイズが限定できない場合にも効果がある。

### 3.3.4.2 複数テンプレートの探索実験

マージン  $m$ , グループ化時のテンプレート数  $N$  を変化させたときの探索時間を測定し, ほかの手法と比較する. 実験は

1. 回転角のみを変化させたテンプレート
2. スケールのみを変化させたテンプレート

の2種類に分けて行った. 本節では 3.3.3 節の拡張を含む AWS 法を拡張 AWS 法, 3.3.3 節の拡張を含まない AWS 法を単に AWS と記す.

**実験条件** 表 9 に実験仕様を示す. 実験に使用した物体を図 63 に示す.

**回転角を変化させたテンプレートによる探索** 対象物体を 10 度おきに撮影した 36 枚のテンプレートを用いた. テンプレートの大きさは  $48 \times 83$  である. 入力画像の大きさは  $320 \times 240$  である.

閾値は 1 度おきに撮影した 360 枚の入力画像からテンプレートと角度が異なる 324 枚を使用して, 探索を行い, Precision = Recall = 100% となる最小の閾値を測定して求めた値  $1 \times 10^5$  に設定した.

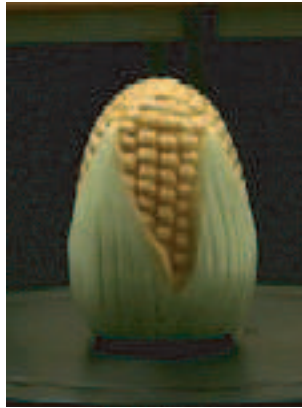
$N$  と  $m$  を変化させて本手法を適用したときの探索時間 (物体 fuku) を図 64 に示す. マージンに関しては,  $m = 13$  付近で探索時間が最短となり, 同時に処理を行うテンプレート数に関しては,  $N = 3$  あたりまで探索時間が急激に減少し, 以後, ほぼ変化がない.

表 9 拡張 AWS の実験に使用した計算機とソフトウェアの仕様

CPU	Intel XEON CPU (2.2GHz)
Memory	3GB
OS	Microsoft Windows XP Professional (SP1)
	MATLAB を使用



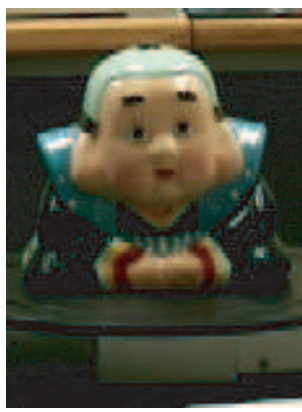
a) cat



b) cone



c) kuma



d) fuku



e) ukurere

図 63 拡張 AWS の実験のために物体の映る向きや大きさの異なるテンプレートを作成した物体

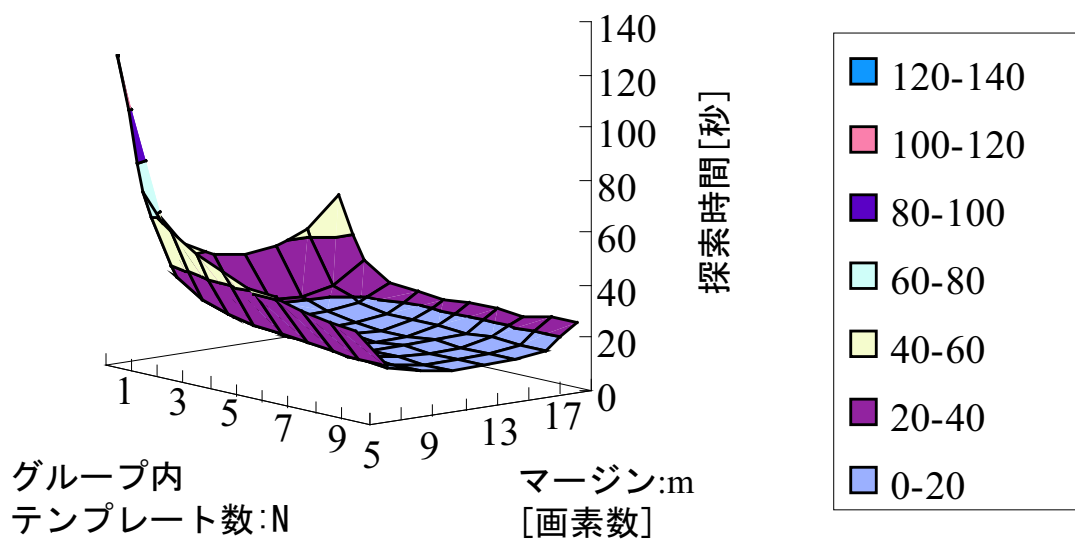


図 64 物体の向きが異なる多数のテンプレートを用いた拡張 AWS の探索時間

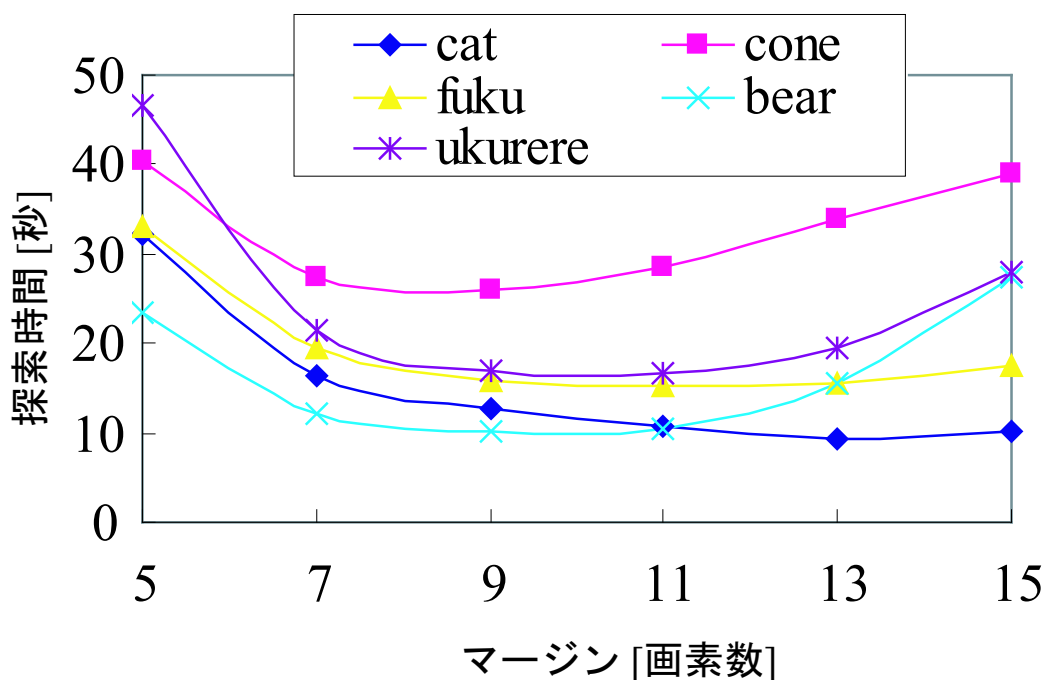


図 65 物体の向きが異なる多数のテンプレートを用いた拡張 AWS をマージン ( $m$ ) を変えて実行した探索時間

図 64 より探索時間が最短となる  $N = 11$  において  $m$  を変化させたときの探索時間を図 65 に示す.  $m$  が増加すると探索時間は急激に減少し, 9[画素] から 13[画素] 付近で最短となって, その後増加する. マージンを大きくとると, 1 回の照合で省略可能な領域が増大するもののサブリファレンス間の距離が大きくなり照合省略の可能性が減少する. 色の違いが少ない物体ほど  $m$  を大きくとることで探索時間が減少している.

同様に図 64 より探索時間が最短となる  $m = 13$  において  $N$  を変化させたときの探索時間を図 66 に示す. 探索時間は  $N = 4$  付近まで急激に減少し, 以後変化が少ない. 各物体とも同じ減少傾向を示している.

比較のため, AWS, SSDA, 部分空間法 [MN94], テンプレートマッチング法 (TM) による探索時間を測定した. AWS および SSDA は前記の閾値を用いた. AWS で

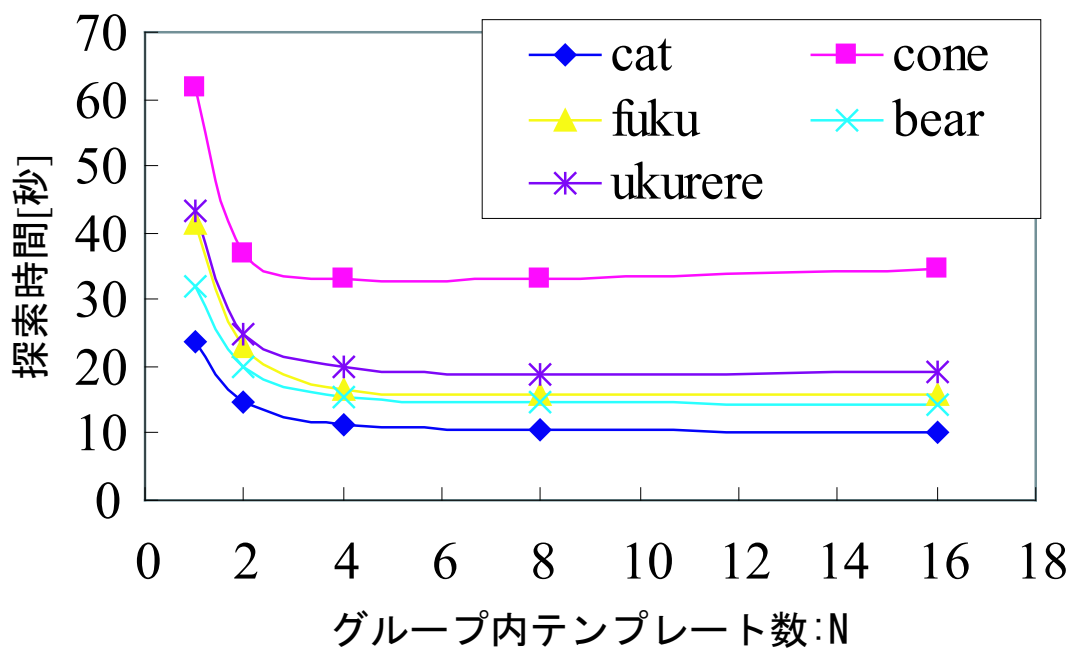


図 66 物体の向きが異なる多数のテンプレートを用いた拡張 AWS をグループサイズ (N) を変えて実行した探索時間

はマージン  $m$  を変化させて、探索時間が最短になるように  $m$  を決定した。また、部分空間法は Precision = Recall = 100% となる最小の次元数で探索時間を測定した。結果を表 10 に示す。拡張 AWS により AWS 法に比べ約 2.7 倍、部分空間法の 7 倍の速度で探索を終えている。

**スケールを変化させたテンプレートによる探索** 撮影画像から切り抜いた  $48 \times 83$  の画像を  $(15/16)^k$  に縮小したものをテンプレートとして用いる。(ただし  $k = 1, \dots, 14$ ) 入力画像には  $320 \times 240$  の撮影画像を縦横 70% に縮小したものをを用いる。入力画像のサイズ, テンプレートのサイズ, 枚数とも表 10 までの実験より小さいため, 計測される時間は短くなる。閾値は Precision = Recall = 100% となる最小の閾値を測定して 1 画素あたり 30 になるように設定した。

物体 fuku に対して拡張 AWS を適用した結果を図 67 に示す。マージンに関しては  $m = 9$  付近で探索時間が最短となり, 同時に処理を行うテンプレート数に関しては,  $N=3$  あたりまで探索時間が急激に減少し, 以後緩やかに増加していることがわかる。

図 64 より探索時間が最短となる  $N = 3$  において  $m$  を変化させたときの各物体の探索時間を図 68 に示す。  $m$  が 7 から 9 付近で各物体とも探索時間が最短となっている。

同様に探索時間が最短となる  $m = 9$  において  $N$  を変化させたときの各物体の探索時間を図 69 に示す。各物体とも  $N = 3$  付近で探索時間が最短となり, その後緩やかに増加している。

他の手法との比較のため, AWS, SSDA, テンプレートマッチング法による探索時間を測定した。部分空間法は一般にはスケール変化には適用できないため, 本

表 10 物体の向きが異なる多数のテンプレートを用いた場合の拡張 AWS と他の手法の探索時間の比較

拡張 AWS	AWS	SSDA	部分空間法	TM
16.625[s]	41.25[s]	534.2[s]	118.6[s]	8464[s]



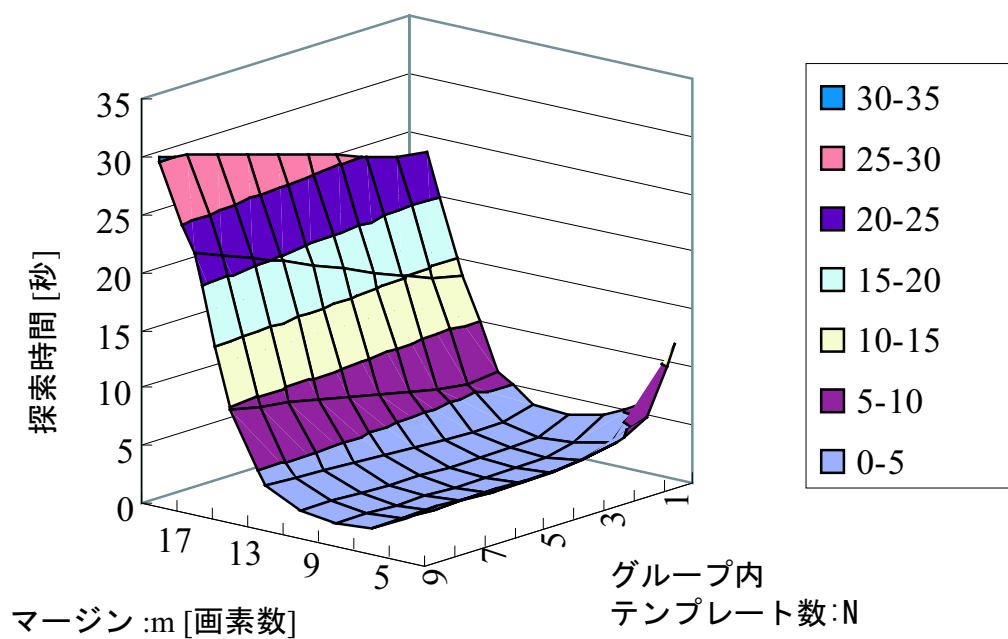


図 67 物体の映る大きさが異なる多数のテンプレートを用いた拡張 AWS の探索時間

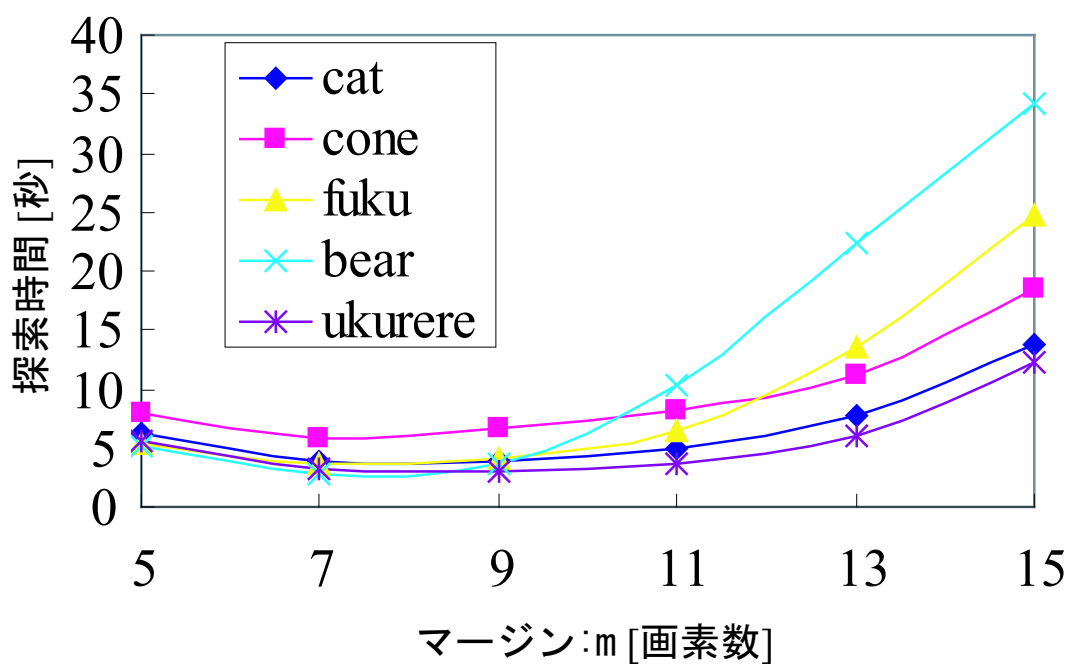


図 68 物体の映る大きさが異なる多数のテンプレートを用いた拡張 AWS をマージン (m) を変えて実行した探索時間

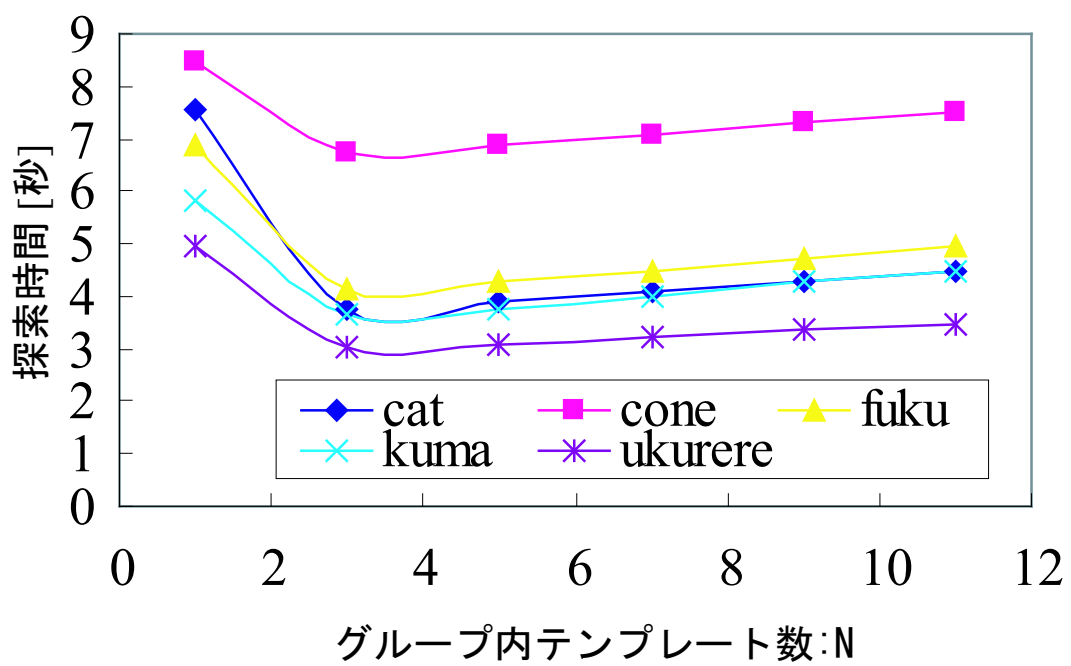


図 69 物体の映る大きさが異なる多数のテンプレートを用いた拡張 AWS をグループサイズ (N) を変えて実行した探索時間

実験の比較対象から外している。拡張 AWS, AWS および SSDA は前記の閾値を用いた。また、AWS ではマージン  $m$  を変化させて、探索時間が最短になるように  $m$  を決定した。結果を表 11 に示す。拡張 AWS 法により AWS 法に比べ約 2 倍, SSDA の 20 倍の速度であった。

### 3.3.5 考察

3.3 節では、テンプレート内のサブテンプレート間の類似性を利用して、テンプレートマッチングを高速に行う適応的ウィンドウスキッピング法 (AWS) を提案した。AWS は図 62 に示したように、 $2^n \times 2^n$  付近の画素数のテンプレートを用い、低い閾値を設定できる場合には BSPA に性能が劣る。BSPA には、テンプレートの大きさが  $2^n \times 2^n$  付近でないと極めて計算量が大きくなり、低い閾値の設定が難しい場合にも同様に計算量が大きいため、適用先が MPEG 圧縮などに限定される。一方、AWS は、BSPA では高速化できない場合にも高速化が可能であり、汎用性が高い。このため AWS は、多様なテンプレートを多様なサイズで探索する、物体探索や画像映像検索の用途には、BSPA に比べ、大きな優位性を持つと考えられる。

また、AWS は、テンプレート内部の類似性という、BSPA とは異なる性質を使った高速化手法であるため、現在 BSPA に劣る MPEG 圧縮のようなアプリケーションに対しても、BSPA と AWS を組み合わせさらに高速な探索を実現できる可能性がある。これは今後の課題である。

表 11 物体の向きが異なる多数のテンプレートを用いた場合の拡張 AWS と他の手法の探索時間の比較

拡張 AWS	AWS	SSDA	TM
1.891[s]	4.094[s]	78.87[s]	145.7[s]

### 3.4 結言

本章では、まず、3.2 節においてパン・チルト・ズームカメラを用いて実環境からの物体探索を可能とする動的アクティブ探索法を提案した。本手法は、物体の位置・姿勢の違いによって照明条件やカメラのパン・チルト・ズームパラメータが異なり、物体の見え方が様々に変わる状況でも高速に物体を検出することが可能である。本手法はヒストグラムを用いた画像探索法であるアクティブ探索法を発展させたもので、あらかじめ収集した多数のテンプレートに対して色分布をベクトル量子化 (VQ) を用いて学習することによって、その精度・速度を改善している。更に、多数のテンプレートでも高速に探索することが可能な並列探索法、併合探索法を提案した。これらをカメラの動的制御に適用しカメラの制御回数を削減した。

次に、3.3 節において一般の距離特徴を使ったテンプレートマッチングの高速化手法である適応的ウィンドウスキッピング法 (AWS) を提案した。本手法はテンプレート内のサブテンプレート間の類似性を利用して、テンプレートマッチングを高速に行う。一般の距離を用いているため、距離が定義された特徴で識別できる任意のアルゴリズムを利用でき、適用範囲が広い。AWS では、リファレンスサブテンプレートとサブウィンドウとを照合し、テンプレートとサブウィンドウを内部に含む近傍のウィンドウとの距離下限値を計算する。距離下限値が、閾値を上回る場合には、そのウィンドウでの照合を省略できる。距離下限値の計算には、サブテンプレート間の距離と、三角不等式の原理を利用する。AWS は、数多くの領域の照合を省略して探索時間を削減するにも関わらず、総当り探索と同じ精度を保証する。本手法は SSDA や SEA といった高速化手法と比べ、数倍高速である。

AWS は任意の距離尺度に対して適用できるため、L1 距離である色ヒストグラム特徴にも適用できる。しかしながら照合の省略が可能な領域がサブテンプレートのマージンの範囲内に制限されてしまうため、アクティブ探索法よりも照合回数は多くなる。一般的な距離特徴を使った場合と色ヒストグラム特徴を使った場合を比較すると、色ヒストグラム特徴は累積特徴であるため、ウィンドウの多少のずれや物体の姿勢の小さな変化などに対して距離特徴に比べ変化量が小さく照合の枝刈効率が低い。このため探索速度は動的アクティブ探索法が適応的ウィン

ドウスキッピング法よりも速い。速度・精度の観点から、あらかじめ動的アクティブ探索法において大まかな物体の位置・姿勢を求めたあと、適応的ウィンドウスキッピング法に用いて探索し、詳細な物体の位置・姿勢を求めるということも考えられる。

本章で提案した2つの手法は物体の見え方の変化に対処するためにテンプレートを増やす手法である。照明の状況や物体の形状によって必要なテンプレート数が異なる。提案手法ではどちらも必要な最小限のテンプレートを選択するのではなく、あらかじめ大量のテンプレートを用意するアプローチを取っている。提案手法では複数のテンプレート間で不要な照合を省略し、計算量への影響が限定的になるためである。しかしながらテンプレートと類似する領域では計算時間がかかるために、入力画像の数・大きさが物体領域に比べて大きくない場合は類似したテンプレートの増加が計算量の増加になる。このため、あまりにも類似するテンプレートは利用しないことが望ましい。テンプレートをいかに効率的に生成・取得するかについては今後さらに検討する必要がある。

## 4. 結論

### 4.1 まとめと考察

本論文では、画像照合による物体検出において問題となる、照明の変化や隠蔽、物体の位置・姿勢の変化に伴う物体の見え方の変化という問題を扱った。

2章では、照明の変化や隠蔽に対しても頑健に、背景モデルに基づいて移動物体領域を検出することを目的とした。従来の方法では、照明の点灯・消灯・調光の生じる環境下で、画像中の背景領域が少ない状況では正しく移動物体領域を検出することができなかった。また、物体が何らかの障害物に隠蔽された場合にも検出することができなかった。

これらの問題に対し、まず、2.2節において室内環境での照明の点灯・消灯といった照明の変化が生じる場合かつ背景の大部分が移動物体領域に占められる場合にも一定数の背景候補を随時更新して追跡を継続する手法を提案した。本手法では、前景と背景を分離できる頑健な距離特徴を導入し、前景の占める面積が大きい場合にも適切に背景差分を行って物体を検出した。しかしながら総当りでこの計算を行った場合、多大な計算量を要する。その計算量を削減するため、確率モデルに基づくサンプリングおよびその更新を行う逐次モンテカルロ法を導入し、各蓄積画像の重み係数の追跡問題として扱った。実験では、室内環境において光源色が異なる2種類の照明を使用し、その明るさを変化させ本手法を評価した。毎秒3フレーム程度の処理で、前景が50%を超える場合にも正しく移動物体領域を求めることができることを示した。

次に、2.3節では、雑音や照明変動等の外乱や隠蔽が生じる状況において、安定かつ継続的な話者の追跡を行うために、画像情報と音情報を用いた移動物体の追跡手法を提案した。円形マイクロホンアレイと全方位カメラを用いて取得した情報を用いて人物を追跡する実験を行い、周辺雑音がある環境において、隠蔽や声の途切れが生じた場合でも、継続的な追跡が実現できることを示した。

3章では、物体の位置・姿勢の変化に伴い画像に映る物体の見え方が多様になった場合についてそれら全ての見え方を用意して検出対象モデルに基づく手法により探索し、静止物体を含め目的の物体を高速に検出することを目的とした。室内

程度の環境でも、画像に映る物体の向きや大きさの変化が膨大であり、かつ撮影すべき画像枚数が膨大になるため、従来手法では現実的な時間で検出対象モデルに基づく手法を使って静止物体を検出することは困難であった。3.2節では、物体の位置・姿勢の変化に対して、多様な物体の見え方に対応し、パン・チルト・ズームカメラを用いて室内環境から物体検出を可能とする検出対象モデルに基づく手法（動的アクティブ探索法）を提案した。動的アクティブ探索法はテンプレート間の異色画素数をあらかじめ求めておきテンプレート間でも照合の省略を行って探索を高速化する。約 5m の範囲内の物体をカメラを制御して探索した場合、平均約 10 秒で検出可能であった。これは従来の探索の手法にアクティブ探索法を用い、単純なカメラ制御を行った場合に比較し 4 倍以上の速度向上に相当する。しかしながら、3.2 節で提案した手法は、物体を色ヒストグラム特徴で識別できる場合にしか適用できず、適用範囲が色が鮮やかな物体に限定される。

次に、色ヒストグラム特徴ではなく、詳細な特徴を評価して物体を見分けるため、3.3節では、一般の距離特徴を利用してテンプレートマッチングを高速に行う適応的ウィンドウスキッピング法 (AWS 法) を提案した。AWS は、数多くの領域の照合を省略して探索時間を削減するにも関わらず、総当り探索と同じ精度を保証する。本手法は SSDA や SEA といった高速化手法と比べ、数倍高速である。また AWS は、 $2^n \times 2^n$  付近の画素数のテンプレートを用い、低い閾値を設定できる場合には BSPA に性能が劣るが他の条件では優れている。このため、多様なテンプレートを多様なサイズで探索する、物体探索や画像映像検索の用途には、BSPA に比べ、大きな優位性を持つと考えられる。

従来手法と同様に本研究の 2 章に述べた手法と 3 章に述べた手法を組み合わせ、移動物体を効率的に検出することも可能である。この場合、単純な組み合わせだけでなく 2 章に述べた手法による前処理の結果、得られる照明条件のパラメータを用いて 3 章に述べた手法のテンプレートを適応的に選ぶことができ、3 章で照合すべき探索領域が絞り込めるだけでなくテンプレートの絞り込みも行え、計算時間を削減できると考えられる。この組み合わせ手法については今後の課題である。

本研究では物体検出技術に関わる照明の変化と隠蔽の課題と物体の位置・姿勢の変化の課題をそれぞれ背景モデルに基づく手法、検出対象モデルに基づく手法



において解決し、物体検出技術の適用範囲を広げた。本研究は、物体検出技術の応用の拡大に寄与すると考えられる。

## 4.2 展望と今後の課題

本論文で提案した手法を具体的なアプリケーションに組み込み新たな研究を開始している。以下に現在研究中の2つの技術について述べる。

### 擬人化エージェントによるコミュニケーション技術

2章で述べたシステムを用いた人物検出と小型の円筒型ディスプレイ [KTMT02a, KTMT02b, KTMT03, KTT+05] とを組み合わせた擬人化エージェントの開発を行っている (図 70)。円筒型ディスプレイは180度全方向から観察でき、また表示画像はリアルタイムでの書き換えが可能であることから、エージェントの顔を話者の方向に向けたり、その視線による方向指示が可能である。提案技術は、大勢の人々が集まる展示会場や繁華街での監視システムやロボットによる案内システムの他、テレビ会議システムや会議映像自動作成システム等への応用が期待される。

### 画像データベースからの部分画像検索技術

テンプレートマッチングとデータベース検索を組み合わせるアプローチ [MWL01, MWH02] と3章で述べた AWS 法を組み合わせ、膨大な画像データベースから、部分画像検索を少ないインデックスサイズで高速に行う手法の開発を進めている [KKK04a, KKK04b, KKK04c, KKK05b, KKOK05]。

高速な検索を実現するためには、ある種のインデックス構造を導入することが必要である。しかし、上記部分画像検索においては、データベース中の各画像から膨大な量の部分領域を抽出してインデックス構造に投入する必要があるため、インデックスの保持に必要な記憶容量が膨大になるという問題点がある。そこで、データベース中の画像について、部分領域を間引いて抽出することにより、インデックスサイズを大幅に削減することを検討する。部分領域の抽出間隔を適切に設定することにより、インデックスを全く削減しない総当たり照合と全く同一の検索結果を理論的に保証することが可能である。インデックスとして得られた領

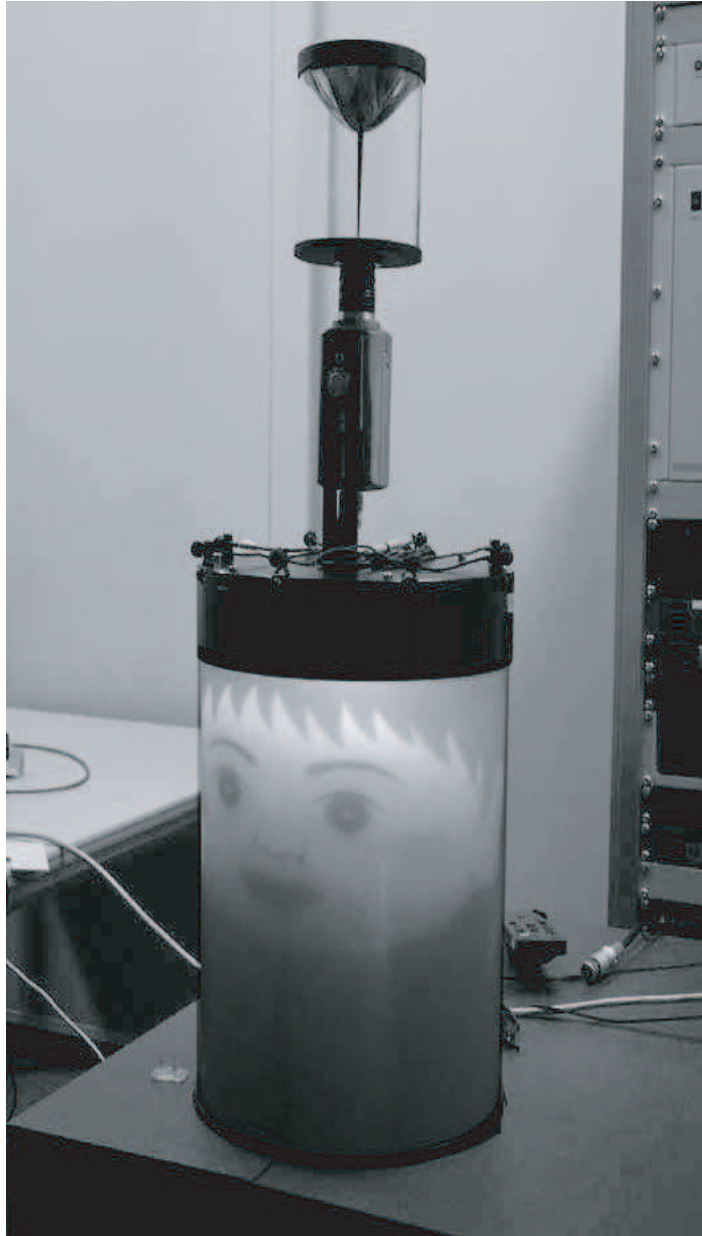


図 70 擬人化エージェントと円筒ディスプレイ

域と周辺領域の検索には, AWS を用いることができるため, 周辺領域の検索時間が極めて小さく, インデックスを削減しない場合と比べインデックス検索時間削減効果が, 周辺領域の検索時間の増加を上回る場合もあり, 非常に効率的である. 今後これらの研究を推し進め, 超膨大な画像データベースから所望の画像を高速に検索する技術の開発を目指す.

## 謝辞

本研究の全過程を通して、懇切なる御指導、御鞭撻を賜った視覚情報メディア講座 横矢 直和教授、山澤 一誠助教授に心より深謝致します。

本研究の遂行にあたり、有益な御助言と御鞭撻を賜った像情報処理学講座 千原 國宏教授、知能情報処理学講座 木戸出 正継 教授に厚く御礼申し上げます。

本研究を進めるにあたり、終始、暖かい御指導をして頂いた NTT コミュニケーション科学基礎研究所 村瀬洋氏（現、名古屋大学大学院情報科学科教授）、高木 茂氏（現、国立沖縄工業高等専門学校情報通信システム工学科教授）に深く感謝致します。

本研究の遂行において多大なる御助言、御鞭撻を賜り、終始御指導を頂いた NTT コミュニケーション科学基礎研究所メディア認識研究グループ 大和淳司氏、柏野 邦夫氏に心から感謝致します。

また、日頃から有益な御意見を頂き御指導して頂いた、NTT コミュニケーション科学基礎研究所メディア認識研究グループ 土田勝氏（現 NTT データ）、黒住 隆行氏、木村昭悟氏、ならびに、本研究の実験・データ計測において、多大な協力を頂いた NTT コミュニケーション科学基礎研究所メディア認識研究グループ事務補佐員 荒井七恵女史、市橋淑江女史、NTT コミュニケーション科学基礎研究所メディア認識研究グループおよび本学視覚情報メディア講座の諸氏に深く御礼申し上げます。

日頃より暖かく支えて下さった視覚情報メディア講座 事務補佐員 守屋 知代女史、元ソフトウェア基礎講座 事務補佐員 福永 博美女史、に厚く感謝致します。

最後に、普段より公私にわたり支えてくれた妻右子に厚く感謝する。

## 参考文献

- [AF98] 天本直弘, 藤井明宏, “画像処理技術による障害物検出と移動物体追跡法”, 電子情報通信学会論文誌 (A), 1998.
- [Aka97] 赤松茂, “コンピュータによる顔の認識—サーベイ—”, 電子情報通信学会論文誌 (D-II), J80-D-II, 8, pp. 2031-2046, Aug. 1997.
- [ASF<sup>+</sup>98] Y. Abe, M. Shikano, T. Fukuda, F. Arai and Y. Tanaka, “Vision based navigation system by variable template matching for autonomous mobile robot,” *Proc. of the IEEE International Conference on Robotics and Automation (ICRA '98)*, pp. 952-957, 1998.
- [BAJ02] M. J. Beal, H. Attias and N. Jojic, “Audio-video sensor fusion with probabilistic graphical models,” *Proc. of the European Conference on Computer Vision*, pp. 736-752, 2002.
- [BJ01] R. Basri and D. Jacobs, “Lambertian reflectance and linear subspaces,” *Proc. of the 8th IEEE International Conference on Computer Vision (ICCV01)*, pp. 383-390, 2001.
- [BS72] D. I. Barnea and H. F. Silverman, “A class of algorithms for fast digital image registration,” *IEEE Transactions on Computers*, C-21, 2, pp. 179-186, Feb. 1972.
- [FDNI85] 深田陽司, 土井博, 永峰啓二, 稲荷隆彦, “山積みされたクランクシャフトの認識”, 三菱電機技報, 59, 10, pp. 30-33, 1985.
- [GA01] M. Gharavi-Alkhansari, “A fast globally optimal algorithm for template matching using low-resolution pruning,” *IEEE Transactions on Image Processing*, 10, 4, pp. 526-533, 2001.
- [Hal94] P. Hallinan, “A low-dimensional representation of human faces for arbitrary lighting conditions,” *Proc. of the 1994 IEEE Computer*

*Society Conference of Computer Vision and Pattern Recognition (CVPR94)*, pp. 995-999, 1994.

- [HH93] C. M. Huang and R. W. Harris, "A comparison of several vector quantization codebook generation approaches," *IEEE Transactions on Image Processing*, 2, 1, pp. 108-112, 1993.
- [HK03] 長谷川修, 金出武雄, "一般道路映像中の移動物体の識別・色の推定と特定対象の検出", *情報処理学会論文誌*, 44, 7, pp. 1795-1807, 2003.
- [HK05] O. Hasegawa and T. Kanade, "Type classification, color estimation, and specific target detection of moving targets on public streets," *Machine Vision Applications*, 16, 2, pp. 116-121, 2005.
- [HKIK85] 長谷部淳, 加藤良平, 伊藤徳一, 菊池敦, "マルチプロセッサ型ビデオ画像処理システム sips", *情報処理学会研究報告*, CV39-5, pp. 1-8, 1985.
- [HMK03] A. Hatabu, T. Miyazaki and I. Kuroda, "Optimization of decision-timing for early termination of SSDA-based block matching," *Proc. of the 2003 IEEE International Conference on Multimedia and Expo. (ICME2003)*, 2, pp. 821-824, 2003.
- [HN83] 原靖彦, 中川清, "局所的特徴の抽出と比較による回路パターンの欠陥検出法", *電子通信学会論文誌 (D)*, 66-D, 8, pp. 970-976, 1983.
- [Hor93] B. K. P. Horn, *ロボットビジョン. 機械は世界をどうみるか.*, 第13章, pp. 329-368. 浅倉書店, 1993.
- [HSY04a] T. Haga, K. Sumi and Y. Yagi, "Human detection in outdoor scene using spatio-temporal motion analysis," *Proc. of the 17th IAPR International Conference of Pattern Recognition (ICPR04)*, IV, pp. 331-334, 2004.

- [HSY04b] 羽下哲司, 鷺見和彦, 八木康史, “変化領域内の動きの時空間特徴に着目した屋外情景における歩行者の検出”, 電子情報通信学会論文誌 (D-II), J87-D-II, 5, pp. 1104-1111, 2004.
- [HT00] 早坂光晴, 富永英義, “動画像からの背景画像生成を用いた移動物体抽出法に関する一検討”, 情報処理学会研究報告, AVM29-1, pp. 1-6, 2000.
- [HWM99] 波部齊, 和田俊和, 松山隆司, “照明変化に対して頑健な背景差分法”, 情報処理学会研究報告, CVIM115-3, pp. 17-24, 1999.
- [IB98a] M. Isard and A. Blake, “Condensation - conditional density propagation for visual tracking,” *International Journal on Computer Vision*, 29, 1, pp. 5-28, 1998.
- [IB98b] M. Isard and A. Blake, “Icondensation: Unifying low-level and high-level tracking in a stochastic framework,” *Proc. of the European Conference on Computer Vision*, pp. 893-908, 1998.
- [ILYH05] 岩井儀雄, S. Lao, 山口修, 平山高嗣, “画像処理による顔検出と顔認識”, 情報処理学会研究報告 (CVIM), 38, pp. 343-368, May 2005.
- [IMG] “Test image database in simplicity paper,”  
<http://wang.ist.psu.edu/IMAGE/>.
- [Ish97] H. Ishiguro, “Distributed vision system: A perceptual information infrastructure for robot navigation,” *Proc. of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, pp. 36-41, 1997.
- [JC98] G. Y. Jiang and T. Y. Choi, “Robust detection of landmarks in color image based on fuzzy set theory,” *Proc. of the 4th International Conference on Signal Processing (ICSP'98)*, pp. 968-971, 1998.

- [KCL<sup>+</sup>98] T. Kanade, R. Collins, A. Lipton, P. Burt and L. Wixson, “Advances in cooperative multi-sensor video surveillance,” *DARPA Image Understanding Workshop*, 1, pp. 3-24, 1998.
- [KFH05] 加藤丈和, 深尾隆則, 羽下哲司, “対象追跡-フレーム間の類似度に着目した手法から動きのモデルに着目した手法まで-”, 情報処理学会研究報告 (CVIM), 88, pp. 185-198, Sep. 2005.
- [KKK04a] A. Kimura, T. Kawanishi and K. Kashino, “Acceleration of similarity-based partial image retrieval using multistage vector quantization,” *Proc. of the 17th IAPR International Conference on Pattern Recognition (ICPR 2004)*, 2, pp. 993-996, Aug. 2004.
- [KKK04b] A. Kimura, T. Kawanishi and K. Kashino, “Similarity-based partial image retrieval guaranteeing same accuracy as exhaustive matching,” *Proc. of the 2004 IEEE International Conference on Multimedia & Expo. (ICME 2004)*, 3, Jun. 2004.
- [KKK04c] 木村昭悟, 川西隆仁, 柏野邦夫, “SPIRE: 総当たり照合と同一の精度を保証する類似部分画像検索”, 画像の認識・理解シンポジウム (MIRU2004) 予稿集, 2, pp. 400-404, Jul. 2004.
- [KKK<sup>+</sup>05a] 川西隆仁, 久野和樹, 木村昭悟, 黒住隆行, 柏野邦夫, 高木茂, “サブテンプレート間距離を用いた適応的ウィンドウスキップによる高速テンプレートマッチング法”, 電子情報通信学会論文誌 (D-II), J88-D-II, 8, pp. 1389-1397, Aug. 2005.
- [KKK05b] 木村昭悟, 川西隆仁, 柏野邦夫, “SPIRE: 総当たり照合と同一の精度を保障する類似部分画像検索”, 電子情報通信学会論文誌 (D-II), J88-D-II, 8, pp. 1712-1719, Aug. 2005.
- [KKM00a] 柏野邦夫, 黒住隆行, 村瀬洋, “ヒストグラム特徴を用いた音や映像の高速 and/or 探索法”, 電子情報通信学会論文誌 (D-II), J83-D-II, 12, 2000.



- [KKM00b] 黒住隆行, 柏野邦夫, 村瀬洋, “時系列アクティブ探索法のための特徴変動に頑健な確率ディザボータリング”, 画像の認識・理解シンポジウム (MIRU2000) 講演論文集, I, pp. 313-318, 2000.
- [KKOK05] 木村昭悟, 川西隆仁, 大塚和弘, 柏野邦夫, “重み付き特徴点照合に基づく高速画像検索”, 電子情報通信学会技術研究報告, PRMU2005-24, DE2005-2, pp. 7-12, Jun. 2005.
- [KKT04] 久野和樹, 川西隆仁, 高木茂, “スキッピングテンプレートマッチングによる3次元物体の高速探索”, 画像の認識・理解シンポジウム (MIRU2004) 予稿集, 2, pp. 435-440, Jul. 2004.
- [KM00] 川西隆仁, 村瀬洋, “アクティブ探索法のパン・チルト・ズームカメラを用いた物体探索への適用”, 画像の認識・理解シンポジウム (MIRU2000) 講演論文集, II, pp. 31-36, Sep. 2000.
- [KM01] 川西隆仁, 村瀬洋, “色ヒストグラムとパン・チルト・ズームカメラを用いた高速物体探索法—動的アクティブ探索法—”, 電子情報通信学会論文誌 (D-II), J84-DII, 8, pp. 1772-1730, Aug. 2001.
- [KMT02] T. Kawanishi, H. Murase and S. Takagi, “Quick 3d object detection and localization by dynamic active search with multiple active cameras,” *Proc. of the 16th IAPR International Conference on Pattern Recognition (ICPR2002)*, 2, pp. 603-606, Aug. 2002.
- [KMTW01a] T. Kawanishi, H. Murase, S. Takagi and M. Werner, “Dynamic Active Search for quick object detection with pan-tilt-zoom camera,” *Proc. of the 2001 IEEE International Conference on Image Processing (ICIP2001)*, 3, pp. 716-719, Oct. 2001.
- [KMTW01b] 川西隆仁, 村瀬洋, 高木茂, Martin Werner, “動的アクティブ探索法と複数の能動カメラを用いた物体の位置推定”, 電子情報通信学会技術研究報告, PRMU2001-81, pp. 125-132, Jul. 2001.

- [KNMU73] 河村貞和, 中谷敏雄, 目瀬道弘, 烏野武, “傷検出のパターン認識とそのディスプレイについて”, 通信学会研究会資料, IE73-3, 1973.
- [KSM99] 柏野邦夫, ガビンスミス, 村瀬洋, “ヒストグラム特徴を用いた高速探索法—時系列アクティブ探索法—”, 電子情報通信学会論文誌 (D-II), J82-D-II, 9, pp. 1365-1374, 1999.
- [KTMT02a] 川西隆仁, 土田勝, 村瀬洋, 高木茂, “擬人化エージェントのための小型円筒ディスプレイとその応用”, 電子情報通信学会技術研究報告, PRMU2002-52, IE2002-38, MVE2002-30, pp. 45-50, Jul. 2002.
- [KTMT02b] 川西隆仁, 土田勝, 村瀬洋, 高木茂, “擬人化エージェント表示のための小型円筒ディスプレイ”, 情報技術レターズ, 1, pp. 131-132, Sep. 2002.
- [KTMT03] T. Kawanishi, M. Tsuchida, H. Murase and S. Takagi, “Small cylindrical display for anthropomorphic agents,” *Proc. of the 2003 IEEE International Conference on Multimedia and Expo. (ICME2003)*, II, pp. 85-88, Jul. 2003.
- [KTT<sup>+</sup>05] T. Kawanishi, M. Tsuchida, S. Takagi, A. Kimura and J. Yamato, “Small cylindrical display using aspherical mirror for anthropomorphic agents,” *Proc. of the International Display Workshop / Asia Display (IDW/AD'05)*, pp. 1755-1758, Dec. 2005.
- [LC97] C. H. Lee and L. H. Chen, “A fast motion estimation algorithm based on the block sum pyramid,” *IEEE Transactions on Image Processing*, 6, 11, pp. 1587-1590, 1997.
- [LS95] W. Li and E. Salari, “Successive elimination algorithm for motion estimation,” *IEEE Transactions on Image Processing*, 4, 1, pp. 105-107, 1995.

- [McL93] G. F. McLean, "Vector quantization for texture classification," *IEEE Transactions on Systems, Man and Cybernetics*, 23, 3, pp. 637-649, 1993.
- [MKN97] 松尾直司, 北側博紀, 長田茂美, "音声情報と画像情報の統合による話者位置検出システム", 第13回ヒューマンインタフェースシンポジウム論文集, pp. 469-474, 1997.
- [MN94] 村瀬洋, Shree K. Nayar, "2次元照合による3次元物体認識-パラメトリック固有空間法-", 電子情報通信学会論文誌 (D-II), J77-D-II, 11, pp. 2179-2187, 1994.
- [MOKS00] 村松彰二, 大塚裕史, 小林芳樹, 清水英志, "テンプレートマッチングの高速化戦略とgaによる最適化", 電子情報通信学会論文誌 (D-II), J83-D-II, 6, pp. 1487-1497, Jun. 2000.
- [MV98] 村瀬洋, V. V. Vinod, "ヒストグラム特徴を用いた高速物体探索法—アクティブ探索法—", 電子情報通信学会論文誌 (D-II), J81-D-II, 9, pp. 2035-2042, 1998.
- [MWH02] Y. Moon, K. Whang and W. Han, "Generalmatch: A subsequence matching method in time-series databases based on generalized windows," *Proc. of the 2002 ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2002)*, pp. 382-393, 2002.
- [MWHT01] 松山隆司, 和田俊和, 波部齊, 棚橋和也, "照明変化に頑健な背景差分法", 電子情報通信学会論文誌 (D-II), J84-D-II, 10, pp. 2201-2211, 2001.
- [MWL01] Y. Moon, K. Whang and W. Loh, "Duality-based subsequence matching in time-series databases," *Proc. of the 17th IEEE International Conference on Data Engineering (ICDE 2001)*, pp. 263-272, 2001.

- [NK88] N. M. Nasrabadi and R. A. King, "Image coding using vector quantization: a review," *IEEE Transactions on Communications*, 36, 8, pp. 957-971, 1988.
- [NMF<sup>+</sup>96] 長屋茂喜, 宮武孝文, 藤田武洋, 伊藤渡, 上田博唯, "時間相関型背景判定法による移動物体検出", 電子情報通信学会論文誌 (D-II), J79-D-II, 4, pp. 586-576, 1996.
- [NNL<sup>+</sup>02] T. Nishiura, M. Nakamura, A. Lee, H. Saruwatari and K. Shikano, "Talker tracking display on autonomous mobile robot with a moving microphone array," *Proc. of the 8th International Conference on Auditory Display*, pp. 244-247, 2002.
- [NYNS00] T. Nishiura, T. Yamada, S. Nakamura and K. Shikano, "Localization of multiple sound sources based on a csp analysis with a microphone array," *Proc. of the 2000 IEEE International Conference on Acoustics, Speech and Singnal Processing (ICASP 2000)*, SP-L4-6, pp. 1053-1056, 2000.
- [Ono87] 尾上守夫 (編), 画像処理ハンドブック, 昭晃堂, 1987.
- [ORP00] N. Oliver, B. Rosario and A. Pentland, "A bayesian computer vision system for modeling human interactions," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22, 8, pp. 831-843, 2000.
- [SC96] B. Schiele and J. L. Crowley, "Probabilistic object recognition using multidimensional receptive field histograms," *Proc. of the 13th IAPR International Conference on Pattern Recognition (ICPR'96)*, pp. 50-54, 1996.
- [Sha92] A. Shashua, *Geometry and Photometry in 3D Visual Recognition* PhD thesis, Dept. Brain and Cognitive Science, MIT, 1992.

- [Sha97] A. Shashua, "On photometric issues in 3d visual recognition from a single 2d image," *International Journal of Computer Vision*, 21, 1/2, pp. 99-122, 1997.
- [SI71] 白井良明, 井上博允, "ビジュアル・フィードバックを利用したロボットの組合せ作業", 電総研彙報, 35, 3, pp. 327-333, 1971.
- [SRS01] S. Spors, R. Rabenstein and N. Strobel, "Joint audio-video object tracking," *Proc. of the 2001 IEEE International Conference on Image Processing*, pp. 393-396, 2001.
- [SSH05] 鷺見和彦, 関真規人, 波部斉, "物体検出 - 背景と検出対象のモデリング -", 情報処理学会研究報告 (CVIM), 88, pp. 79-98, Sep. 2005.
- [THKS88] C. Thorpe, M. Hebert, T. Kanade and Shafer S., "Vision and navigation for the carnegie-mellon navlab," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10, 3, pp. 362-373, 1988.
- [TI77] 壺井芳昭, 井上驥, "視覚付ロボットによる物体の3次元位置方向測定の一方法", 通信学会論文誌, 60-D, 8, pp. 586-593, 1977.
- [TKK97] 高藤政雄, 北村忠明, 小林芳樹, "空間微分および差分処理を用いた車両抽出法", 電子情報通信学会論文誌 (D-II), J80-D-II, 11, pp. 2976-2985, 1997.
- [TKMT03a] M. Tsuchida, T. Kawanishi, H. Murase and S. Takagi, "Sequential monte-carlo estimation of background image for background subtraction under changing illumination," *Proc. of the 3rd IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP2003)*, pp. 421-425, Sep. 2003.
- [TKMT03b] 土田勝, 川西隆仁, 村瀬洋, 高木茂, "人物検出を目的とした照明変化に応じた背景画像の生成", 電子情報通信学会技術研究報告, PRMU2002-176, pp. 43-48, Jan. 2003.

- [TKMT04a] M. Tsuchida, T. Kawanishi, H. Murase and S. Takagi, “Joint audio-visual tracking based on dynamic-weighted linear combination of probability state density,” *Journal of Advanced Computational Intelligence & Intelligent Informatics (JACII3)*, 8, 2, pp. 190-199, Mar. 2004.
- [TKMT04b] 土田勝, 川西隆仁, 村瀬洋, 高木茂, “背景差分法による物体検出を目的とした逐次モンテカルロ法による背景推定”, 電子情報通信学会論文誌 (D-II), J87-D-II, 5, pp. 1062-1070, May 2004.
- [TSOO00] 田中弘一, 佐野睦夫, 大原秀一, 奥平雅士, “パラメトリックテンプレート法とその高精度ロバストマッチングへの応用”, 電子情報通信学会論文誌 (D-II), J83-D-II, 4, pp. 1119-1130, 2000.
- [VR77] G. J. Vanderbrug and A. Rosenfeld, “Two-stage template matching,” *IEEE Transactions on Computers*, C-26, 4, pp. 384-393, 1977.
- [WADP97] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, “Pfinder: Real-time tracking of the human body,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 7, pp. 780-785, 1997.
- [WLW01] J. Z. Whang, J. Li and G. Wiederhold, “Simplicity: semantics-sensitive integrated matching for picture libraries,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 9, pp. 947-963, 2001.
- [WM96] T. Wada and T. Matsuyama, “Appearance Sphere: Background model for pan-tilt-zoom camera,” *Proc. of the 13th IAPR International Conference on Pattern Recognition (ICPR'96)*, pp. 718-722, 1996.
- [WTVF93] H. P. William, Saul A. T., William T. V. and Brian P. F., Numerical recipes in C (日本語版), 第2章. 技術評論社, 1993.

- [YIE74] 依田晴夫, 池田貞弘, 江尻正員, “回転形パターンマッチングによる形状選別の試み”, 計測自動制御学会論文誌, 10, 3, p.284, 1974.
- [YYY95] K. Yamazawa, Y. Yagi and M. Yachida, “Obstacle detection with omnidirectional image sensor hyperomnivision,” *Proc. of the 1995 IEEE International Conference on Robotics and Automation (ICRA'95)*, pp. 1062-1067, 1995.
- [ZDD02] D. N. Zotkin, R. Duraiswami and L. S. Davis, “Joint audio-visual tracking using particle filters,” *EURASIP Journal of Applied Signal Process*, 2002(11), pp. 1154-1164, 2002.
- [ZF01] T. Zhao and L. Fengjun, “Segmentation and tracking of multiple humans in complex situations,” *Proc. of the 2001 IEEE International Conference on Computer Vision and Pattern Recognition*, 2, pp. 194-201, 2001.

## 研究業績

### 学術論文誌

1. 川西 隆仁, 村瀬 洋, “色ヒストグラムとパン・チルト・ズームカメラを用いた高速物体探索法 -動的アクティブ探索法-”, 電子情報通信学会論文誌 (D-II), J84-D-II, 8, pp. 1722-1730, Aug. 2001. (3章に関連)
2. 土田 勝, 川西 隆仁, 村瀬 洋, 高木 茂, “背景差分法による物体検出を目的とした逐次モンテカルロ法による背景推定”, 電子情報通信学会論文誌 (D-II), J87-D-II, 5, pp. 1062-1070, May 2004. (2章に関連)
3. M. Tsuchida, T. Kawanishi, H. Murase and S. Takagi, “Joint audio-visual tracking based on dynamic-weighted linear combination of probability state density,” Journal of Advanced Computational Intelligence & Intelligent Informatics (JACI3), 8, 2, pp. 190-199, Mar. 2004. (2章に関連)
4. 川西 隆仁, 久野 和樹, 木村 昭悟, 黒住 隆行, 柏野 邦夫, 高木 茂, “サブテンプレート間距離を用いた適応的ウィンドウスキップによる高速テンプレートマッチング法”, 電子情報通信学会論文誌 (D-II), J88-D-II, 8, pp. 1389-1397, Aug. 2005. (3章に関連)
5. 木村 昭悟, 川西 隆仁, 柏野 邦夫, “SPIRE: 総当り照合と同一の精度を保障する類似部分画像検索”, 電子情報通信学会論文誌 (D-II), J88-D-II, 8, pp. 1712-1719, Aug. 2005.

### レター

1. 川西 隆仁, 土田 勝, 村瀬 洋, 高木 茂, “擬人化エージェント表示のための小型円筒ディスプレイ”, 情報技術レターズ, 1, pp. 131-132, Sep. 2002.
2. 川西 隆仁, 黒住 隆行, 柏野 邦夫, 高木 茂, “参照画像内の部分領域間の残差情報を利用した高速テンプレート照合法”, 情報技術レターズ, 2, pp. 175-176, Sep. 2003. (3章に関連)



3. 土田 勝, 川西 隆仁, 高木 茂, 柏野 邦夫, “複数の光源色情報を用いた鏡面反射成分の除去”, 情報技術レターズ, 2, pp. 159-160, Sep. 2003.

## 解説記事・報告

1. 川西 隆仁, 村瀬 洋, 高木 茂, “動的アクティブ探索法 室内中の物体をすばやく探す技術”, 画像ラボ, 13, 7, pp. 47-51, Jul. 2002. (3章に関連)
2. 蔵田 武志, 岩井 儀雄, 川西 隆仁, 北原 格, 北本 朝展, 孟 洋, 斎藤 英雄, “PRMU 若手プログラム 2002 報告”, 電子情報通信学会技術研究報告, PRMU2002-134, pp. 31-36, Dec. 2002.
3. 川西 隆仁, 田辺 暁弘, 松井 健一, 深津 真二, 堀越 崇, 森永 徹, 伊藤 康之, 中岡 聡, “IP 映像配信サービス「4th MEDIA」”, 映像情報メディア学会誌, 60, 6, pp. 847-852, Jun. 2006.

## 国際会議

1. T. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura and N. Yokoya, “Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors,” Proc. of the 14th IAPR International Conference on Pattern Recognition (ICPR'98), I, pp. 485-489, Aug. 1998.
2. T. Kawanishi, H. Murase, S. Takagi and M. Werner, “Dynamic active search for quick object detection with pan-tilt-zoom camera,” Proc. of the 2001 IEEE International Conference on Image Processing (ICIP2001), 3, pp. 716-719, Oct. 2001. (3章に関連)
3. T. Kawanishi, H. Murase and S. Takagi, “Quick 3D object detection and localization by dynamic active search with multiple active cameras,” Proc. of the 16th IAPR International Conference on Pattern Recognition (ICPR2002), 2, pp. 603-606, Aug. 2002. (3章に関連)

4. T. Kawanishi, M. Tsuchida, H. Murase and S. Takagi, "Small cylindrical display for anthropomorphic agents," Proc. of the 2003 IEEE International Conference on Multimedia & Expo. (ICME2003), II, pp. 85-88, Jul. 2003.
5. M. Tsuchida, T. Kawanishi, H. Murase and S. Takagi, "Sequential monte-carlo estimation of background image for background subtraction under changing illumination," Proc. of the 3rd IASTED International Conference on Visualization, Imaging, and Image Processing, pp. 421-425, Sep. 2003. (2章に関連)
6. T. Kawanishi, T. Kurozumi, S. Takagi and K. Kashino, "Skipping template matching guaranteeing same accuracy with exhaustive search," Proc. of the 5th International Conference Advances on Pattern Recognition (ICAPR 2003), pp. 209-212, Dec. 2003. (3章に関連)
7. M. Tsuchida, T. Kawanishi and S. Takagi, "Synthesizing an image invariant to illumination geometry when the illumination spectrum cannot be measured," Proc. of the IS&T's 2nd European Conference on Color in Graphics, Imaging and Vision (CGIV 2004), pp225-228, Apr. 2004.
8. B. Raytchev, H. Murase and T. Kawanishi, "VQ-faces: Unsupervised multi-view face recognition by pairwise clustering," Proc. of the 6th Asian Conference on Computer Vision (ACCV2004), 1, pp. 79-84, Jan. 2004.
9. A. Kimura, T. Kawanishi and K. Kashino, "Similarity-based partial image retrieval guaranteeing same accuracy as exhaustive matching," Proc. of the 2004 IEEE International Conference on Multimedia & Expo. (ICME2004), 3, pp. 1895-1898, Jun. 2004.
10. A. Kimura, T. Kawanishi and K. Kashino, "Acceleration of similarity-based partial image retrieval using multistage vector quantization," Proc. of the 17th IAPR International Conference on Pattern Recognition (ICPR2004), 2, pp. 993-996, Aug. 2004.

11. T. Kawanishi, T. Kurozumi, K. Kashino and S. Takagi, "A fast template matching algorithm with adaptive skipping using inner-subtemplates' Distances," Proc. of the 17th IAPR International Conference on Pattern Recognition (ICPR2004), 3, pp. 654-657, Aug. 2004. (3章に関連)
12. T. Kawanishi, M. Tsuchida, S. Takagi, A. Kimura and J. Yamato, "Small cylindrical display using aspherical mirror for anthropomorphic agents," Proc. of the International Display Workshop / Asia Display (IDW/AD'05), pp. 1755-1758, Dec. 2005.

## 研究会・シンポジウム

1. 川西 隆仁, 山澤 一誠, 岩佐 英彦, 竹村 治雄, 横矢 直和, "六角錐ミラーを用いた全方位ステレオ画像センサによる高解像度ステレオパノラマ画像の作成", 電子情報通信学会技術研究報告, PRMU97-118, pp. 51-58, Oct. 1997.
2. 川西 隆仁, 山澤 一誠, 岩佐 英彦, 竹村 治雄, 横矢 直和, "六角錐ミラーを用いた全方位画像センサによる全周ステレオパノラマ動画像の作成", 画像の認識・理解シンポジウム (MIRU'98) 講演論文集, I, pp. 177-182, Jul. 2001.
3. 横矢 直和, 竹村 治雄, 山澤 一誠, 尾上 良雄, 川西 隆仁, "全方位ビデオストリームからの視線追従型透視投影画像生成によるテレプレゼンス", 画像の認識・理解シンポジウム (MIRU'98) 講演論文集, I, pp. 27-32, Jul. 1998.
4. 山澤 一誠, 尾上 良雄, 川西 隆仁, 竹村 治雄, 横矢 直和, "全方位画像センサを用いたテレプレゼンス", 映像メディア処理シンポジウム (IMPS'98) 資料, pp. 107-108, Oct. 1998.
5. 川西 隆仁, 村瀬 洋, "アクティブ探索法のパン・チルト・ズームカメラを用いた物体探索への適用" 画像の認識・理解シンポジウム (MIRU2000) 講演論文集, 2, pp. 31-36, Jul. 2000. (3章に関連)
6. 川西 隆仁, 村瀬 洋, 高木 茂, Martin Werner, "動的アクティブ探索法と複数の能動カメラを用いた物体の位置推定", 電子情報通信学会技術研究報告,

PRMU2001-81, IE2001-61, MVE2001-60, pp. 125-132, Jul. 2001. (3章に関連)

7. 川西 隆仁, 土田 勝, 村瀬 洋, 高木 茂, “擬人化エージェントのための小型円筒ディスプレイとその応用”, 電子情報通信学会技術研究報告, PRMU2002-52, IE2002-38, MVE2002-30, pp. 45-50, Jul. 2002.
8. 土田 勝, 川西 隆仁, 村瀬 洋, 高木 茂, “人物検出を目的とした照明変化に応じた背景画像の生成”, 電子情報通信学会技術研究報告, PRMU2002-176, pp. 43-48, Jan. 2003. (2章に関連)
9. 川西 隆仁, 黒住 隆行, 柏野 邦夫, 高木 茂, “サブテンプレート間距離を用いた適応的スキップによる高速テンプレートマッチング法 — スキッピングテンプレートマッチング法 —”, 画像の認識・理解シンポジウム (MIRU2004) 予稿集, 2, pp. 19-23, Jul. 2004. (3章に関連)
10. 木村 昭悟, 川西 隆仁, 柏野 邦夫, “SPIRE, 総当たり照合と同一の精度を保証する類似部分画像検索”, 画像の認識・理解シンポジウム (MIRU2004) 予稿集, 2, pp. 400-404, Jul. 2004.
11. 久野 和樹, 川西 隆仁, 高木 茂, “スキッピングテンプレートマッチングによる3次元物体の高速探索”, 画像の認識・理解シンポジウム (MIRU2004) 予稿集, 2, pp. 435-440, Jul. 2004. (3章に関連)
12. 木村 昭悟, 川西 隆仁, 大塚 和弘, 柏野 邦夫, “重み付き特徴点照合に基づく高速画像検索”, 電子情報通信学会技術研究報告, PRMU2005-24, DE2005-2, pp. 7-12, Jun. 2005.

## 全国大会

1. 川西 隆仁, 山澤 一誠, 岩佐 英彦, 竹村 治雄, 横矢 直和, “六角錐ミラーを用いた全方位画像センサにおけるレンズ中心の位置合わせ”, 電子情報通信学会ソサイエティ大会, D-12-87, Sep. 1997.

2. 川西 隆仁, 前田 英作, 村瀬 洋, “色ヒストグラムを用いた実環境物体探索のための一検討”, 電子情報通信学会全国大会春季, D-12-3, p.176, Mar. 1999. (3章に関連)
3. 川西 隆仁, 村瀬 洋, “パン・チルト・ズームカメラによる物体探索のための最適撮影間隔の決定法”, 電子情報通信学会全国大会春季, D-12-105, p.275, Mar. 2000. (3章に関連)
4. 土田 勝, 川西 隆仁, 村瀬 洋, 高木 茂, “全方位カメラと複数のマイクロホンを用いた話者の検出”, FIT2002 一般講演論文集, 3, pp. 77-78, Sep. 2002. (2章に関連)

## 特許

1. 川西 隆仁, 村瀬 洋, “物体検出方法および装置およびこの方法を記録した記録媒体”, 特許公開 2001-224014.
2. 川西 隆仁, 村瀬 洋, 高木 茂, “物体検出/位置測定方法およびその装置、ならびにこの方法の実行プログラムおよびその記録媒体”, 特許公開 2003-022442.
3. 高木 茂, 川西 隆仁, 村瀬 洋, “物体検出装置、物体検出方法、プログラムおよび記録媒体”, 特許公開 2003-346156.
4. 川西 隆仁, 村瀬 洋, 高木 茂, “全周映像描画装置およびその方法”, 特許公開 2004-012477.
5. 土田 勝, 川西 隆仁, 村瀬 洋, 高木 茂, “物体抽出装置、物体抽出方法、プログラム及び記録媒体”, 特許公開 2004-213572.
6. 高木 茂, 川西 隆仁, 土田 勝, 村瀬 洋, “物体追跡装置と方法、及びそのプログラムとそのプログラムを記録した記録媒体”, 特許公開 2004-362210.
7. 川西 隆仁, 黒住 隆行, 柏野 邦夫, 高木 茂, “テンプレートマッチング方法、装置、プログラム及びその記録媒体”, 特許公開 2005-063206.

8. 土田 勝, 川西 隆仁, 高木 茂, “物体認識装置、物体認識方法、プログラム及び記録媒体”, 特許公開 2005-070026.
9. 土田 勝, 川西 隆仁, 村瀬 洋, 高木 茂, “物体追跡方法、物体追跡装置、物体追跡システム、プログラム、および、記録媒体”, 特許公開 2005-141687.
10. 木村 昭悟, 川西 隆仁, 柏野 邦夫, “部分画像検索システム及び方法並びにプログラム”, 特許公開 2005-352990.
11. 木村 昭悟, 川西 隆仁, 柏野 邦夫, “信号照合システム及び方法並びにプログラム”, 特許公開 2006-004153.
12. 木村 昭悟, 川西 隆仁, 柏野 邦夫, “部分画像検索システム及び方法並びにプログラム”, 特許公開 2006-011622.
13. 川西 隆仁, 黒住 隆行, 柏野 邦夫, 高木 茂, “テンプレート照合装置および方法”, 特許公開 2006-031392.
14. 木村 昭悟, 川西 隆仁, 柏野 邦夫, “映像検索装置, 方法及びプログラム並びにプログラムを記録した記録媒体”, 特許公開 2006-060796. 映像描画装置及び映像描画方法特許公開 2006-171578
15. 川西 隆仁, 土田 勝, 高木 茂, 村瀬 洋, 大和 淳司, “映像描画装置及び映像描画方法”, 特許公開 2006-171578.

## 表彰

1. FIT2003 第2回情報科学技術フォーラム FIT ヤングリサーチャー賞受賞, Sep. 2004.