The University of Osaka
Institutional Knowledge Archive

Title	対話を用いた物体認識に関する研究
Author(s)	槇原, 靖
Citation	大阪大学, 2005, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/1954
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

対話を用いた物体認識に関する研究

大阪大学大学院 工学研究科 電子制御機械工学専攻
槇原 靖

2004 年 12 月

論文概要

本論文では，対話を用いた物体認識とユーザの対話の負担を減らすための照明変化への対応手法に述べた．

第 1 章では，コンピュータビジョンによる物体認識におけるユーザとの対話を用いた物体認識の位置付け及び必要性と，照明条件変化への対応の必要性について述べた．

第 2 章では，ユーザとの対話を用いた物体認識システムについて述べた．最初に，ユーザから指定された物体を容易に認識できるように，物体をどの方向から見ても認識でき，他の物体との区別が付くような最小限度の特徴を持つような物体モデルを作成した．次に，自動物体認識を行った結果，複数の候補が検出されたり，対象物体が一部隠蔽されて認識に失敗するような場合に，ユーザから必要な情報を得られるように認識結果の表示と報告を行い，ユーザから得られた情報に基づいて最終的に認識を成功させる手法を提案した．

第 3 章では，ユーザの対話の負担を減らすためのグローバルな外部照明変化を考慮した色変換について述べた．物理的な光の反射モデルに基づく色のモデルを有限次元線形モデルによって離散化することにより，異なる照明条件下での線形な色変換を定式化した．次いで，物体モデル作成時の基準照明条件から認識時の未知照明条件への色変換を，認識時に観測される一つの参照物体の色を用いて推定する手法を定式化した．色変換パラメタを学習するためには様々な照明条件下で観測した物体と参照物体の色が必要であるので，最初は明度補正のみを考慮したパラメタを用いて色変換を行い，ユーザとの対話によって正しい物体認識結果を得る毎に他のパラメタを追加して更新することで，新しい環境へ適応する手法を提案した．最後に，本手法によって効率的に色変換を学習できることを実験によって確認した．

第 4 章では，外部照明と内部照明の 2 種類の光源によるシーン内でのローカルな照度変化を考慮した場合の色変換について述べた．外部照明については，照度と冷蔵庫内の位置との関係を学習するものとした．内部照明の照度については，光源との位置関係に加えて，物体自身の光の透過率や照明を遮る物体の有無による影響があり，冷蔵庫内の位置が同じでも物体の種類や物体配置が異なる場合にはその値が変化するため，観測毎に物体モデルの色に最も合う照度を当てはめるものとした．学習した色変換を用いて様々な照明条件下での冷蔵庫内シーンにおける物体認識実験を行い，本手法の有効性を確認した．

最後に，第 5 章において，本論文の成果のまとめと今後の課題について述べた．

目 次

第 1 章　序論 **4**
1.1　画像センサを利用した物体認識 4
1.2　画像センサと他の情報を組み合わせた物体認識 6
1.3　物体認識における照明条件の影響 8
1.4　本論文の構成 8

第 2 章　対話を用いた物体認識システム **11**
2.1　システムの枠組み 11
2.2　物体モデルの作成 14
2.2.1　各種類の形状モデル 15
2.2.2　投影画像の作成 17
2.2.3　物体の特徴の登録 28
2.2.4　特徴にばらつきのある物体モデルの作成 32
2.3　自動物体認識 33
2.3.1　候補領域の抽出 34
2.3.2　物体の種類の照合・決定 35
2.3.3　物体モデルとの照合 47
2.4　対話を用いた物体認識 49
2.4.1　認識結果に応じた対話 49
2.4.2　色の範囲がずれている物体の認識 51
2.4.3　部分的に隠蔽された物体の認識 52
2.4.4　対話によるモデルの学習 55
2.5　ユーザにとって選択しやすい認識結果表示法 55
2.5.1　認識結果の候補の表示方法 55
2.5.2　認識結果の階層的選択 61
2.6　まとめ 67

第 3 章　グローバルな外部照明変化を考慮した色変換の学習 **69**
3.1　関連研究との関わり 69
3.2　光の反射の物理モデルに基づく色変換のモデル 71
3.3　参照色を用いた色変換の推定 75

3.4 学習セットが十分にある場合の色変換の学習 76
3.4.1 色変換の学習結果 77
3.5 新しい環境での色変換の学習 79
3.6 照明変化を考慮した物体認識 84
3.7 実験 85
3.7.1 候補領域抽出実験 85
3.7.2 新しい環境への適応実験 86
3.8 まとめ 88

第 4 章 ローカルな照度変化を考慮した色変換の学習 92
4.1 関連研究との関わり 92
4.2 ローカルな照度変化を考慮した色変換の推定 93
4.2.1 色変換の問題設定 93
4.2.2 半透明物体の色のモデリング 94
4.2.3 照度変化を含む未知照明下での物体の色推定 96
4.3 ローカルな照度変化を考慮した色変換の学習 100
4.3.1 色変換の学習方法 100
4.3.2 色変換の学習結果 102
4.4 ローカルな照度変化を考慮した物体認識 107
4.5 実験 107
4.5.1 物体の検出実験 108
4.5.2 物体の識別実験 109
4.5.3 総合評価 110
4.6 まとめ 110

第 5 章 結論 113

付 録 A 冷蔵庫領域の抽出 121

付 録 B 色に基づく領域分割 123

付 録 C 物体の 3 次元位置の取得 125

第1章　序論

近年，高齢社会の到来により介護を必要とする人が増加しているため，社会福祉関係の労働力に対する需要が高まってきている．しかし一方では，少子化の進行に伴う労働力の不足により，十分な介護を受けられない人が出てくることも予想される．そのような事態を想定して，現在，被介護者・介護者の作業を補助・代行する機器の研究開発が数多く行われている．例えば，筋力の衰えを補助するための装着型のパワーアシスト装置 [1]，リハビリ歩行支援器 [2]，食事支援ロボット [3] や日常品をユーザのところに持ってくるサービスロボット [4]，外出時の移動支援のための知的車椅子 [5] や段差のある市街地での移動支援ロボット [6] 等がある．

そのようなロボットの機能の中で，ユーザに指定されたものを持って来るという機能は，寝たきりの人や足の不自由な人にとって非常に重要なものである．このようなサービスロボットを実現するためには，ユーザとの音声対話機能，指定された物体を認識してその位置を知る機能，認識した物体を操作する機能，目的地までの自律移動機能等が必要となる．中でも指定された物体を認識してその位置を知る機能は重要である．

物体を検出するための方法としては，物体に薄い磁石を添付しておき磁気センサ [7] を使って物体を検出することが考えられるが，物体の識別ができず，また検出距離も短いことから，実用は困難である．また，バーコードを利用して物体の検出と識別を行う方法が考えられるが，物体の向きによってはバーコードが背面に来るために，検出できないことがある．また，バーコードより記憶容量の大きな RF タグ [8] を物体に添付することで，物体の検出と識別を行うことが考えられるが，物体のおおよその位置は分かっても，正確な物体の境界が分からないため，実際にマニピュレータで取りに行く際には問題となる．

1.1　画像センサを利用した物体認識

画像センサを用いてシーンを理解する技術は一般にコンピュータビジョンと呼ばれる．コンピュータビジョンに関する研究の初期段階では，画像のノイズやボケを修正する画質向上のための画像処理の研究や，画像上の一様色領域やエッジを利用して物体の輪郭線を抽出することでシーンを理解する研究等が行われた．これ以降，パターン認識の技術として，認識対象物体の画像をテンプレートとして持っておいて入力された画像からその物体を検出するテンプレートマッチング，高

次元の画像情報を主成分分析に基づいて特徴を良く表す低次元空間に投影して効率的に認識を行う固有空間法 [9]，直線や円等の幾何的特徴を投票原理によって抽出するハフ変換 [10] 等が研究された．また，カラー画像を扱うようになってからは，色彩に関する研究や物理的な光の反射現象に基づく色モデルの考案等がなされた．加えて，移動物体の追跡などを目的として，時間的に連続した複数枚の画像を処理する動画像処理 [12] の研究も行われ，連続差分画像 [13] やオプティカルフロー [14] 等の手法が開発された．こういったコンピュータビジョン技術の中で，サービスロボットの物体認識のために使える技術の一つとして，距離画像センサを用いた物体の位置や形状の認識が挙げられる．

距離画像センサの一つとして，三角測量法に基づくステレオ視 [15] や光投影法 [16] 等が挙げられる．ステレオ視は，装置として最低限 2 台のカメラがあれば良いということで非常に単純であるため，比較的古くから研究が行われている．例えば，ステレオ視による自律移動ロボットのため障害物回避やナビゲーション [17, 18, 19]，ステレオ視を用いた物体認識 [20, 21, 22] 等に幅広く用いられている．また，最近では実際にステレオ視による距離画像取得機能を実装した製品 [23][24] も発売されている．但し，ステレオ視の問題点として，距離が遠くなるにつれて距離精度が悪くなるという性質があり，またオクルージョン等が原因で左右の画像での対応付けの曖昧さが生じることがあるため，生活空間における冷蔵庫や戸棚等の複雑なシーンから距離情報に基づいて完全な三次元再構成を行うことは困難である．光投影法の場合は，画像の対応付けの曖昧性が無くなるという利点があるが，画素値がある一定以上の明るさを超えているような場合には，投影した光を画像上で明確に観測できない場合がある．

他の距離画像センサとしては，光飛行時間測定法に基づくレンジファインダ [25] が挙げられる．この手法ではステレオ視と異なり，オクルージョンによる死角が生じないという特長がある．その中でも，レーザ距離センサ (例えば [26][27]) は非常に精度が良く，最近では，十数メートルの測定範囲に対して，5〜10cm 程度の距離精度を持つものもあり，自律移動ロボット [29][30] や研究用の自律走行車 [28] に搭載されている．また，レーザ距離センサ等によって得られた距離情報から，平面や曲面を当てはめて物体の 3 次元形状を認識する研究 [31][32] もなされている．しかし，レーザ距離センサはある水平断面に対する距離計測を行うため，高さ方向へのスキャンをさせるためには装置が大がかりになってしまう．

これらの特別な装置を用いずに，一枚の画像から物体の三次元形状を復元するために画像の濃淡変化を利用する手法 [33] やテクスチャの変化を利用する手法 [34] もあるが，これらの手法はノイズに対して非常に敏感である上，物体表面内の相対的な位置を取得するのみで，絶対位置については取得できない．

また，これらの距離画像センサを単独で用いた場合には，同じ形状でテクスチャの異なる物体同士の識別はできない．これに対して，2 次元画像上での物体の検出と識別を画像特徴 (色やエッジ) を利用して行い，それを 3 次元位置に投影して物体の位置を取得する方法が考えられる．但し，2 次元位置から 3 次元位置を算出す

る場合には，1 次元分の制約知識が必要となる．本手法では，物体が接地している水平面の高さを事前知識として与えることによって，画像上の物体の接地点を実世界の水平面上の点に投影して物体の 3 次元位置を求めるものとする (詳細は付録 C 参照).

一方，これまでに挙げたようなコンピュータビジョン技術は，先に挙げた例に加えて，工場における部品の種類判別，位置合わせ，欠陥検出の自動化といったような工業分野を中心に応用が進んでいる．しかし，これらの多くは照明条件を一定にすることや，扱う部品形状の種類を少数に限定することや，部品同士の区別がつくように部品が重ならないようにすること等の制約を必要としている．一方，サービスロボットに対するコンピュータビジョン技術の応用を考えると，ユーザが暮らす環境で照明を一定にするという制約や物体が互いに重ならないように配置されているという制約を加えることは困難である．また，照明変化や物体の重なりによる画像特徴の変化を全て考慮して物体を認識しようとすると，他の物体との誤認識の可能性が増加する．このような場合には，画像情報以外の情報を補助として用いて物体認識を行う方法が考えられる．

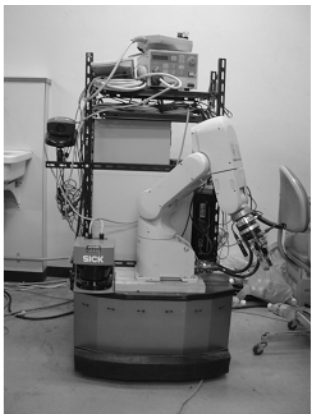
1.2　画像センサと他の情報を組み合わせた物体認識

例えば，画像情報とユーザからのマウス入力情報を組み合わせた研究としては，マウスポインティングで花弁等を選択することで植物の認識に利用する研究 [41] やユーザが画像から物体輪郭をマウスでなぞって切り出す際に画像上のエッジ等を利用して多少のずれを補正するインテリジェントシザーの研究 [42] 等がある．

画像情報とテキストによる言語情報を組み合わせた研究としては，画像情報からその画像に対する説明文を生成する研究 [43][44] や，植物図鑑の花や果実を，図に添付された説明文を利用して認識するシステム [45] がある．

ここで，サービスロボットに対して与えられる画像情報以外の情報について考えてみると，ユーザとの音声対話による情報が最も自然であると考えられる．このことは，実際にある人が冷蔵庫からある色の缶を取って欲しいと頼まれたときに，その色の缶が 2 本入っている場合に，頼んだ人にどちらを取るかを尋ねること，即ち頼んだ人と対話することで，取ってくる缶を決定することからも妥当であることが分かる．

画像情報とユーザとの音声対話を組み合わせた研究としては，ユーザの発話の内容に一致する特徴を画像から抽出することで，物体を認識するシステム [46][47] 等がある．しかし，いずれも認識が失敗したときにそこから再度認識を試みて成功させることは行っていない．高橋ら [48] は，言語とジェスチャによる対話を通して，りんごや本を認識するロボットの研究を行った．ロボットはユーザがジェスチャによって指し示した方向を見て，音声で指定された物体を画像から認識し，複数の物体が発見された場合には，ユーザとの対話によって物体を選択する．し



(a) 本研究で使用するロボット

(b) ユーザの使用環境

図 1.1: システムの概要

(a) 他の物体による隠蔽のあるシーン

(b) 同じ物体による隠蔽のあるシーン

図 1.2: 自動物体認識が困難なシーン

かし，物体が発見されなかった場合に再認識を試みることは行っていない．

本研究では，冷蔵庫から必要なものを取って来るサービスロボット (図 1.1(a)) において，認識失敗時にもユーザとの対話を通して認識を成功に導くことを行う．本システムに対して，ユーザは最初に手元のディスプレイを見ながら，音声でロボットに指示を出す (図 1.1(b))．ここで，必要以上にユーザを煩わせないために，最初は，あらかじめ登録された物体モデルを用いて可能な限り自動で認識を試みる．得られた認識結果は，ディスプレイと音声を用いてユーザに伝えられる．図 1.2 のように隠蔽等が生じて自動物体認識が失敗した場合や，ディスプレイで結果を見たユーザが間違いを指摘した場合には，ユーザとの音声による対話を通して認識の補助となるような情報を得て，それを基に物体認識を試みる．

1.3　物体認識における照明条件の影響

一般に物体認識を行う視覚システムにとって，照明条件の影響に対する考慮も重要である．一つの方法として，工場の例と同様に，照明条件を完全に一定にコントロールすることで，観測色の変化を防ぐことが考えられるが，ユーザの生活環境を一定の照明に保つためには，常に外からの光を遮断して，電灯を点けたままもしくは消したままにしておく必要があり，ユーザに対して負担をかけることとなる．また，ロボット自身が光を照らすことで照明条件をコントロールすることが考えられるが，やはりロボット自身の光以外の光源を完全に遮断する必要があり，ユーザの生活空間に制約を加えることとなる．

他の方法としては，認識時の照明条件を時刻や電灯の ON・OFF 情報を取得することで事前に完全に把握しておき，その照明条件に合った色の補正を行うことが考えられる．しかし，天候・カーテンの開け閉め・季節の変化による外からの光強度の変化までを考慮することは困難である．

他には，照明条件の変化によって生じる観測色のばらつきを学習しておき，そのばらつきを考慮して物体領域の抽出や照合を行うことが考えられる．しかし，認識すべき色の種類が多い場合や，照明変化の幅が大きい場合には，照合時に他の物体と誤認識する可能性が高くなる．

よって，システムはその時々に応じて照明条件を推定して，観測色の変化に対応していく必要がある．そこで，本研究では認識時に観測される一つの参照色を利用して，認識時の照明条件を推定して，適切な色変換を求めることで，観測色の変化に対応する．また，システムを新しい環境へ適応する場合には，ユーザとの対話を用いて認識を行いつつ，そこで得た認識結果を元に参照色と色変換の関係を更新していくことで，徐々に色変換精度を上げて認識失敗によるユーザの対話の負担を軽減する．

1.4　本論文の構成

本論文では，対話を用いた物体認識とその際のユーザの対話の負担を可能な限り減らすための照明変化への対応手法に述べる．

第 2 章では，ユーザとの対話を用いたサービスロボットの物体認識システムについて述べる．最初に，本研究におけるシステムの枠組みについて触れ，物体認識におけるユーザとの対話の役割を述べる．次に，ユーザから指定された物体を容易に認識できるように，物体をどの方向から見ても認識でき，他の物体との区別が付くような最小限度の特徴を持つような物体モデルの作成方法について述べる．次に，そのモデルをもとにした自動物体認識を行った後，ユーザにとって選択しやすいように認識結果を表示する方法について述べる．その結果に対してユーザによる複数候補からの選択や間違いの指摘を受けた場合に，その助言に基づいて再度認識を試みることで，最終的に認識を成功させる対話を用いた認識につい

て述べる．

第 3 章では，第 2 章で述べたシステムにおいて，ユーザの対話の負担を減らすためのグローバルな外部照明変化を考慮した色変換の学習について述べる．物体モデル作成時の基準照明条件から認識時の未知照明条件への色変換を，認識時に観測される一つの参照物体の色を用いて推定する手法を定式化する．色変換と参照物体の色の関係については，物体モデルの色と様々な照明条件下で観測したその物体の色と参照物体の色の組を用いて学習する．システムがこのような学習用の色の組を持たない新しい環境で用いられる場合には，最初は適当な関係を仮定した色変換を行い，ユーザとの対話によって正しい物体認識結果を得る毎にその関係を更新することで，新しい環境への適応を実現する．

第 4 章では，第 3 章で述べた手法を，外部照明と内部照明の 2 種類の光源によるシーン内でのローカルな照度変化を考慮した場合に拡張する．外部照明の照度変化については光源との位置関係による影響が強いため，冷蔵庫内の位置との関係を学習する．内部照明の照度については，光源との位置関係に加えて，物体自身の光の透過率や照明を遮る物体の有無による影響があり，冷蔵庫内の位置が同じでも物体の種類や物体配置が異なる場合にはその値が変化するため，観測毎に最も合う内部照明照度を当てはめて対応する．様々な照明条件下での冷蔵庫内シーンにおける物体認識実験を行い，本手法の有効性を示す．

最後に，第 5 章において，本論文の成果のまとめと今後の課題について述べる．

本論文では，対話を用いた物体認識とユーザの対話の負担を減らすための照明変化への対応手法に述べる．

第 2 章では，ユーザとの対話を用いた物体認識システムについて述べる．最初に，ユーザから指定された物体を容易に認識できるように，物体をどの方向から見ても認識でき，他の物体との区別が付くような最小限度の特徴を持つような物体モデルを作成する．次に，自動物体認識を行った結果，複数の候補が検出されたり，対象物体が一部隠蔽されて認識に失敗するような場合に，ユーザから必要な情報を得られるように認識結果の表示と報告を行い，ユーザから得られた情報に基づいて最終的に認識を成功させる手法を提案する．

第 3 章では，ユーザの対話の負担を減らすためのグローバルな外部照明変化を考慮した色変換について述べる．物理的な光の反射モデルに基づく色のモデルを有限次元線形モデルによって離散化することにより，異なる照明条件下での線形な色変換を定式化する．次いで，物体モデル作成時の基準照明条件から認識時の未知照明条件への色変換を，認識時に観測される一つの参照物体の色を用いて推定する手法を定式化する．色変換パラメタを学習するためには様々な照明条件下で観測した物体と参照物体の色が必要であるので，最初は明度補正のみを考慮したパラメタを用いて色変換を行い，ユーザとの対話によって正しい物体認識結果を得る毎に他のパラメタを追加して更新することで，新しい環境へ適応する手法を提案する．最後に，本手法によって効率的に色変換を学習できることを実験によって確認する．

第 4 章では，外部照明と内部照明の 2 種類の光源によるシーン内でのローカルな照度変化を考慮した場合の色変換について述べる．外部照明については，照度と冷蔵庫内の位置との関係を学習するものとする．内部照明の照度については，光源との位置関係に加えて，物体自身の光の透過率や照明を遮る物体の有無による影響があり，冷蔵庫内の位置が同じでも物体の種類や物体配置が異なる場合にはその値が変化するため，観測毎に物体モデルの色に最も合う照度を当てはめるものとする．学習した色変換を用いて様々な照明条件下での冷蔵庫内シーンにおける物体認識実験を行い，本手法の有効性を確認する．

最後に，第 5 章において，本論文の成果のまとめと今後の課題について述べる．

第2章　対話を用いた物体認識システム

本節では，最初に 2.1 節でサービスロボットのシステムの枠組みについて述べる．2.2 節では認識対象の物体モデルの作成方法について述べ，2.3 節ではそのモデルに基づく自動物体認識について述べる．2.4 節では，ユーザとの対話を用いた物体認識について述べ，2.5 節ではユーザが選択しやすいような認識結果の表示方法について述べる．

2.1　システムの枠組み

サービスロボットのための物体認識の問題設定は，ユーザの指定した物体に一致するものを画像内から探して，その画像上の位置を取得するということである．このような問題設定における変数としては，表 2.1 のようなものが挙げられる．

表 2.1: 物体認識における変数とその依存対象

(a) 画像で直接観測される変数

変数	依存対象
画像上の色	物体名 (ID)，物体の向き，照明変化
画像上の大きさ	物体の状態，カメラ位置
画像上の位置	物体名 (ID)，カメラ位置

(b) 画像では直接観測されない変数

変数	依存対象
物体名 (ID)	ユーザの指定
物体の向き	冷蔵庫内での置き方
照明条件	時刻，電灯の ON・OFF，冷蔵庫内の位置
物体の状態	隠蔽，同色物体の重なり
カメラ位置	冷蔵庫に対するロボットの位置・姿勢

これに対してユーザから提供され得る情報についてまとめると，表 2.2 のようなものが考えられる．

表 2.2: ユーザが与えることのできる情報

情報	例
物体名 (ID)	商品名: ポカリスエット，十六茶
	種類名: お茶，オレンジジュース，コーヒー
物体の状態	重なっている，隠されている
画像上の色	赤，橙，黄，緑
画像上の位置	上の棚の左，青い缶の右

ここで，ユーザの指定した冷蔵庫内の物体について，画像上の色，画像上の大きさがモデル登録時と同じである場合には，単純に画像上のモデルに合うものを抽出すれば認識は終了する．しかし，これらの変数は実際には表 2.1 の依存対象によってばらつきを持つことになる．

もし，これらの変数のばらつきを考慮せずに認識を行うと，隠蔽などが生じた場合には認識に失敗する．一方，これらの変数のばらつきを全て含むようにして認識を行うと，対象物体以外の候補が多数検出され，ユーザに選択してもらう手間が必要となる．

そこで，まずこれらの変数のばらつきを可能な限り小さくするために，物体の向きが変化しても認識できる物体モデルを作成して (2.2 節参照)，照明条件を推定して画像上の色のばらつきを小さくする (第 3, 4 章参照)．そして，隠蔽等の物体の状態の影響のように，普段はあまり起こらないようなことが原因となっている画像上の大きさについては，最初はそのばらつきが小さいものと仮定して認識を試みて，それで失敗した場合にユーザからの助言をもとにして，必要な変数を変化させて再認識を試みる．

例えば，ユーザに対象物体が「隠されています.」と言われれば，画像上の大きさが小さいものも許可して認識を試みること (変数のばらつきを広げての再探索) や,「上の棚の右にあります.」と言われれば，画像上の位置を上の棚の右に限定して，他の変数のばらつきを広げて再探索することを行う．

このような方針を採ることで，隠蔽などが起こっていない通常時は，小さいばらつきのもとで探索を行って，不必要な候補を提示しないことでユーザの選択の手間を省き，隠蔽などの大きなばらつきが起こって認識が失敗した場合には，全変数のばらつきを広げるのではなく，ユーザとの対話を通して得られた助言に基づいて，必要のある変数のばらつきのみを広げることで，効率良く物体の再認識を試みる．

これをサービスロボットのシステム全体に対する処理の流れの中で考えてみると，図 2.1 の様になり，システムは大きく分けると以下のようになる．

- ロボット

カメラによる画像取得や物体の把持やユーザへの手渡し，目的地への移動を行う．

- 物体認識部

  物体認識条件によって必要な物体認識モジュール (通常認識，色拡張認識，隠蔽物体認識等) を起動して，認識結果を取得する．

- 音声対話部

  ユーザからの発話の音声認識やシステムの発話の音声合成を行う．

- 制御部

  ロボット・物体認識部・音声対話部を束ね，各部への依頼や各部からの結果を基にして状況判断を行い，タスク達成のための行動計画を決定する．

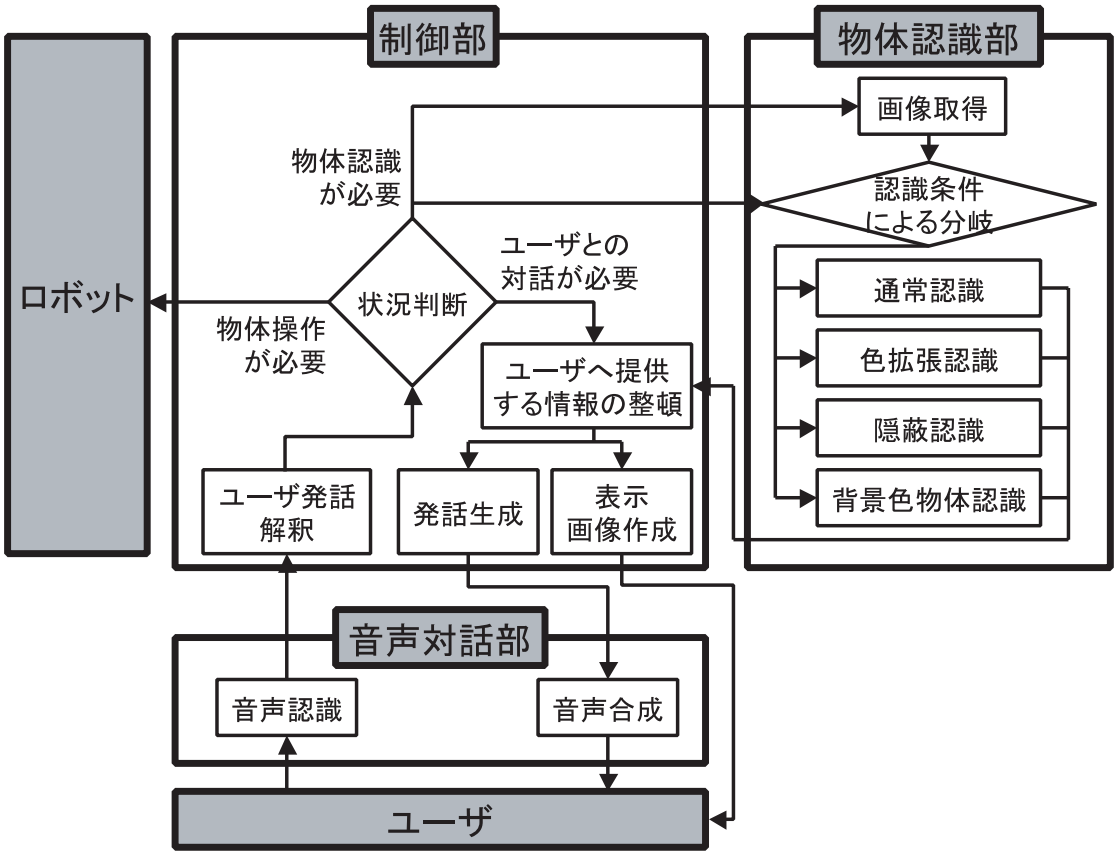


図 2.1: システムの処理の流れ

まず，ユーザがシステムに対して取って欲しい物体を音声によって依頼する．制御部では音声認識された発話内容を解釈して，状況判断を行う．ユーザの発話が誤認識されているような場合には，もう一度ユーザと対話をして正しい情報を得

る必要があると判断して，ユーザに必要な情報を尋ねる．ユーザの発話が正しく認識されている場合には，依頼された物体を認識する必要があるので，その物体を認識条件に設定する．物体認識部は，画像取得後，認識条件によってどの物体認識モジュールを起動するかを判断して，その認識結果を制御部へ返す．制御部はその認識結果を基にユーザへ提供する情報の整頓を行い，その情報をシステムの発話として生成して音声でユーザに伝えると同時に，認識結果として表示画像を作成してディスプレイに表示してユーザに示す．

ユーザはディスプレイと音声で示された認識結果や質問文を解釈して，確認・選択・助言等の返答を行う．制御部では再びユーザの応答を基に状況判断を行い，その結果，画像上で把持や除去等の操作をすべき物体が一つに定まれば，ロボットにその物体の 3 次元位置を渡して物体操作を依頼する．

一方，画像上で操作すべき物体が一つに定まらなかった場合には，ユーザとの対話によって候補を絞れば良いのか，再度物体認識が必要 (隠蔽などによって認識が失敗した場合) なのかを決定する．候補を絞る場合には，現在の状況からどのような質問をするかを決定する．また，再度認識が必要な場合は，ユーザの助言に基づいて認識条件を決定 (例えば，隠蔽に関する情報が与えられたら隠蔽認識に設定し，単に場所を限定されれば色拡張認識に設定) して，再び物体認識部に依頼する．ユーザとの対話を利用してこのループを繰り返すことで，最終的にタスクを達成することができる．

ここでは，音声対話や移動マニピュレータを用いた物体操作に関しての説明を省いたが，その詳細について，音声対話システムは文献 [70, 71] を，物体操作システムは [72] を参照されたい．

2.2　物体モデルの作成

本システムは，自動物体認識を行うために，あらかじめその物体のモデルを作成・登録しておく．登録を行う際には，ロボットはユーザに物体名を教えてもらい，その物体をマニピュレータに手渡してもらうと，その物体をカメラで撮影して特徴を自動的に登録する．

先に述べた物体認識の枠組みにおける変数のばらつきの中で，物体モデル作成時に対応しておく必要があるものは物体の向きである．これは，ユーザから物体の方向に関する補助的情報をもらうことは考えにくいためである．よって，物体モデルはどの方向から見ても認識可能であるように作成する必要がある (但し，どの方向であるかを認識できるようにする必要はない)．このことを達成するためには，物体を各方向から見た場合の特徴を登録しておく必要がある．ここで，最初に考えられる方法としては，方向毎のテンプレート画像を登録しておいて，その全てを使ってテンプレートマッチングを行う方法がある．しかし，この方法では物体を検出するために，画像上の位置の走査とテンプレート画像上での相違度計

算のための走査を方向毎のテンプレート画像に対して行う必要があり，計算量が非常に多くなる．

このような手間を省くためには，できるだけ簡単に物体を検出できるようにすることが望ましい．そこで物体の特徴を，以下のように分けて登録する．

- 物体を検出するための最低限の特徴
- 他の物体との識別のための特徴

また，ここで検出するための特徴として，方向毎にあまり変化しないものを選ぶと，特徴が変化しない方向の区間毎に一まとめにした特徴だけを用いれば良く，全方向に対して各々検出を試みる場合に比べて計算量が減少する．

このような要求に基づいて，物体モデルの特徴として以下のものを用いる (図 2.2 参照).

1. 物体の実際の大きさ

   次節で述べる形状毎のモデルに基づく各種寸法 (缶の場合は図 2.2 のように幅と高さ) とする．ここで，画像上の大きさから実世界における物体の大きさを求めるために，あらかじめカメラをキャリブレーションしておき，そのカメラから既知の位置にマニピュレータを用いて物体を置くものとする．

2. 代表色

   最大面積を持つ一様色領域の色とする．

3. 2 次特徴

   代表色以外の一様色領域の色，上下端位置，横幅，面積とする．ここで，2 次特徴として左右端位置ではなく横幅を選択したのは，物体上のある領域は，その物体が回転することでその左右端位置が変化するが，横幅に関してはほぼ一定のままになるからである．

ここで，1,2が検出に最低限必要な特徴であり，3が識別用の特徴である．結局，物体の実際の大きさと代表色については全ての物体について登録しておき，2 次特徴については物体同士の区別が付くための最小限度のものだけを登録しておくという方針を採る．

一方，缶・瓶・ペットボトルとは異なり，果物のように大きさや代表色といった特徴にばらつきを持つ物体を登録する場合には，そのばらつきについても考慮する必要がある．このような物体の登録については 2.2.4 で述べる．

2.2.1 各種類の形状モデル

- 缶

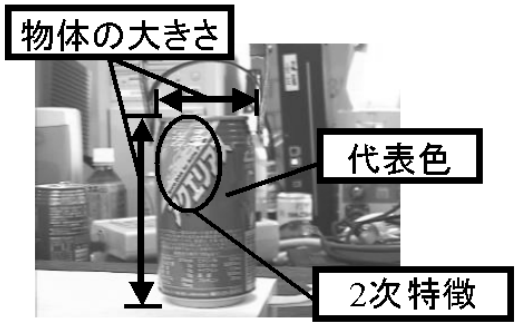


図 2.2: 物体の特徴

缶は，細長い円筒形と見ることができ，今回扱う物体の中で最も基本的な形状である．円筒形物体を斜め上から見下ろしたとき，上面の円が画像上では楕円として映り，側面の線がその楕円の長軸の両端から下に続くほぼ垂直な線分として映るので，上面の楕円と側面の垂直線分を抽出することで円筒物体を認識することができる．しかし，冷蔵庫内の缶を認識する際には，以下の問題点がある．

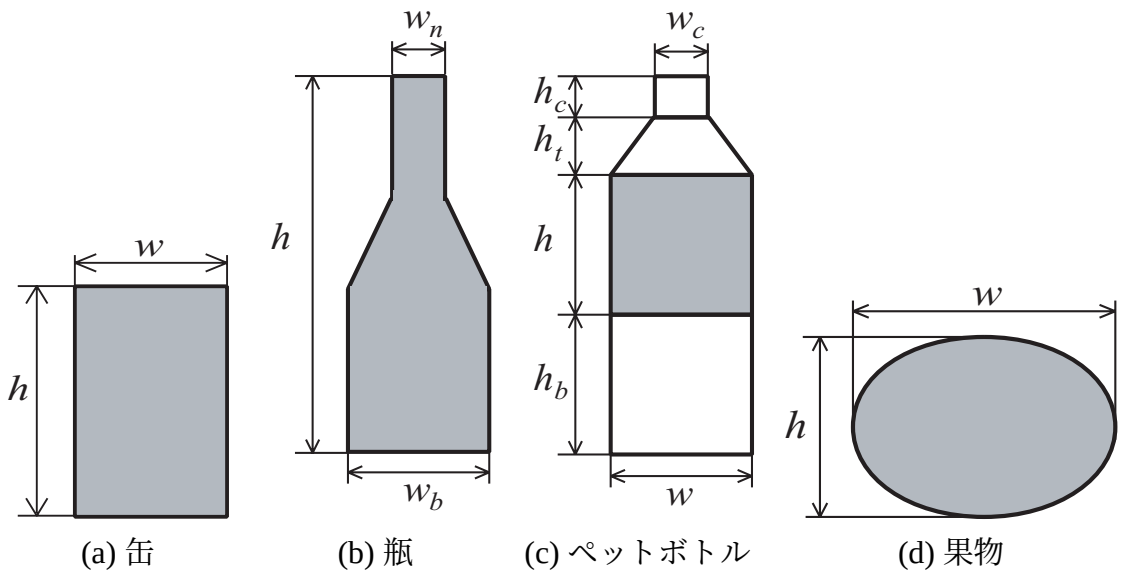
- 冷蔵庫全体を画像に収めるので，飲料用の缶はそれ程大きく映らないため，上面の楕円は画像上では小さく映る．
- あまり上から見下ろすと冷蔵庫の奥の方にある物体が見えないことから，比較的水平に近い角度で見るために，上面の楕円の短軸が長軸に比べて非常に短く映る．
- 飲料用の缶は上面が鏡面反射が支配的な光沢を持つ金属であるので，冷蔵庫の照明によって部分的にハイライトが生じるために，一様な領域として抽出することが困難である．

これらの問題から，冷蔵庫内の缶の認識に上面の楕円を利用することは困難である．代わりに缶を長方形近似して，図 2.3(a) のように高さ h と幅 w を登録する．

- 瓶

  瓶には様々な色や形状をしたものがあるが，ここでは透明な瓶は扱わず，ビール瓶のような色のついた瓶で，胴体部分と細い首の部分がある瓶について扱う．瓶の形状特徴としては，図 2.3(b) のような胴体部分の 1 組の垂直線分と間隔の狭い首部分の 1 組の垂直線分とが考えられる．そこで，瓶全体の高さ h に加えて，胴体部分の幅 w_b と首部分の幅 w_n を登録する．

- ペットボトル



(a) 缶　(b) 瓶　(c) ペットボトル　(d) 果物

図 2.3: 各種類の形状モデル

標準的な飲料用ペットボトルは，図 2.3(c) のようにラベル・キャップ・ラベルの下の容器部分の 3 つの部分から成り立っている．そこで，各部分に着目して，ラベルの高さ h と幅 w，キャップの高さ h_c と幅 w_c，そしてラベルからキャップまでの高さ h_t とラベルから下端までの距離 (ラベル下の容器部分の高さ)h_b を登録する．また，ペットボトルの代表色や 2 時特徴を登録する場合には，内容液の有無によっても殆ど見えの変化しないラベル部分を用いる．

- 果物

  ここでは，蜜柑や林檎等，図 2.3(d) のように画像上で楕円近似可能な物体について取り扱い，その高さ h と幅 w を登録する．

2.2.2　投影画像の作成

物体をどの方向からでも認識できるようにするためには，多方向から見た画像より特徴を抽出する必要があるが，効率化のために代わりに図 2.4 のような 1 枚の投影画像を作成して特徴を抽出する．ここで言う投影画像とは，物体に外接する仮想的な円筒にその表面の模様を中心軸から投影して，その円筒を長方形平面に展開したものである．また，物体を撮影するときの照明条件やカメラの絞りの変化によって，画像全体の色が変化することや物体表面にハイライトが生じることもあるので，投影画像の作成の際にはそれらを補正する必要がある．以下では，この投影画像の作成方法について述べる．

図 2.4: 物体の投影画像

画像全体の色補正

物体の撮影時には，外部照明の変化やカメラの絞りの変化等によって図 2.5(a) のように画像全体の色が変化する．ここでは，このような撮影時の照明条件の補正を行うために，3.2 節で定義する以下のような色の線形変換を考える．

$$\begin{bmatrix} \hat{R} \\ \hat{G} \\ \hat{B} \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \\ 1 \end{bmatrix} \tag{2.1}$$

ここで，$[R, G, B, 1]^T$ は変換前の色，$[\hat{R}, \hat{G}, \hat{B}, 1]^T$ は変換後の色，a_{ij} は変換パラメタを表すものとする．モデル撮影時には，変換パラメタを推定するためにモデルの撮影場所に参照白色 (図 2.5(a) の黒枠) を用意しておき，観測された参照白色 $[R_{ref}, G_{ref}, B_{ref}, 1]^T$ が基準の白色 $[R_{max}, G_{max}, B_{max}, 1]^T = [255, 255, 255, 1]^T$ になるような変換パラメタを求める．しかし，参照白色を用いて求まるパラメタは 3 個しかないので，式 (2.1) の変換パラメタの内，最も重要な対角成分の変換パラメタのみを考慮した以下の変換を考える．

$$\begin{bmatrix} \hat{R} \\ \hat{G} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.2}$$

これより変換パラメタは以下のように求まる．

$$a_{11} = \frac{R_{max}}{R_{ref}},\ a_{22} = \frac{G_{max}}{G_{ref}},\ a_{33} = \frac{B_{max}}{B_{ref}} \tag{2.3}$$

これにより，式 (2.2) と式 (2.3) の変換パラメタを用いて，画像の RGB 成分の補正を行う．図 2.5(b) は図 2.5(a) のそれぞれの画像を補正したものであり，図 2.5(a) の照明条件の異なる 2 枚の画像における物体の色の差が殆ど無くなっていることが分かる．

(a) 補正前の画像　　(b) 補正後の画像

図 2.5: 参照白色を用いた画像全体の色補正 (黒枠内:参照白色)

断片画像の作成

投影画像を作成するためには，いくつかの方向から撮影した物体の断片をつなぎ合わせる必要がある．ここでは，その断片画像の作成方法について述べる．

まず，各方向から物体を撮影した画像から，物体のおおよその範囲を抽出する．ここでは，あらかじめカメラのキャリブレーションを行っておき，マニピュレータを用いて物体を既知の場所に置くことから，画像上に映る物体の位置を予想することができる．ここでは，画像上での物体への外接矩形の予想位置を

$$S_{rect} = (x_1, y_1, x_2, y_2) \tag{2.4}$$

として定義しておく．ここで，x_1, y_1 は矩形の左上の点の x, y 座標，x_2, y_2 は右下の点の x, y 座標である．この予想位置に対して，適当な探索範囲を設定してエッジを用いて縦・横それぞれの線を抽出することで，物体の範囲を決定する．

ここでは，左側面を例にとって，探索範囲を設定して物体側面の縦線の位置，即ち x 座標の値を求める具体的な方法について説明する．

- 図 2.6(a) のような探索範囲の矩形を

$$S_l = (x_{l1}, y_{l1}, x_{l2}, y_{l2})$$

$$= \quad (x_1 - x_s, y_1, x_1 + x_s, y_2)$$

として設定する．ここで，x_s は探索範囲の大きさを決めるものであり，実験的に $x_s = 5$ とする．

- $[x_{l1}, x_{l2}]$ に含まれる各 x に対して，以下の式に従ってエッジ強度を計算する．

$$E_v(x) = \frac{1}{y_{l2} - y_{l1} + 1} \sum_{y=y_{l1}}^{y_{l2}} D(\mathbf{c}(x-1, y), \mathbf{c}(x+1, y)) \tag{2.5}$$

ここで，$\mathbf{c}(x, y)$ は点 (x, y) での画素値 $[R, G, B]^T$ を表しており，また D は以下の式で定義されるユークリッドノルムを表している．

$$D(\mathbf{c_1}, \mathbf{c_2}) = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2} \tag{2.6}$$

- 算出された $E_v(x)$ を最大にする x，即ち

$$x_v = \arg\max_x \{E_v(x)\} \tag{2.7}$$

を求める縦線の位置とする．

右側面の縦線については，図 2.6(b) のような探索範囲を以下のように設定することで，同様に求める．

$$\begin{aligned} S_r &= (x_{r1}, y_{r1}, x_{r2}, y_{r2}) \\ &= (x_2 - x_s, y_1, x_2 + x_s, y_2) \end{aligned}$$

また，上側の横線の位置については，以下のように求める．

- 図 2.6(c) のように，探索範囲の矩形を

$$\begin{aligned} S_t &= (x_{t1}, y_{t1}, x_{t2}, y_{t2}) \\ &= (x_1, y_1 - y_s, x_2, y_1 + y_s) \end{aligned}$$

として設定する．ここで，y_s は探索範囲の大きさを決めるものであり，実験的に $y_s = 5$ とする．

- $[y_{t1}, y_{t2}]$ に含まれる各 y に対して，以下の式に従ってエッジ強度を計算する．

$$E_h(y) = \frac{1}{x_{t2} - x_{t1} + 1} \sum_{x=x_{t1}}^{x_{t2}} D(\mathbf{c}(x, y-1), \mathbf{c}(x, y+1)) \tag{2.8}$$

- 算出された $E_h(y)$ を最大にする y，即ち

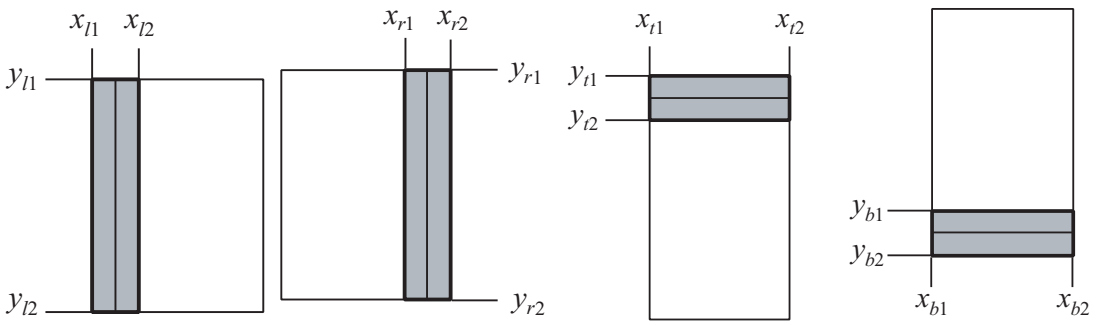
$$y_h = \arg\max_y \{E_h(y)\} \tag{2.9}$$

を求める縦線の位置とする．

下端の横線については，図 2.6(d) のような探索範囲を以下のように設定することで，同様に求める．

$$\begin{aligned} S_b &= (x_{b1}, y_{b1}, x_{b2}, y_{b2}) \\ &= (x_1, y_2 - y_s, x_2, y_2 + y_s) \end{aligned}$$

以上の方法によって求められた物体のおおよその輪郭の例を図 2.7 に示す．



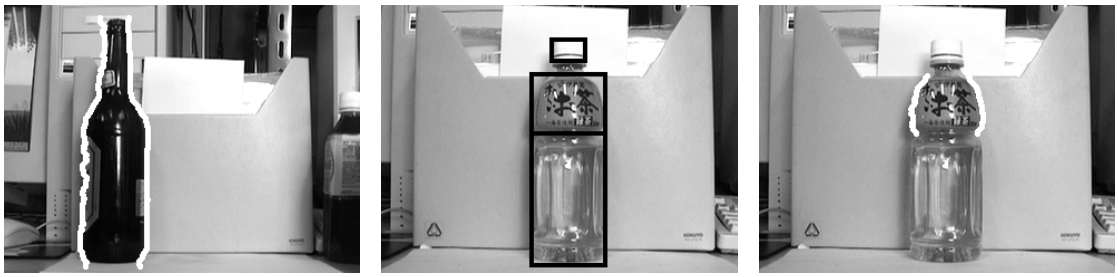
(a) 左端　(b) 右端　(c) 上端　(d) 下端

図 2.6: 線分の探索範囲の模式図

図 2.7: 抽出された物体の輪郭 (白線)

缶のように円筒形近似できるものについてはこれで輪郭抽出が終了するが，瓶やペットボトルについては物体の各部寸法を得るために更に詳細な輪郭を抽出する必要がある．

瓶については，缶と同様にしておおよその範囲を決定した後，その範囲内で左右の下端から上端に向かって連続したエッジ点を抽出することで，図 2.8(a) のように詳細な輪郭を取得する．また，ペットボトルについては，おおよその範囲を決定した後，まずキャップとラベルの範囲をエッジを用いて求める．ここで，キャップやラベルの大きさはあらかじめある程度分かっているものとして，その大きさをもとに画像上での予想位置を求めて，そこから缶の場合と同様に与えた探索範囲をもとにしてエッジを抽出することで，図 2.8(b) のように求める．次いで，ラベル部分の輪郭を瓶の場合と同様に図 2.8(c) のように求める．



(a) 瓶の輪郭　(b) ペットボトルの各部分　(c) ペットボトルのラベルの輪郭

図 2.8: 瓶・ペットボトルに対する物体の輪郭抽出

最後に，求められた物体の輪郭から物体の断片画像を作成する方法について述べる．投影画像は円筒に投影された物体の模様を長方形に展開したものであり，断片画像を重ね合わせて投影画像を作成するためには，断片画像も平面に展開したものにしておく必要がある．ここで，物体の大きさがカメラと物体の距離に比べて小さいとして正投影を仮定すると，図 2.9 の点 P は画像上では点 P_p に映る．ここでは画像上での物体の中心位置を x 座標の原点として，正投影画像上での x 座標の値を x_p，断片画像上での x 座標の値を x_d とすると，物体の中心から θ のところにある点 P の x_p と x_d はそれぞれ以下のようになる．

$$x_p = r\sin\theta \tag{2.10}$$

$$x_d = r\theta \tag{2.11}$$

ここで，r は物体の半径であり，先に求めた物体の左右の輪郭位置をそれぞれ x_1, x_2 とすると以下の式で求まる．

$$r = \frac{x_2 - x_1}{2} \tag{2.12}$$

式 (2.10) と式 (2.11) から θ を消去すると，以下のように x_p と x_d の関係が求まる．

$$x_d = r\arcsin\frac{x_p}{r} \tag{2.13}$$

この式により，観測された画像から展開された断片画像を作成することができる．

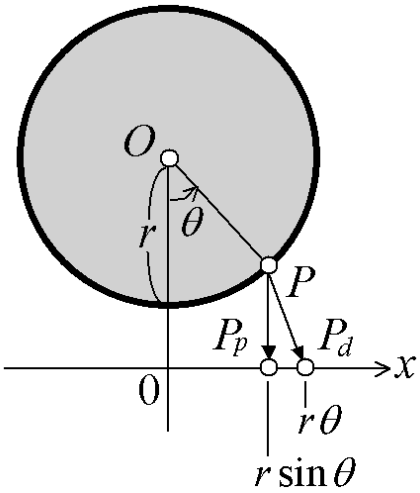


図 2.9: 正投影画像と展開画像の関係の模式図

瓶・ペットボトルについては高さ毎に半径が異なるので，以上の式を高さ毎に適用して断片画像を作成する．また，ペットボトルの内，四角型のものについては，円筒形に投影したものではなく，四つの平面を展開したものとして作成するため，8 方向から撮影した画像の内，面に対して正面から撮影した 4 枚の断片画像のみを用いる．

また，一般に物体をある方向から撮影した一枚の画像からは，物体中心から 180deg 部分が観測可能であるが，左右の端の方では歪みが大きく十分な解像度が得られないので，断片画像の作成には中心付近のみを利用する．具体的には，中心からそれぞれ 60deg，全体で 120deg 分の断片画像を作成する (図 2.10).

図 2.10: 作成された断片画像

断片画像上での部分的な色補正

一般に物体の反射光には，光源の色を直接反映する鏡面反射と物体の色を反映する拡散反射の 2 種類がある [64]．光源からの光を物体表面で鏡面反射してカメラに到達して観測されるものはハイライトと呼ばれ，模式的な図 2.11 では点 P 付近で強いピークを持つ．図 2.12(a) は，光沢のある物体をほぼ正面からの光源下で撮影したものであるが，中心付近に顕著なハイライトが生じていることが分かる．

一方，拡散反射については光源からの入射光に対して垂直な表面においてピークを持つが，鏡面反射のような強いピークは無く，ピークから離れるに従って徐々に暗くなっていくという性質を持つ．図 2.12(b) は，比較的光沢の少ない物体であるが，中心付近で最も明るくなっていて，そこから左右の端近づくに従って徐々に暗くなっていることが分かる．実際には，この 2 つの反射の重ね合わせが画像として観測されることになる．

前節では断片画像の作成方法について述べたが，その断片画像にはこのような反射の分布の影響による部分的な色の変化が残されたままになっている．しかし，投影画像はどの方向から見ても一様な照明条件であるものが望ましいので，このような部分的な色の変化を可能な限り除去する必要がある．以下では，その具体的方法について述べる．

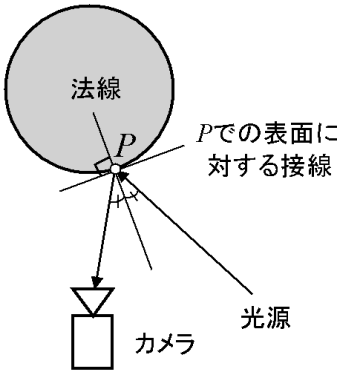
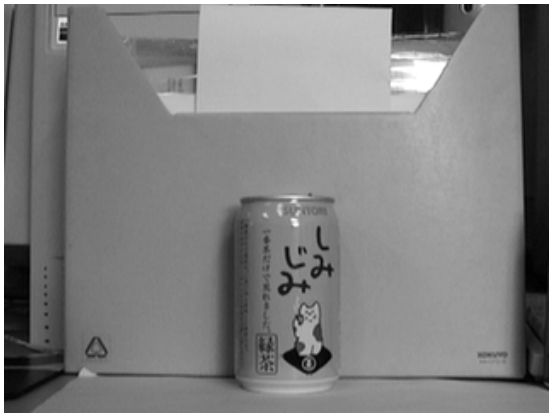


図 2.11: 物体表面での反射の模式図

(a) 光沢のある物体 (鏡面反射)



(b) 光沢の少ない物体 (拡散反射)

図 2.12: 物体表面での反射光の分布例

1. ハイライト候補領域の抽出

   ハイライト領域は光源の色を反映するが，一般的に光源の色は白に近いので，ハイライト領域はその他の領域に比べて輝度値が大きくなる．そこで，断片

画像上である程度の輝度値を持ち，かつ周囲 (今回は左右とした) より明るい部分をハイライトの候補として抽出する．具体的には，ある点 (x,y) における輝度値を $Y(x,y)$ と表したときに，以下の条件を満たす点 (x,y) の集合をハイライトの候補領域とする．

$$Y(x,y) > Y(x-d,y) \tag{2.14}$$

$$Y(x,y) > Y(x+d,y) \tag{2.15}$$

$$Y(x,y) > Y_{hl} \tag{2.16}$$

ここで，d は左右のどれだけ離れた点と輝度値を比較するかを表しており，極大値を持つ点に加えてその付近の明るい領域も候補として挙げるために，ここでは少し大きめな値として $d=6$ とした．また，Y_{hl} はハイライトとしての輝度値の下限値であり，実験的に $Y_{hl}=100$ とした．図 2.10 の断片画像に対して，この式により抽出されたハイライト候補領域を図 2.13 に示す．図を見ると，ハイライト以外のテクスチャの部分でもハイライトの候補として抽出されている部分があるが，これについては次節以降で述べる断片画像の位置合わせやモザイキングを行う際に候補を絞ることができるので，そちらを参照されたい．

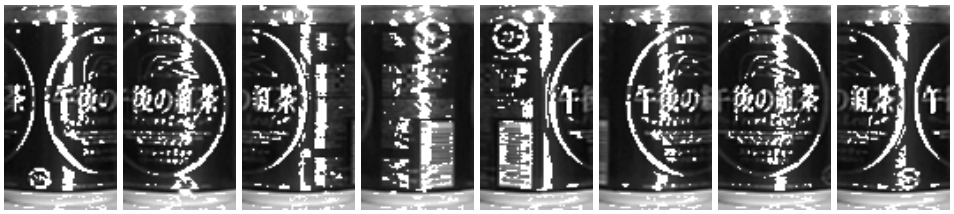


図 2.13: 抽出されたハイライト候補領域 (白領域)

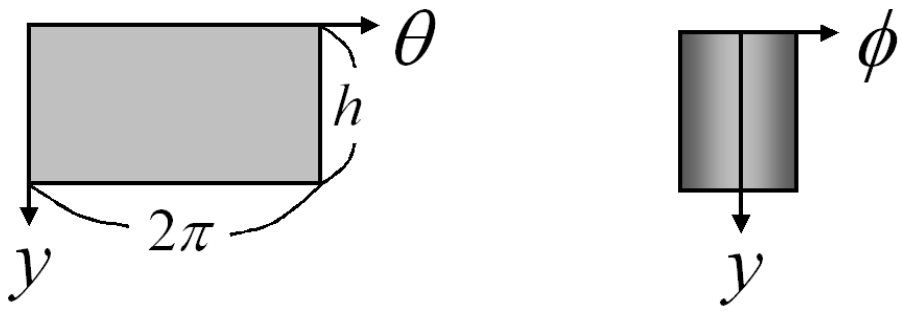
2. 拡散反射成分の照度分布による明るさの正規化

一様な照度分布の投影画像の輝度値を $I_p(\theta,y)$ とし (図 2.14(a) の座標系を参照)，その画像上のある θ を中心として見たときの一様でない照度分布の物体の画像の輝度値を $I_o(\phi,y;\theta)$ とする (図 2.14(b) の座標系を参照)．物体の画像は，投影画像に照度分布 $l(\phi)$ を掛けたものであるので，

$$I_o(\phi,y;\theta) = l(\phi)I_p(\theta+\phi,y) \tag{2.17}$$

が成り立つ．よって，$l(\phi)$ を求めることで照度分布の補正を行うことができる．まず，全ての方向から見た物体の画像の輝度値を用いて，見る方向 θ と高さ y に関して積分した $S(\phi)$ を計算すると以下のようになる．

$$\begin{aligned} S(\phi) &= \int_0^h \int_0^{2\pi} I_o(\phi,y;\theta)d\theta dy \\ &= l(\phi)\int_0^h \int_0^{2\pi} I_p(\theta+\phi,y)d\theta dy \end{aligned} \tag{2.18}$$



(a) 投影画像　　(b) 物体の断片画像

図 2.14: 投影画像と物体の断片画像の座標系

投影画像の周期性より式 (2.18) の積分項はいずれの ϕ に対しても投影画像全体を積分したもの，即ち定数となるので，$S(\phi)$ から $l(\phi)$ の分布が分かる．しかし，実際には物体の画像は有限の枚数しかないので，式 (2.18) の積分は以下のような離散的な計算で近似する．

$$\int_0^h \int_0^{2\pi} I_o(\phi, y; \theta) d\theta dy \simeq \sum_{i=1}^{H_p} \sum_{j=1}^{N_\theta} I_o(\phi, i\Delta y; j\Delta\theta)\Delta y \Delta\theta; \tag{2.19}$$

ここで，H_p は断片画像の高さ方向の画素数，N_θ は物体を撮影する方向の数であり，ここでは 8 方向から撮影するので $N_\theta = 8$ である．また，$\Delta y, \Delta\theta$ はそれぞれ y, θ の刻み幅である．この近似では，画像枚数が非常に少ないため，得られる分布には大きな誤差が含まれる．そこで，照度分布のモデルを導入して当てはめを行う．ここでは，光源のモデルとしてある一つの方向から到来する光による余弦項 (ランバートの余弦則による) と，壁などで反射した光の集合で物体全体を均一に照らす定数項 (周囲光) を考慮して，以下のような分布のモデルを考える．

$$\hat{l}(\phi) = a\cos(\phi + \phi_0) + b \tag{2.20}$$

ここで，a, b, ϕ_0 は分布のパラメタである．ここで，ハイライトの候補領域が抽出された方向を除いて，誤差 $\delta(\phi) = \hat{l}(\phi) - l(\phi)$ の自乗和を最小とするパラメタを求める．そして，物体表面上の最大照度で照らしたときの色に合わせるように補正係数 $(a+b)/(a\cos(\phi+\phi_0)+b)$ を各物体の画像の画素値に掛ける．

図 2.15 は，図 2.10 の断片画像を以上の手順で補正したものである．図より拡散反射成分の照度分布がほぼ一様になるように補正されていることが分かる．

断片画像の位置合わせ

これまでに作成した断片画像を用いて投影画像を作成するためには，まず断片画像同士を重ね合わせる位置を決める必要がある．最初に，図 2.16 のように 2 枚

図 2.15: 拡散反射成分の照度分布補正

の断片画像 Piece1,Piece2 の位置合わせを考える．高さ・幅が H_p, W_p の Piece1 に対して，x, y 方向にそれぞれ d_x, d_y だけずらして Piece2 を重ね合わせたとする．この時の重なり部分の内，ハイライト候補領域以外の部分において 2 枚の断片画像の色の差の平均を以下のように計算する．

$$S(d_x, d_y) = \frac{1}{A_c} \sum_{(x,y) \in R_c} D(\mathbf{c}_1(x,y), \mathbf{c}_2(x - d_x, y - d_y)) \tag{2.21}$$

ここで，$\mathbf{c}_1, \mathbf{c}_2$ はそれぞれ Piece 1，2 の色であり，R_c は重なり部分かつハイライト候補領域以外の領域であり，A_c はその面積である．結局，2 枚の合わせ位置 (d_x, d_y) は $S(d_x, d_y)$ を最小にする (d_x, d_y) として求める．これを，全ての隣り合う 2 枚の断片画像について行うことで，全体の断片画像の位置合わせを行う．

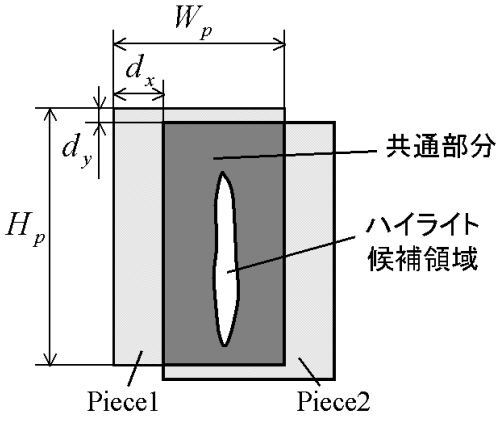


図 2.16: 断片画像の位置合わせの模式図

断片画像のモザイキング

断片画像の位置合わせができると，それの位置を元に断片画像をモザイキングして投影画像を作成する．モザイキングの方法は，以下の通りである．

- 2 枚の断片画像が重なっていない部分については，各断片画像の画素値をそのまま用いる．

- 2 枚の断片画像が重なっていて，ハイライトの候補領域でない部分については，2 枚の断片画像の平均値にする
- 2 枚の断片画像が重なっていて，ハイライトの候補領域である部分については，実際にその点での 2 枚の断片画像の色の差をユークリッドノルムで求めて，それが 50 以上の値であればハイライトであったと見て，輝度値の小さい方の色にして，そうでなければ平均値にする．

これにより，図 2.17 のような投影画像が作成される．

図 2.17: 断片画像の重ね合わせによって作成された投影画像

2.2.3　物体の特徴の登録

ここでは，先に述べた物体の特徴である実際の物体の大きさ，代表色，2 次特徴の登録方法について述べる．

画像上での物体の大きさは，カメラからの距離によって変化するので，実際の物体の大きさを登録しておく必要がある．本研究では，登録時にマニピュレータで既知の距離に物体を置くことができるので，あらかじめカメラをキャリブレーション [38][39] しておくと，その距離を用いて画像上の物体の大きさから実際の物体の大きさを求めることができる．また，回転非対称の物体は見る方向によって大きさが変化するので，物体の大きさに範囲を与えて登録しておく．

代表色や 2 次特徴も，見る方向に依存するので，ある方向から見たときに投影画像上で見える範囲を少しずつずらしながら特徴を抽出して，登録する．以下では，代表色及び 2 次特徴の登録方法について述べる．

区間の設定

投影画像から特徴を抽出する際には，その物体をある方向から見たときに見える範囲を設定する必要がある．ここでは，それを単位区間と呼ぶものとする．この単位区間は，本来見える範囲の 180deg 分を設定することも考えられるが，2.2.2 節でも述べたように，端の方では十分な解像度での観測が得られないため，特徴として用いるのは困難である．そこで，断片画像の作成のときと同様，物体を見ている方向を中心として，左右 60deg 分，全体で 120deg 分を単位区間として設定す

る．次に，この単位区間を全体でいくつ設定するかを決定する必要があるが，今回は 50 区間を設定した．これは即ち 7.2deg ずつ見る方向をずらして特徴を登録することに相当するが，代表色や 2 次特徴という特徴に対しては，7.2deg という角度は十分に細かいので区関数としては十分である．また，これらの特徴は 50 区間全てに対して個別に登録するのではなくて，ある範囲に渡って同じ特徴が観測される場合には，その単位区間をまとめて一つの区間として定義して登録しておく．例えば，図.2.18 のような物体があったとすると，その代表色としては，白色と青色の二つの区間 I_1, I_2 を登録することになる．これにより，50 の単位区間全てを登録する場合と比べて，登録データを縮小することができ，また認識の際にも 50 の単位区間全てとマッチングする必要が無くなり，より少ない区間との比較で済む．

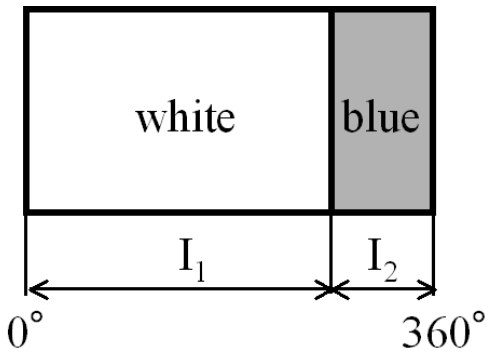


図 2.18: 区間の設定

2 次特徴の追加方針

本研究では，先にも述べたように，登録のデータを減らし認識を簡単にするために，特徴は必要最小限のものを登録しておくという方針を採る．よって，ある物体を登録するときに，同じ代表色を持つ物体が他に存在しなければ，物体の大きさと代表色のみを登録しておき，そうでなければ同じ代表色を持つ各物体に対して区別可能な 2 次特徴を登録する．

具体例を図 2.19 を用いて述べる．現在，代表色が黒の物体 A と代表色が白と淡灰の物体 B が登録されていたとする．この段階では，物体 A と B は代表色のみで区別が付くので，2 次特徴は登録しない．ここで，代表色が淡灰と濃灰の物体 C が新たに登録されたとすると，物体 B と C の淡灰の代表色の区間が区別不可能になる．そこで，各区間から 2 次特徴を抽出し，その特徴毎に区間を分割して登録する．もし，この 2 次特徴で区別可能であれば，特徴の追加作業を終了する．そうでなければ，区別の付かない区間に対して，更に 2 次特徴の追加作業を繰り返す．

このような 2 次特徴の追加作業を行うために，システムは特徴の情報のみではなく，投影画像そのものもデータとして蓄えておく．

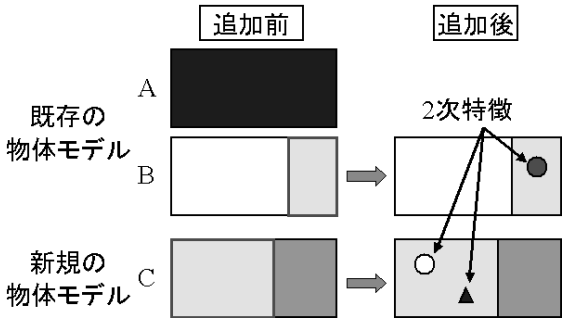


図 2.19: 2 次特徴の追加

代表色の登録

代表色の登録は，以下の手順に従って行う．

1. 投影画像を YIQ 成分のヒストグラムを用いて一様色領域に分割する (図 2.20(b)).
2. 投影画像上である方向から見たときの単位区間を設定して，面積が最大である領域の色を代表色とする．
3. 見る方向を 7.2deg ずらして単位区間を再設定して同様にして代表色を取得し，前の単位区間の色と似ていればその代表色の区間を延長して，そうでなければ新しい色の代表色の区間を作成する．
4. 残りのの単位区間全てについて手順 3 を繰り返し行う．

手順 1 については，付録 B の方法によって行う．ただし，ここで分割に YIQ 成分を用いたのは，輝度成分である Y と色成分である I,Q に分けた方が，RGB よりも明確なヒストグラム上の分割点が得られるためである．

手順 2 については，投影画像上での代表色の領域に含まれる色が，色空間上である程度のばらつきを持つので，代表色にある程度の範囲を与えて登録しておく必要がある．ここでは，投影画像上での代表色の領域内の色が RGB 空間内で正規分布すると仮定して，その範囲 R_{rep} を以下のように定義する．

$$R_{rep} = \{\mathbf{c} \in \mathbf{R}^3 | (\mathbf{c} - \mu)^T \Sigma^{-1} (\mathbf{c} - \mu) \leq k^2\} \tag{2.22}$$

ここで，$\mathbf{R}^3$ は RGB 空間，μ は領域内の画素値の平均，Σ はその共分散行列，k は信頼区間の大きさを決定するものであり，ここでは 3 次元正規分布の 90%の信頼区間に対応する値の 2.51 とする．

手順 3 については，比較する前の区間の代表色がはじめて作成されたときの範囲 R_{rep} を記憶しておき，現在の区間の代表色の平均がその範囲に入れば同じ区間として，そうでなければ新しい代表色とする．

以上の手順によって，図 2.20(c) のように代表色が登録される．

(a) 投影画像　(b) 領域分割された投影画像　(c) 登録された代表色 (下帯)

図 2.20: 代表色の登録

2 次特徴の登録

2 次特徴が必要な場合には，単位区間毎に以下のようにして登録する

1. 投影画像 (図 2.21(a)) の領域分割画像から代表色の領域を除いて，代表色以外の一様色領域を取得する (図 2.21(b)).
2. 得られた領域の中から，最低限の面積を持つものを選び，既に登録された特徴の色との差が大きい領域を 2 次特徴として選択する．

(a) 投影画像　(b) 代表色以外の一様領域

図 2.21: 2 次特徴の候補領域

手順 1 については，単純に連結領域を取得すると，物体に書いてある文字の部分などで，微小な多数の領域となってしまう (図 2.22(a)). また，このような領域は画像によっては結合されることもあるので，安定した領域として利用することが難しい．そこで，微小領域に関しては近傍に似た色の領域が無いかどうかを探し，存在すればその領域と結合を行う．これにより，図 2.22(b) のように文字の部分などが一まとまりの領域として得られる．

手順 2 については，最低限の面積を 10pixel とする．また，2 次特徴にはその領域の色・上下端・横幅・面積が含まれているが，ノイズ等を考慮してそれぞれにある程度の範囲を与えておく必要がある．色については代表色とまったく同様にして範囲を与えておく．それ以外の特徴については，その領域を 1 回膨張させたとき，1 回収縮させたときに変化する範囲を与えておくものとする．これは，領域の縁では色が遷移する領域であるので，認識時には領域に含まれることも含まれないこともあるために，このような範囲を設定した．図 2.23 は登録された特徴の例

(a) 結合前のラベリング画像　　(b) 結合後のラベリング画像

図 2.22: 近接領域の結合

である．この図では，上帯が代表色, 中帯が一つ目の 2 次特徴の色，下帯が二つ目の 2 次特徴の色を表す.

(上帯:代表色，中帯:一つ目の 2 次特徴，下帯:二つ目の次特徴 2)

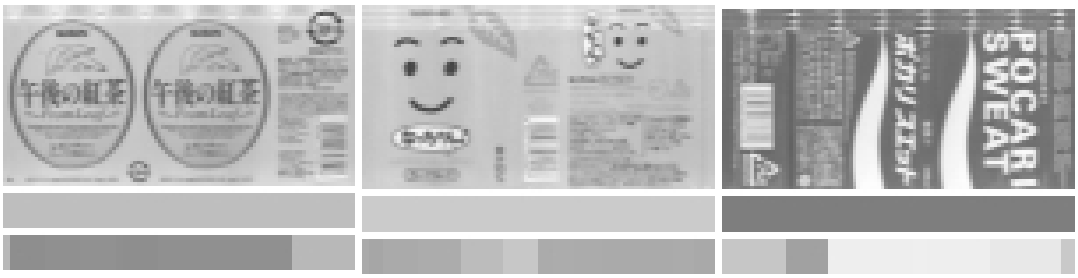
図 2.23: 2 次特徴の登録結果

また，その他の缶，瓶，ペットボトルについての投影画像と特徴の登録結果の例を図 2.24 に示す．投影画像の下にある帯の意味は図 2.23 と同じ意味である.

2.2.4　特徴にばらつきのある物体モデルの作成

果物等の特徴にばらつきのある物体 (図 2.25 参照) のモデルを作成する場合には，そのばらつきを考慮する必要がある．ここでは，その色や大きさを登録方法について述べる.

まず，第一回目の登録，即ちシステムが初めてその物体を見たときには，缶・瓶・ペットボトルの時と同様にして輪郭を抽出した後，1 回目の観測として大きさ (高さ h_1 と幅 w_1) と代表色 $\mathbf{c}_1$ を抽出して，それを物体モデルの特徴として登録しておく．その後，物体認識時に大きさや代表色の異なる同名の物体を観測した場合には，モデルとは一致しないために認識に失敗することがある．このような場合には，後の 2.4 節で述べる対話を用いた認識によって再度認識を試みる．そのようにして認識に成功したときに得られる物体の大きさ h_i, w_i と代表色 $\mathbf{c}_i$ を i 回目の観測値として記録しておき，物体の大きさの最小 h_{min}, w_{min}，最大 h_{max}, w_{max} と代表色の平均値 $\bar{\mathbf{c}}$，共分散 Σ を更新する．認識を行う度に以上の操作を行うことで，物体の大きさと代表色の特徴のばらつきを登録することができる.



(a) 缶

(b) 瓶

(c) ペットボトル

図 2.24: 投影画像と特徴の登録例

2.3 自動物体認識

本システムは，ユーザとの対話を用いた物体認識を行うが，必要以上にユーザを煩わせないように，登録された物体モデルを元にして，可能な限り自動的な物体認識を行う．ユーザから受け付けるものとして，物体名や製品名などを具体的に指定されたときにその物体を認識することと，ユーザが冷蔵庫内に何が入っているかを尋ねたときに，分かる範囲で全物体の認識を試みることの二つがある．後者については，ユーザがディスプレイを見ながら指令を出す場合には不必要であるが，将来的にディスプレイを使わないで音声対話のみで情報を伝える場合に必要となるので，今回ここで取り扱うこととした．いずれの認識においても，以下のような手順で行う．

1. 冷蔵庫領域を抽出する．

(a) 緑色系の蜜柑 (下の棚)

(b) 橙色系の蜜柑 (上の棚)

図 2.25: 色や大きさにばらつきを持つ物体の例

2. 物体の候補領域を抽出する.
3. 抽出された候補領域の種類 (缶・瓶・ペットボトル) を照合又は決定して，物体の領域を確定する.
4. 抽出された領域と物体モデルの照合を行う.

手順 1 については付録 A で述べられているので，そちらを参照されたい．以下では，残りの手順について説明する.

2.3.1 候補領域の抽出

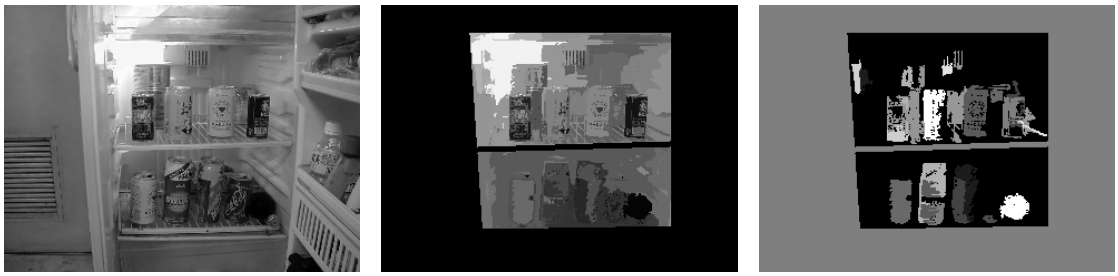
- 指定された物体の認識時
  ここではモデルの代表色の領域を候補領域として抽出するが，その際には物体モデルの持つ色の共分散を用いて，正規分布の 90%に入る領域を，原画像 (2.26(a)) の冷蔵庫内の領域より抽出する．しかし，このような色に基づく抽出方法では，図 2.26(b) のように物体の領域以外の不必要な領域を含むことがある．そこで，冷蔵庫の枠や扉に隣接している領域や物体としてはあまりにも大きすぎる領域を不適切な領域として削除すると，図 2.26(c) のような候補領域が得られる.

- 全物体の認識時
  原画像 (図 2.27(a)) の冷蔵庫内の全体の領域を 2.2.3 節と同様の方法で一様色領域に分割して (図 2.27(b))，先の場合と同様にして物体として不適切な領域を削除すると図 2.27(c) のような候補領域が得られる．全物体の認識方法としては，登録された全ての物体が指定されたと見て，全代表色に当てはまる領域を抽出する方法も考えられるが，登録された物体の数が多くなると時間

もかかり，かつ代表色が色空間の全体を覆ってしまうと，冷蔵庫内全体を抽出してしまう可能性があるため，今回は領域分割による候補領域の取得を採用した．

(a) 原画像　(b) 代表色の領域　(c) 候補領域

図 2.26: 指定された物体に対する候補領域



(a) 原画像　(b) 領域分割画像　(c) 候補領域

図 2.27: 全物体に対する候補領域

2.3.2　物体の種類の照合・決定

本論文では，大きく分けて缶・瓶・ペットボトル・果物の 4 種類の物体を取り扱っている．そこで，指定された物体の認識時には，抽出された候補領域を指定された物体の種類と照合を行い，全物体の認識時には抽出された候補領域がどの種類と一致するかを決定することになる．ここで，2 次特徴の照合よりも先に種類の照合・決定を行う理由は，2 次特徴には色だけでは無く，その上下端の位置や横幅も含まれていて，これらの照合には物体の正確な範囲を必要とし，その範囲が種類の照合・決定を行う際に後述のようにエッジ等を用いて求められるためである．以下では，各種類の物体の認識の方法について述べる．

缶の認識

缶の認識時には，まず側面や上下端の線分のエッジを抽出して，それらの線分で囲まれた長方形領域の大きさが，登録された物体の大きさと比べて適当かどうかをチェックする．ここで，物体モデルに登録されている実際の大きさから画像上での大きさに変換する必要があるが，あらかじめ与えておいた実際の冷蔵庫の棚の長さ w_{shelf} と画像上で抽出された冷蔵庫の棚の長さ $w_{i,shelf}$ の比 $m_{shelf} = w_{i,shelf}/w_{shelf}$ を用いて，以下のように近似的に変換するするものとする．

$$
\begin{aligned}
w_i &= m_{shelf} w \\
h_i &= m_{shelf} h
\end{aligned}
$$

これ以降，添え字 i をつけた各種物体の寸法は，この比 m_{shelf} を乗じて画像上の大きさに変換したものを表すものとする．

また，缶としての評価値は 1.0 を最大として条件を満たさない項目がある毎に減点していく方式で計算する．

最初に，缶のエッジの抽出方法について述べる．基本的にはモデル登録時と同様にして，図 2.28 のようにして領域に外接する長方形の端点 $S = (x_1, y_1, x_2, y_2)$ を取得して，その矩形に対して図 2.6 のような探索範囲を設定して，その探索範囲内でエッジ強度が最大になる位置として抽出する．

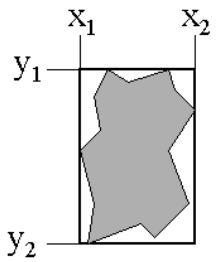


図 2.28: 領域の端点

しかし，缶が隣接して並んでいる場合には，探索範囲内に隣の缶のエッジを含むことがあるので，単純にエッジ強度が最大になる線分を探すと，隣の缶のエッジを抽出してしまうことがある．そこで，候補領域の平均色 $\bar{\mathbf{c}}$ とエッジの内側の色とのノルムを計算すると，指定した缶のエッジに対しては小さい値が得られるのに対して，隣の缶のエッジに対しては内側にその色が存在しない (背景色が存在する) ために大きな値が得られるので，このノルムを用いて判別することができる (但し，指定した缶の色と背景色が似ているときには，ノルムが常に小さくなるので，判別できないことがある).

この方法で缶の左端を抽出する場合の具体的な手順を以下に示す.

1. 図 2.6(a) にあるようなエッジの探索範囲をモデル登録時と同様に

$$\begin{aligned} S_l &= (x_{l1}, y_{l1}, x_{l2}, y_{l2}) \\ &= (x_1 - x_s, y_1, x_1 + x_s, y_2) \end{aligned}$$

として設定する．

2. $[x_{l1}, x_{l2}]$ に含まれる各 x に対して，縦エッジ成分の平均強度 $E_v(x)$ を式 (2.5) によって計算する．

3. $[x_{l1}, x_{l2}]$ に含まれる各 x に対して，エッジ強度を計算した x の 1 画素だけ缶の内側の $x+1$ における候補領域の平均色 $\overline{\mathbf{c}}$ との差の平均 $B_v(x)$ を計算する．$B_v(x)$ は具体的には以下のような計算式になる．

$$B_v(x) = \frac{1}{y_{l2} - y_{l1} + 1} \sum_{y=y_{l1}}^{y_{l2}} D(\mathbf{c}(x+1, y), \overline{\mathbf{c}}) \tag{2.23}$$

4. 集合 $\{x_b\}$ を

$$\{x_b\} = \{x | B_v(x) \leq B_{thresh}\} \tag{2.24}$$

として定義する．閾値 B_{thresh} はエッジの内側に缶があるかどうかを判断するための閾値であり，実験的に $B_{thresh} = 30$ とする．

5. 集合 $\{x_b\}$ 内からエッジ位置 x_{v1} を，

$$x_{v1} = \arg\max_{x \in \{x_b\}} \{E_v(x)\} \tag{2.25}$$

として取得する．但し，$\{x_b\}$ が空集合である場合や，$E(x_{v1})$ が閾値 E_{thresh} 以下である場合には，次の手順に進む．ここで，閾値 E_{thresh} は缶のエッジに必要な最低限のエッジ強度を表し，$E_{thresh} = 30$ とする．

6. 缶としての評価値を 0.1 下げて，閾値 B_{thresh}, E_{thresh} を $B_{thresh} = 30 \rightarrow 40$, $E_{thresh} = 30 \rightarrow 20$ というように少しずつ甘くして，同様の方法でエッジを抽出する．

7. それでも見つからなかったときには，さらに評価値を 0.1 下げて，缶の端の位置 x_{v1} を候補領域の端 x_1 の位置で代用する．

本論文では，縦線・横線の抽出にこのような方法を多用するので，以下の 4 つの場合について省略用の表記法を設定する．(記号については図 2.6 と上記の手順を参照とし，探索範囲を $S_s = (x_{s1}, y_{s1}, x_{s2}, y_{s2})$ とする．)

1. 縦線の抽出

   候補領域の平均色との差を考慮しない場合

$$x_v = V([x_{s1}, x_{s2}], [y_{s1}, y_{s2}], E_{thresh}) \tag{2.26}$$

   候補領域の平均色との差を考慮した場合

$$x_v = V_b([x_{s1}, x_{s2}], [y_{s1}, y_{s2}], E_{thresh}, B_{thresh}) \tag{2.27}$$

2. 横線の抽出

   候補領域の平均色との差を考慮しない場合

$$y_h = H([y_{s1}, y_{s2}], [x_{s1}, x_{s2}], E_{thresh}) \tag{2.28}$$

   候補領域の平均色との差を考慮した場合

$$y_h = H_b([y_{s1}, y_{s2}], [x_{s1}, x_{s2}], E_{thresh}, B_{thresh}) \tag{2.29}$$

横線の抽出については，条件を縦横に関して入れ替えたものである．

右側面や上下端のエッジについても同様の方法で求めることができる．左端も含めて具体的な条件を，先に述べた省略法を用いて表すと以下のようになる．

- 左端:$x_{v1} = V_b([x_1 - x_s, x_1 + x_s], [y_1, y_2], E_{thresh}, B_{thresh})$
- 右端:$x_{v2} = V_b([x_2 - x_s, x_2 + x_s], [y_1, y_2], E_{thresh}, B_{thresh})$
- 上端:$y_{h1} = H_b([y_1 - y_s, y_1 + y_s], [x_{v1}, x_{v2}], E_{thresh}, B_{thresh})$
- 下端:$y_{h2} = H_b([y_2 - y_s, \min\{y_2 + y_s, y_{d2} - 1\}], [x_{v1}, x_{v2}], E_{thresh}, B_{thresh})$
  (但し，y_{d2} は冷蔵庫領域の下端である．)

下端については，冷蔵庫の枠をエッジとして抽出しないように，探索範囲を y_{d2} より上側になるように設定している．そのためには，冷蔵庫の領域を抽出しておく必要があるが，その手法については，付録 A を参照のこと．

このようにして抽出された長方形領域の高さ $h' = y_{h2} - y_{h1} + 1$ と幅 $w' = x_{v2} - x_{v1} + 1$ を計算して，物体モデルから求まる高さ h_i と幅 w_i に対して以下の式を満たしている場合には，正しい缶領域として認識する．

$$|h' - h_i| < \max\{0.1h_i, h_{tol}\} \tag{2.30}$$

$$|w' - w_i| < \max\{0.1w_i, w_{tol}\} \tag{2.31}$$

ここで，h_{tol}, w_{tol} は最小限の許容範囲を表しており，実験的に $h_{tol} = w_{tol} = 3$ とする．

以上の手法を用いて缶を抽出した結果を図 2.29 に示す．この図は，上の棚の白い缶を指定したときの過程と認識結果を示したものである．図 2.29(b) は，候補領域である．図 2.29(c) は，左側面のエッジの抽出過程を示した図であり，それぞれ以下のように対応している．

- 黒色のグラフ:$E_v(x)$
- 白色のグラフ:$B_v(x)$
- グラフ上の上の方の灰色の線:E_{thresh}
- グラフ上の下の方の灰色の線:B_{thresh}

この画像では，指定した缶(白い缶)の左隣に青い缶が隣接して，エッジ強度$E_v(x)$については青い缶のエッジの方が大きい値になっている．しかし，青い缶のエッジは$B_v(x)$が閾値B_{thresh}を超えているので採用せず，$B_v(x)$が閾値を超えていない指定した缶のエッジの方を採用している．図2.29(d)は，右側面と上下端のエッジを同様にして求めた場合の缶の認識結果である．

冷蔵庫内の缶については，以上の方法で認識するが，冷蔵庫の扉にある缶については下側が隠れているので，見えている上側の部分を使って認識しなければならないので，物体の高さhの条件を扉の環境に合わせて小さく設定する．側面等のエッジの条件については，全く同様にして扱う．

(a) 原画像　(b) 候補領域(白領域)

(c) エッジの抽出　(d) 缶の認識結果

図 2.29: 缶の認識

瓶の認識

瓶の形状特徴は，図 2.3(b) のような胴体部分の 1 組の垂直線分と間隔の狭い首部分の 1 組の垂直線分である．そこで，候補領域に対して，以下の処理を行うことで瓶を認識する．

1. 領域の境界を境界追跡法 [40] によって求め，境界の情報から各境界点上での方向と曲率とを求める．方向を求める方法としては，現在の点の前後の点を結ぶベクトルの方向 (角度) を用いる方法があるが，境界は必ずしも滑らかに変化しているとは限らないので，スムージングをかけるために前後 5 ピクセル離れた点を結ぶベクトルの方向 (角度) を用いるものとする．曲率については，現在の点と前 5 ピクセルの点を結ぶベクトルと現在の点と後ろ 5 ピクセルの点を結ぶベクトルのなす角度を曲率とする．

2. 境界点から，方向が垂直方向に近くかつ曲率が小さい一連の点を垂直線分として抽出する．具体的には，方向に関しては $90 \pm 10[deg]$ または $-90 \pm 10[deg]$ の範囲に入り，曲率に関しては絶対値が $10[deg]$ 以下である点を垂直線分の一部とする．

3. 垂直線分から胴体部分と首部分のペアを選択し，それぞれの位置関係が瓶として適当かどうかをチェックする．実際には，得られた垂直線分の内，線分の重心が最も左側・右側にあるものを胴体部分のペアとして選択し，残りの線分から線分の重心が領域の重心に最も近いものを，首部分のペアとして左右両側から 1 本ずつ選択する．各線分が抽出された場合には，胴体部分の左右の線分の水平方向の平均位置 x_{b1}, x_{b2} と首部分の線分の水平方向の平均位置 x_{n1}, x_{n2} を求めて，そこから各々の部分の幅 $w'_b = x_{b2} - x_{b1} + 1$，$w'_n = x_{n2} - x_{n1} + 1$ を算出する．また，物体の高さ h' を候補領域の上端と下端の差 $h' = y_2 - y_1 + 1$ として求める．これらがモデル登録されたサイズに対して，以下の式を満たした場合に正しい瓶として認識する．

$$|h' - h_i| < \max\{0.1h_i, h_{tol}\} \tag{2.32}$$

$$|w'_b - w_{i,b}| < \max\{0.1w_{i,b}, w_{tol}\} \tag{2.33}$$

$$|w'_n - w_{i,n}| < \max\{0.1w_{i,n}, w_{tol}\} \tag{2.34}$$

以上の手順で条件を満たせば，瓶としての評価値を 1.0 としておく．しかし，これはあくまでも候補領域が瓶の全体に行き渡っていることを前提としているので，候補領域の中にラベルによって欠けた部分があると，垂直線分がうまく求まらずに瓶を認識できないことがある．

そこで，候補領域に対して，領域の左右の端 x_1, x_2，上下の端 y_1, y_2 を取得して，$[x_1, x_2]$ に含まれる各 x に対して，上端 y_1 から下端 y_2 に向かって候補領域に当たるまで走査していって，当たったらそれより下側を瓶の領域として埋めるというラ

ベルの補間処理を行う (図 2.30)．この補間処理を施した領域に対して，上で述べた方法を適用すると，瓶を認識することができる．

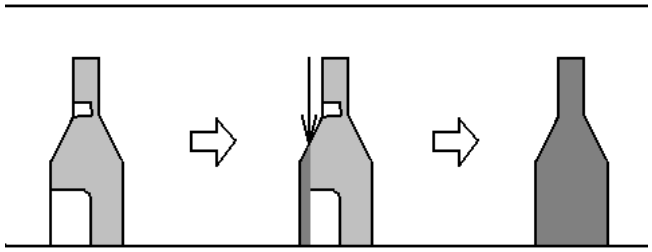


図 2.30: ラベルの補間

図 2.31 は，上記の方法を用いて瓶を認識した結果である．図 2.31(b) は候補領域であり，それに補間処理を施した領域が図 2.31(c) の黒領域である．また，図 2.31 (c) の境界の白線が垂直線分であり，それらの組み合わせによって図 2.31 (d) のような瓶が認識される．

ペットボトルの認識

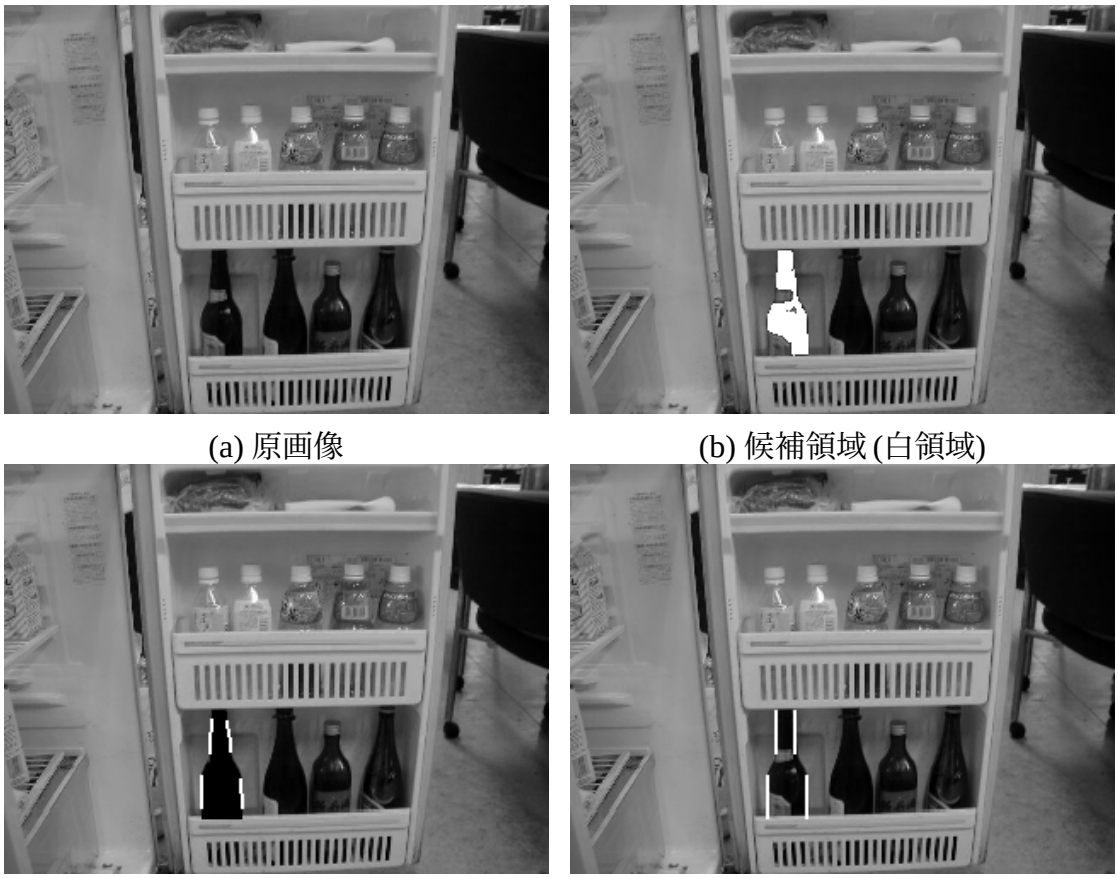
ペットボトルはラベルが候補領域として抽出されるので，残りの部分についてエッジを用いて抽出する．

1. ラベルの上方に探索範囲を設けて，エッジを用いてキャップを抽出する．まず，キャップ上下端を抽出するためにラベル上方に設定する探索範囲 $S_e = (x_{e1}, y_{e1}, x_{e2}, y_{e2})$ を，ラベルの端点 $S = (x_1, y_1, x_2, y_2)$ と冷蔵庫領域の上端 y_{d1} を用いて，以下のように設定する．

$$x_{e1} = g_x - w_{i,c}/2 - x_{tol} \tag{2.35}$$
$$x_{e2} = g_x + w_{i,c}/2 + x_{tol} \tag{2.36}$$
$$y_{e1} = \max\{y_1 - h_{i,t} - h_{i,c} - y_{tol}, y_{d1} + 1\} \tag{2.37}$$
$$y_{e2} = \min\{y_1 - h_{i,t} + y_{tol}, y_1 - 1\} \tag{2.38}$$
$$g_x = (x_1 + x_2)/2 \tag{2.39}$$

まずキャップの上下端の候補を抽出すために，$[y_{e1}, y_{e2}]$ に含まれる各 y に対して，横線としてのエッジ強度の平均 $E_{ch}(y)$ を，

$$E_{ch}(y) = \frac{1}{x_{e2} - x_{e1} + 1} \sum_{x=x_{e1}}^{x_{e2}} D(\mathbf{c}(x, y-1), \mathbf{c}(x, y+1)) \tag{2.40}$$



(a) 原画像　(b) 候補領域 (白領域)

(c) 補間処理 (黒領域) と垂直線分 (白線) の抽出　(d) 瓶の認識結果 (白線)

図 2.31: 瓶の認識

として求め，極大かつ閾値$E_{h,thresh}(=30)$以上となる横線位置$y_c(j)(j=1,2,3...)$に記憶しておく．次に，候補として求まった 1 つの横線$y_c(j)$とその 1 つ下の横線$y_c(j+1)$を各々キャップの上下端y_{c1},y_{c2}としておき，キャップの左右端x_{c1},x_{c2}を以下のように求める．

$$x_{c1} = V([x_{e1},\frac{x_{e1}+x_{e2}}{2}],[y_{c1},y_{c2}],E_{v,thresh}) \tag{2.41}$$

$$x_{c2} = V([\frac{x_{e1}+x_{e2}}{2},x_{e2}],[y_{c1},y_{c2}],E_{v,thresh}) \tag{2.42}$$

ここで，閾値は実験的に $E_{v,thresh}=20$ とする．これらのエッジに囲まれたキャップ領域の高さ $h'_c=y_{c2}-y_{c1}+1$ と幅 $w'_c=x_{c2}-x_{c1}+1$ が以下の式を満たせば，正しいキャップとして判断して，満たさなければ次の横エッジ候補について同様の処理を行う．

$$|h'_c-h_{i,c}| < \max\{0.1h_{i,c},y_{tol}\} \tag{2.43}$$

$$|w'_c-w_{i,c}| < \max\{0.1w_{i,c},x_{tol}\} \tag{2.44}$$

キャップが抽出されなかったら，ペットボトルとしての評価値を 0.2 下げておいて，探索範囲を以下のように設定し直して，再度探索する．

$$x_{e1} = g_x - w_{i,c}/2 - x_{ext,tol} \quad (2.45)$$
$$x_{e2} = g_x + w_{i,c}/2 + x_{ext,tol} \quad (2.46)$$
$$y_{e1} = \max\{y_1 - h_{i,t} - h_{i,c} - y_{ext,tol}, y_{d1} + 1\} \quad (2.47)$$
$$y_{e2} = \min\{y_1 - h_{i,t} + y_{ext,tol}, y_1 - 1\} \quad (2.48)$$

ここで，実験的に $x_{ext,tol} = y_{ext,tol} = 5$ とする．さらに，その際のエッジの閾値を甘く ($E_{h,thresh} = 20, E_{v,thresh} = 10$) しておく．それでも見つからなかったときには，ペットボトルの評価値を 0 にする．これは，キャップがペットボトルによって特徴となるものであり，比較的抽出しやすいものであるにもかかわらず抽出されなかったので，ペットボトルとして信頼できないからである．

2. エッジ情報を用いてラベルの下の部分の下端を抽出する．
具体的には，下端 y_b については冷蔵庫領域の下端 y_{d2} を用いて，

$$y_b = H([y_2 + h_{i,b} - y_{tol}, \min\{y_2 + h_{i,b} + y_{tol}, y_{d2} - 1\}], [x_1, x_2], E_{v,thresh}\} \quad (2.49)$$

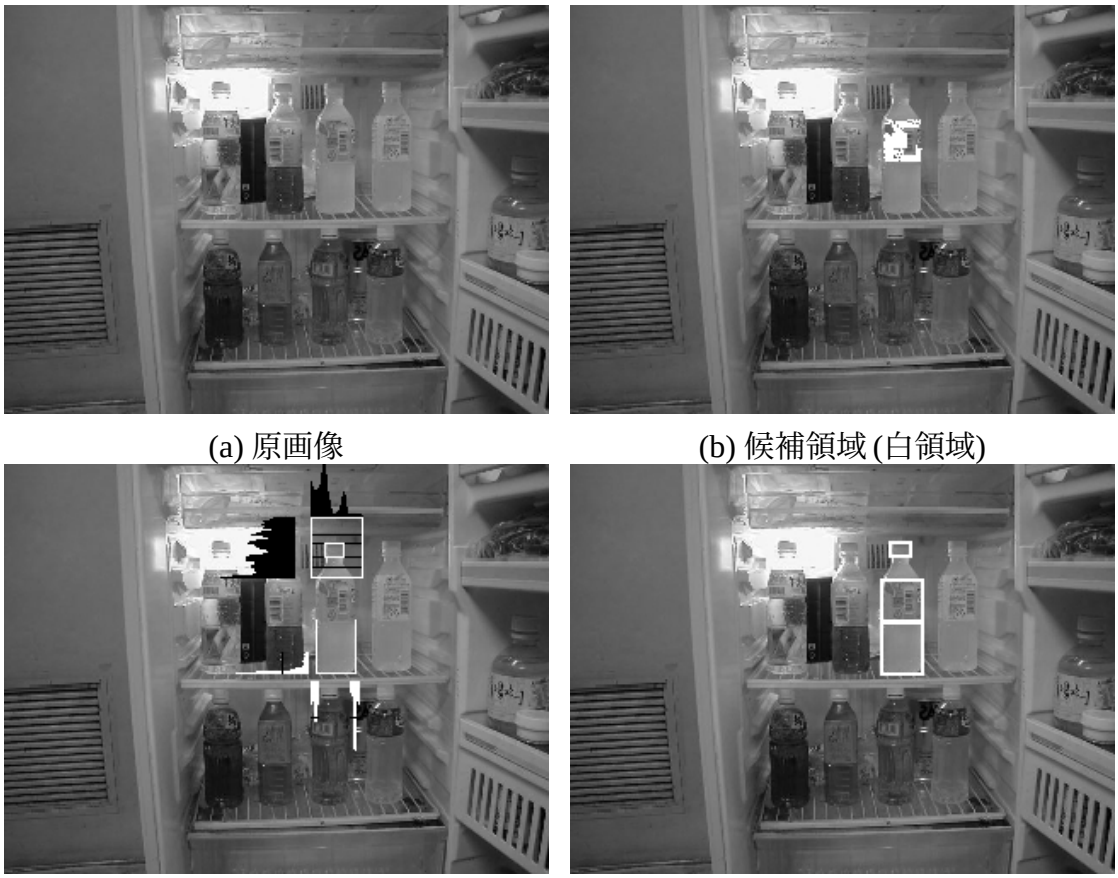
として抽出する．但し，ラベルの下の部分のエッジについては，水やスポーツドリンク等の透明な液体が入っているときには抽出できないときがあるので，キャップが見つからなかったときのようにペットボトルの評価値 0 にするのではなくて，評価値を 0.1 下げるだけにしておき，そのときの下端位置をモデルの値を用いて $y_b = \min\{y_2 + h_{i,b}, y_{d2} - 1\}$ としておく．

図 2.32 は，上記の方法を用いてペットボトルを認識したものである．ラベルの領域を求めると同図 (b) のようになり，ラベルの位置を基準として同図 (c) にあるようにキャップやラベルの下の部分のエッジを抽出する．ここで，キャップ付近にある大きい方の白矩形が探索範囲を表しており，付近の黒グラフが探索範囲内のエッジ強度を表しており，探索範囲内の白矩形が抽出されたキャップのエッジである．また，ラベルの下の容器部分付近にあるグラフが探索範囲内でのエッジ強度を表しており，容器部分の白線が抽出されたエッジを表している．これにより，同図 (d) のような認識結果を得る．

冷蔵庫内では以上の認識方法を用いるが，冷蔵庫の扉の環境ではラベルの下の部分が見えないことがあるので，ラベルの下の部分のエッジは抽出せずに，ラベルとキャップを抽出することで，ペットボトルを認識する．

果物の認識

果物は楕円体近似可能な物体について取り扱うため，画像上でハフ変換 [11] による楕円当てはめを行い形状としての認識を行う．ハフ変換では，以下の式のよ



(a) 原画像　(b) 候補領域 (白領域)

(c) キャップ等の抽出　(d) ペットボトルの認識結果

図 2.32: ペットボトルの認識

うに直立の楕円のみを扱い，楕円の中心 c_x, c_y と x, y 軸方向の軸長 a, b の 4 つのパラメタの投票空間を作成する．

$$\frac{(x-c_x)^2}{a^2} + \frac{(x-c_y)^2}{b^2} = 1 \tag{2.50}$$

ここで，各パラメタについて投票空間上での範囲 (最小・最大) を決定する必要があるが，まず x, y 軸方向の軸長の最小値 a_{min}, b_{min} と最大値 a_{max}, b_{max} については，果物のモデルとして登録された物体の高さと幅の最大・最小値を画像上での長さにそれぞれ変換した $h_{i,min}, h_{i,max}, w_{i,min}, w_{i,max}$ を用いて以下のように定まる．

$$a_{min} = \frac{w_{i,min}}{2} \tag{2.51}$$

$$a_{max} = \frac{w_{i,max}}{2} \tag{2.52}$$

$$b_{min} = \frac{h_{i,min}}{2} \tag{2.53}$$

$$b_{max} = \frac{h_{i,max}}{2} \tag{2.54}$$

また，楕円中心の x,y 座標 c_x,c_y の最小値 $c_{x,min},c_{y,min}$ と最大値 $c_{x,max},c_{y,max}$ は，各候補領域に果物の領域全体が含まれているとすると，候補領域の外接矩形の左上の点 (x_1,y_1) と右下の点 (x_2,y_2) を用いて以下のように定めることができる．

$$c_{x,min} = x_1 + a_{min} \tag{2.55}$$
$$c_{x,max} = x_2 - a_{min} \tag{2.56}$$
$$c_{y,min} = y_1 + b_{min} \tag{2.57}$$
$$c_{y,max} = y_2 - b_{min} \tag{2.58}$$

しかし，実際には果物の周辺付近の領域は影やハイライトによって領域を完全には抽出することが困難であるため，両軸に対して各々x_{tol},y_{tol} ピクセル分の余裕を持たせて以下のように定める．

$$c_{x,min} = x_1 - x_{tol} + a_{min} \tag{2.59}$$
$$c_{x,max} = x_2 + x_{tol} - a_{min} \tag{2.60}$$
$$c_{y,min} = y_1 - y_{tol} + b_{min} \tag{2.61}$$
$$c_{y,max} = y_2 + y_{tol} - b_{min} \tag{2.62}$$

また，ハフ投票空間では正しいパラメタの周辺で投票数の高い候補が多数存在するが，ここでは近接したパラメタ候補をグルーピングして，そのグループ中で最大の投票数を持つ楕円を採用することで候補を限定する．この際，候補として採用する投票数の最小値を定める必要があるが，これは楕円の大きさによる投票数の違いを考慮するために，各軸長 a,b に対して定まる楕円の周長の画素数の 20%以上の投票数を閾値とする．

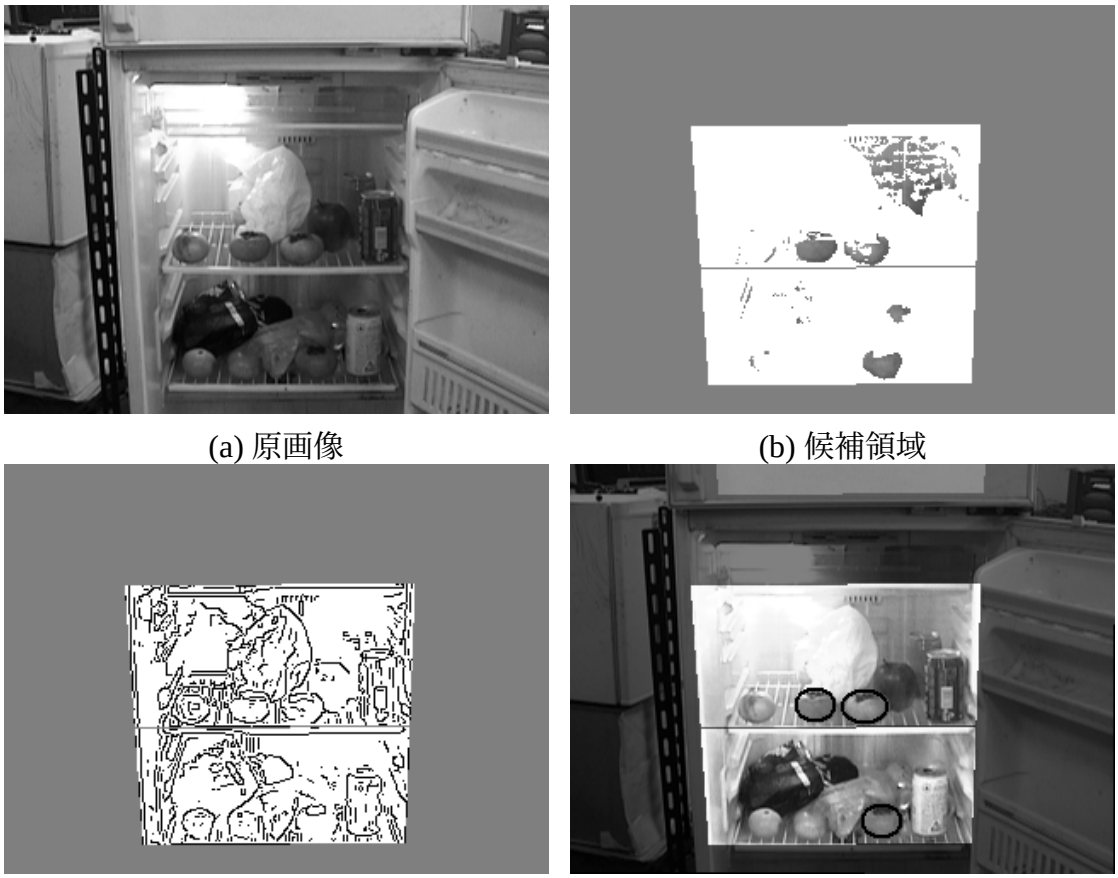
図 2.33 は，上記の方法を用いて果物を認識したものである．候補領域は同図 (b) のようになり，エッジ画像 (同図 (c)) を元にハフ変換により楕円のパラメタを求めると，同図 (d) のような認識結果を得る．

果物としての信頼性の評価値としては，検出された楕円上でのエッジ点を再抽出して，それが楕円周上で存在する割合を用いる．ここでは楕円近似による誤差をある程度考慮するために，エッジ点 (x,y) の内，以下の式を満たすものを楕円に属する点と考える．

$$\left| \frac{(x-x_c)^2}{a^2} + \frac{(y-y_c)^2}{b^2} - 1 \right| < \delta \tag{2.63}$$

ここで δ は誤差を表す値であり，実験的に $\delta = 0.1$ とする．以下に図 2.33 で検出された果物の信頼性を評価するために，楕円周上のエッジ点 (黒色) を再抽出した図 2.34(a) を示す．

さらにノイズによってエッジ点が途切れることや，楕円を構成しないエッジ点が抽出されることを考慮して，楕円周上で抽出されたエッジを周沿いの 1 次元座標に設定して，1 次元での膨張・収縮を行うことで，連続したエッジ点の楕円弧と



(a) 原画像　(b) 候補領域

(c) エッジ画像　(d) 楕円のハフ変換の検出結果 (黒)

図 2.33: 果物の認識

エッジの途切れた楕円弧に分けて，楕円周長に対するエッジ点の楕円弧の長さの割合によって評価する．ここで，膨張・収縮処理は以下のような手順で行う．

1. 膨張を n 回行う
2. 収縮を $2n$ 回行う
3. 膨張を n 回行う

ここでは実験的に $n = 2$ とする．この処理を行ったものが図 2.34(b) であり，エッジ点の楕円弧が黒で，エッジが途切れた楕円弧を白で表している．

形状評価に対する判断

種類の照合を行う際には，上記の評価値が閾値 (ここでは 0.7 とした) 以上であれば，照合されたものと見なす．また，種類の決定を行う際には，全種類に対する評価値を計算して，最も高い評価値が閾値以上であれば，その物体の種類とし

(a) 楕円周上で再抽出されたエッジ点 (黒)

(b) 楕円周上での連続したエッジ点 (黒) と途切れ (白)

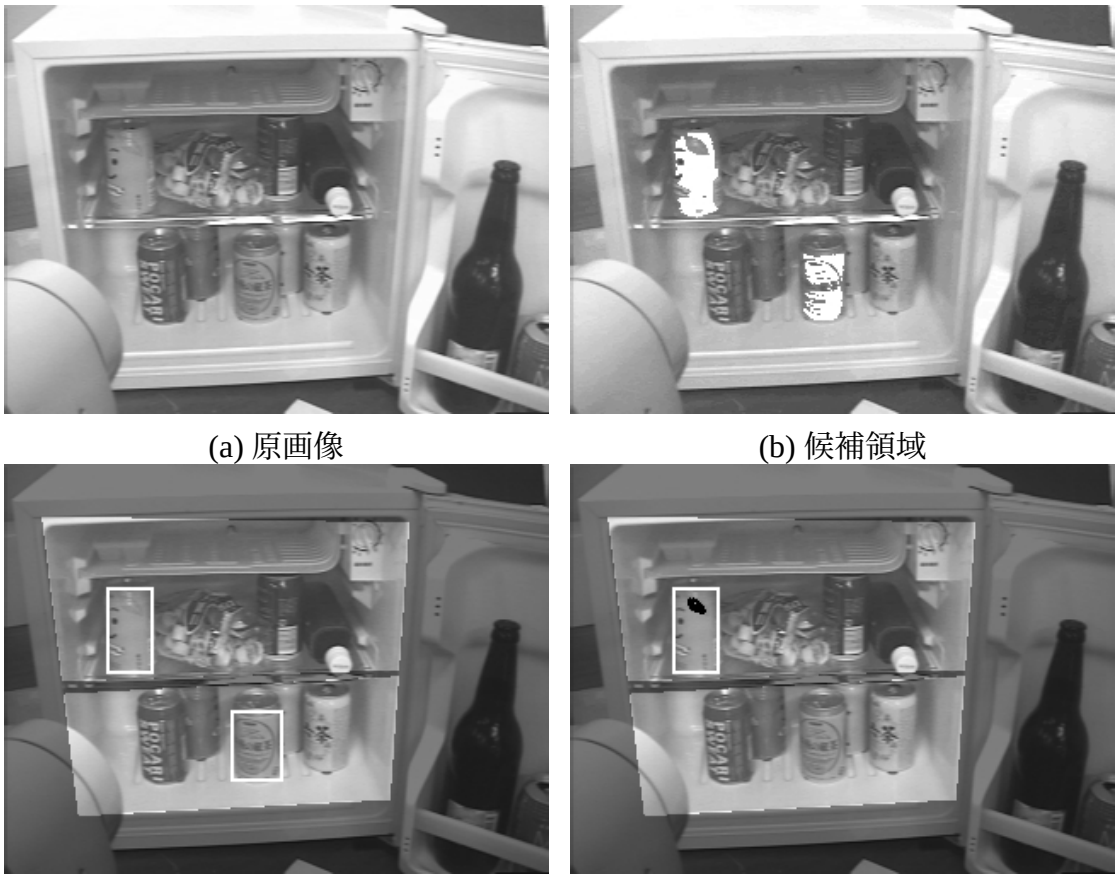
図 2.34: 楕円の検証

て決定する．いずれの場合も閾値を超えなかった場合には，物体の領域ではないものと見なす.

2.3.3　物体モデルとの照合

- 指定物体の認識時

  代表色については候補領域の抽出時に用いているので，ここで抽出された物体については既に指定された物体の代表色を含んでいるので，照合の必要性は無い．また，大きさについても形状の照合時に用いているので，これも照合の必要性はない．よって，ここでは 2 次特徴の照合を行えば良い．対象物体に 2 次特徴が登録されていなければ，代表色と大きさだけ他の物体と区別可能であるということなので，この時点で抽出された領域を指定物体として認識する．2 次特徴が登録されている場合には，その色に当てはまる領域を物体の範囲内から抽出して，登録時と同様に近接領域の結合を行った後，面積，横幅，上下端位置の照合を行い，照合されれば指定物体として認識する．もし，登録された 2 次特徴全てとマッチする物体が無かったときには，最も多くの 2 次特徴とマッチした物体の候補としておく．図 2.35 は物体モデルとの照合例である．この例ではユーザは左上のグレープフルーツを指定するが，図 2.35(b)(c) のように候補領域の抽出・種類の照合を行うと，右下のレモンティーも同じ代表色を持っているために，一緒に抽出されている．しかし，左上の缶からはグレープフルーツジュースの 2 次特徴が図 2.35(d) の黒色の



(a) 原画像　(b) 候補領域

(c) 缶の領域　(d) 抽出された 2 次特徴 (緑) と認識結果

図 2.35: 物体モデルとの照合例

領域のように抽出されるのに対して，右下の缶からは抽出されないので，左上の缶が指定された物体であると認識できる．

- 全物体の認識時

  この場合は，単純に一様色領域を候補領域としていて代表色との照合をまだ行っていないので，まず代表色が一致する物体を検索してくる．後は先の場合と同様にして，検索された物体について 2 次特徴との照合を行い，認識を行う．

また，物体モデルとの照合を行う際に，照合度合いの評価値を計算しておく．これは，もし指定した物体が複数発見された場合に，どれを優先的に持っていくか，また，候補が複数得られた場合に，どれを優先してユーザに提示するかを決める際に用いられる．ここでは色の類似度を用いて評価値を表すものとし，以下の計算式を与えるものとする．

$$S = \frac{w_0(1-d_{m0}) + \sum_{i=1}^{n} w_i(1-d_{mi})}{(w_0 + \sum_{i=1}^{n} w_i)} \tag{2.64}$$

$$d_{m0} = \frac{1}{k^2}(\mathbf{c}'_{m0} - \mathbf{c}_{m0})^T \Sigma_{m0}^{-1}(\mathbf{c}'_{m0} - \mathbf{c}_{m0}) \tag{2.65}$$

$$d_{mi} = \frac{1}{k^2}(\mathbf{c}'_{mi} - \mathbf{c}_{mi})^T \Sigma_{mi}^{-1}(\mathbf{c}'_{mi} - \mathbf{c}_{mi}) \tag{2.66}$$

$$w_0 = 1 \tag{2.67}$$

$$w_i = \frac{1}{1+i} \tag{2.68}$$

ここで，n は登録されている 2 次特徴の数，$\mathbf{c}'_{m0}, \mathbf{c}_{m0}, \Sigma_{m0}$ は代表色の観測値，モデルの平均値，モデルの共分散行列であり，$\mathbf{c}'_{mi}, \mathbf{c}_{mi}, \Sigma_{mi}$ は i 番目の 2 次特徴の観測値，モデルの平均値，モデルの共分散行列である．結局，d_{mo}, d_{mi} は各々の色のマハラノビス距離を示していることになる．また，k は信頼区間の幅に対応する係数で，90%の信頼区間に対応する値 $k = 2.51$ とした．よって，モデルの色に近い程 d_{m0}, d_{mi} の値は小さくなるので，逆に評価値の値は大きくなる．また，w_0, w_i は各特徴に対する重みであり，主な特徴ほど重みが大きくなるようにして定義している．

2.4 対話を用いた物体認識

2.3 節では，自動物体認識について述べたが，全ての場合において自動で認識できるわけではない．例えば，認識しようとした物体が奥の方に置かれたときに暗く映ってしまい，予想した色の範囲から外れてしまう場合や，別の物体に部分的に隠蔽された場合など，自動では認識が困難なケースが存在する．このような場合には，ユーザとの対話を通して，認識の補助となるような情報，例えば指定した物体が隠蔽されていることを得ることによって，その情報を元にした再認識を試みることで成功に結びつけることができる．以下では最初に認識結果に応じた対話について述べた後，各対話による物体認識方法について述べる．

2.4.1 認識結果に応じた対話

システムが指定物体を認識できないときや，認識結果をユーザに表示したところでユーザからの間違いの指摘があるときには，システムは対話を用いた物体認識を試みる．ここでは，どのような場合に対話を用いた物体認識を行うかを，図 2.36 のように認識結果に応じて場合分けして示す．

指定物体が 1 個認識された場合には，ユーザにそれが正しいかどうかを確認して，肯定されれば指定物体を確定してその物体を持って行く．また，指定物体が複数認識された場合は，ユーザに取って欲しい物体を選択してもらい，残りの提示物体の中に指定物体に当てはまらないものが無いかどうかを尋ねて，あれば物体モデルの訂正 (学習) を行い (詳細は 2.4.4 参照)，その後，指定物体を確定する．

指定物体が認識されなかった場合，ユーザの確認・選択時に指定物体が含まれていないということを伝えられた場合には，候補領域が存在するかどうかを確認

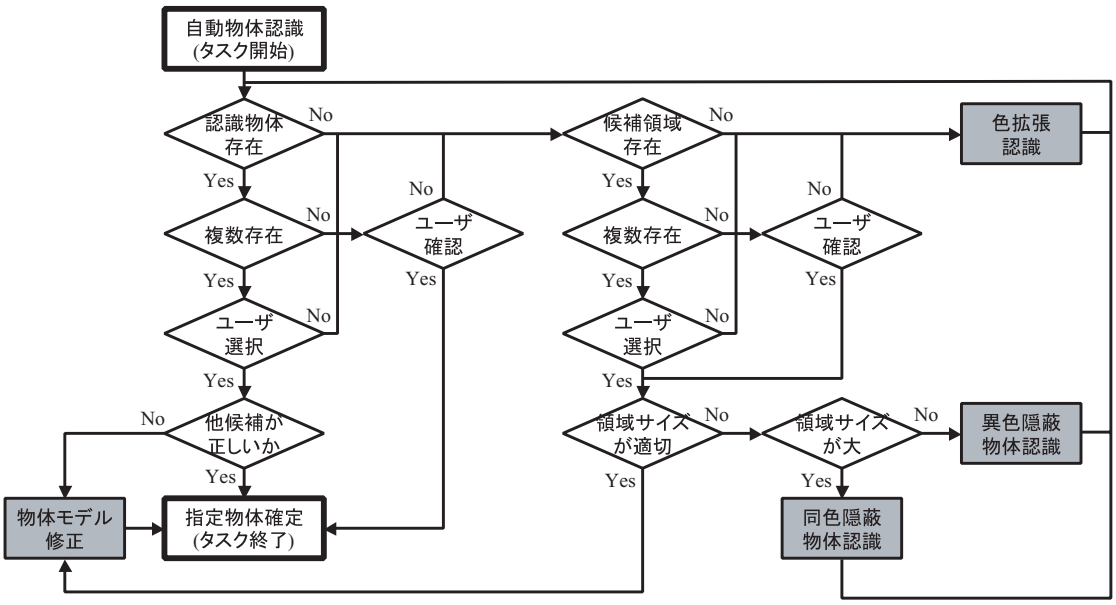


図 2.36: 認識結果に応じた処理の流れ

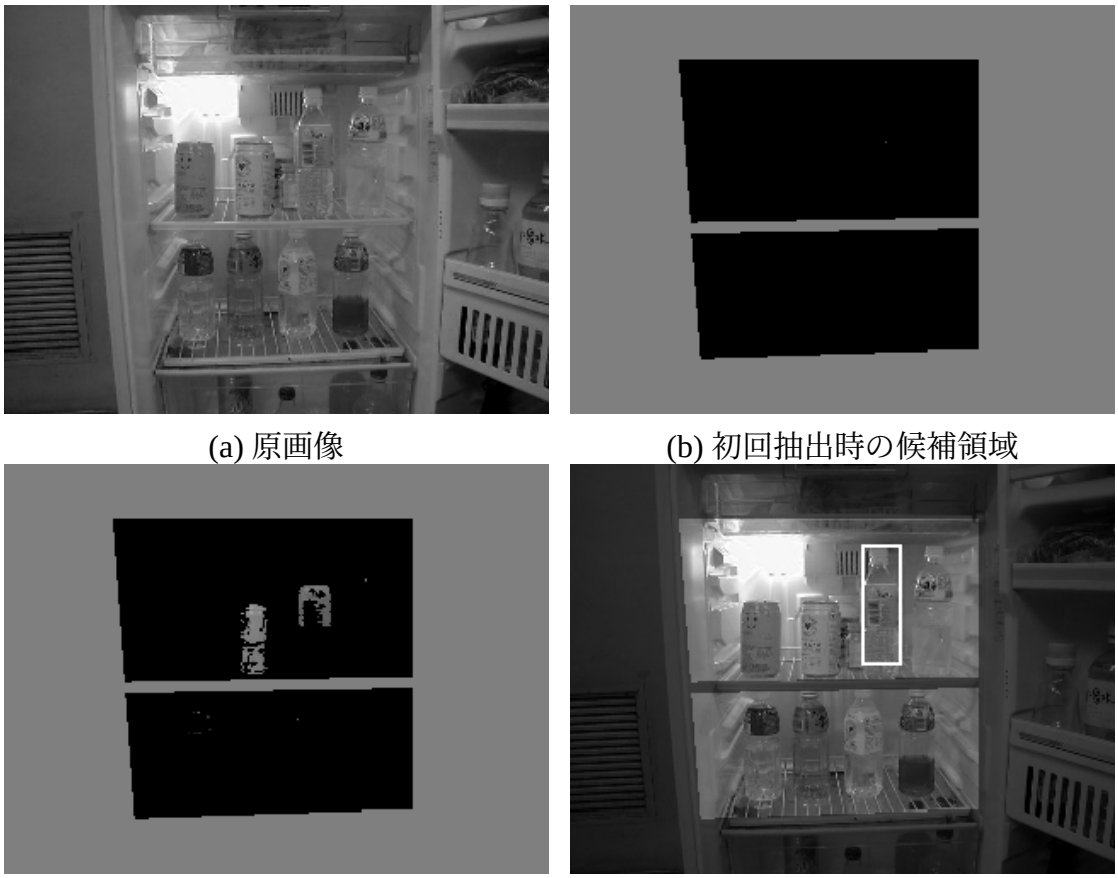
する．ここで，指定物体としては認識されずに候補領域止まりになった場合としては，以下の場合が考えられる．

- 候補領域サイズは物体として適切だが，物体モデルの色誤差によって照合に失敗した場合
- 候補領域サイズが物体として大きすぎた場合
- 候補領域サイズが物体として小さすぎた場合

そこで，候補領域が存在する場合には，それをユーザに提示して確認・選択をしてもらい，それがなされた場合には，候補領域サイズをチェックして処理を分岐する．

領域サイズが物体として適切な場合には，物体モデルの訂正 (詳細は 2.4.4 参照) 後に指定物体を確定する．領域サイズが大きすぎる場合には，同じ色の物体が重なっていて一つの大きな領域を形成している可能性があると考えて，ユーザに重なりに関する情報を求めて，重なっている物体の認識を試みる (詳細は 2.4.3 参照). また，領域サイズが小さすぎる場合には，異なる色の物体に隠蔽されている可能性があると見て，ユーザに隠蔽に関する情報を求めて，それを元に再認識を試みる (詳細は 2.4.3 参照),

候補領域も抽出されなかった場合や，候補領域の選択・確認時に指定物体が含まれていないことを伝えられた場合には，物体モデルの色誤差か色変換誤差 (第 3 章，第 4 章参照) によって指定物体の候補領域が抽出されなかったものと考え，色の範囲を調整して他の場所から再度認識を試みる (詳細は 2.4.2 参照).



(a) 原画像　(b) 初回抽出時の候補領域

(c) 再抽出時の候補領域　(d) 認識結果

図 2.37: 色の範囲を拡張したときの認識

2.4.2　色の範囲がずれている物体の認識

この場合には，物体モデルの色の範囲を広めに設定して，再抽出を試みることになる．具体的には，色の範囲で用いていた信頼区間を表す k の値を 90%に対応する 2.51 から，99%に対応する 3.37 に変えることで，より広い範囲の色の領域を抽出する．

図 2.37 は，この場合の例である．ここでは，水色のペットボトルを認識しようとしているが，このラベルの色が推定した代表色の範囲内に入らなかったために，一度目の抽出では図 2.37(b) のように候補領域が抽出されない．そこで，ユーザがもう一度よく探すように命令したときには，色の範囲を甘くする必要があると考えて，代表色の範囲を広げて抽出すると，図 2.37(c) のような領域が抽出され，最終的に図 2.37(d) のような認識結果が得られる．

2.4.3　部分的に隠蔽された物体の認識

異なる色の物体に隠蔽された物体の認識

ここでは，隠蔽している物体を A，隠蔽されている物体 (ユーザが最初に指定した物体) を B とする．システムが「(候補領域サイズが小さいので) この物体は他の物体に隠されているかもしれませんが，宜しいですか？」と尋ねると，ユーザ「A に隠されています.」や「A の後ろです.」といったような助言を与えてくれるものと考えられる．このような助言を受けた場合は，まず隠蔽している物体 A を今までに述べた通常の方法で認識して，続いて，隠蔽された物体を以下に述べる方法を用いて認識する．ここで，一般にある物体の後ろと指定された場合には，左後ろと右後ろの両方が含まれているので，隠蔽している物体の左右両側に対して隠蔽物体が無いかどうかを調べる必要がある．

隠蔽された物体が缶の場合は，隠蔽している物体 (端点を図 2.38 のように設定する) の左右に対して，隠蔽された缶の領域があるかどうかを調べる．最初に探索範囲を $S = (x_1 - 0.8w_i, y_1 - y_{tol}, x_2 + 0.8w_i, y_2)$ として，そこから隠蔽された物体の候補領域を抽出する．そして，その領域が存在した方に対して，エッジを抽出することで認識する．

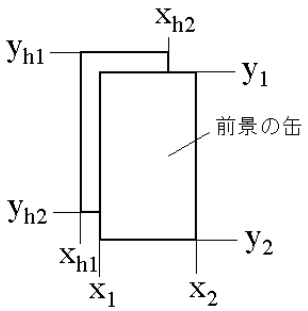


図 2.38: 隠蔽された缶のモデル

例えば左側にその領域が抽出されたとすると，側面のエッジ x_{h1} を，物体モデルの幅 w_i を用いて

$$x_{h1} = V([x_1 - 0.8w_i, x_1 - 0.2w_i], [y_1, y_2], E_{ocl,thresh}) \tag{2.69}$$

として抽出する．ここで，閾値は実験的に $E_{ocl,thresh} = 20$ とする．x_{h1} が抽出された場合には，続いて缶の下端のエッジ y_{h2} を

$$y_{h2} = H([y_2 - 12, y_2 - 2], [x_{h1}, x_1], E_{ocl,thresh}) \tag{2.70}$$

として抽出し，それも抽出されれば，上端 y_{h1} を物体モデルの高さ h_i を用いて，

$$y_{h1} = H([y_{h2} - h_i - y_{tol}, y_{h2} - h_i + y_{tol}], [x_{h1}, x_1], E_{ocl,thresh}) \tag{2.71}$$

として抽出する．もう一方の側面の位置 x_{h2} は，図 2.38 のように，上端のエッジ y_{h1} が隠蔽している物体の上端 y_1 より上側にあれば，

$$x_{h2} = V([x_{h1} + w_i - x_{tol}, x_{h1} + w_i + x_{tol}], [y_{h1}, y_1], E_{ocl,thresh}) \tag{2.72}$$

として抽出するが，抽出されなければ $x_{h2} = x_{h1} + w_i$ で代用する．

右後ろにその領域が抽出された場合も同様にして認識して，両方ともに存在する場合にはそれらをユーザに候補として示して選択してもらう．

例として，図 2.39(a) から MIU の後ろにある午後の紅茶を認識することを考える．図 2.39(b) は通常の方法で抽出された MIU の領域である．図 2.39(c) は隠蔽している物体付近から指定物体の候補領域を抽出したものである．ここでは，左側に抽出されたので，そちら側でエッジを抽出することで，図 2.39(d) のような認識結果を得る．

(a) 原画像　(b) 隠蔽している物体の認識

(e) 探索範囲内の候補領域　(f) 認識結果

図 2.39: 隠蔽された缶の認識

同じ色の物体に隠蔽された物体の認識

この場合には，システムが「(候補領域が大きいので) 複数の物体が重なっていませんか？」と尋ねると，ユーザから「はい，重なっています.」というような確認の返事が得られるものと考えられる．そこで，まず候補領域の端点 $S = (x_1, y_1, x_2, y_2)$ を，図 2.40 のように取得する．ここで左右のいずれの物体が隠蔽している物体 (手前の物体) であるかを知るために，候補領域の下端を辿ってみて，右側の方がより下にあれば右側が隠蔽している物体，左側の方がより下にあれば左側が隠蔽している物体であるとみて，物体を分けるエッジを抽出する．ここで左右のどちら側がより下にあるかということについては，下端の点列に直線を当てはめて，下がっている側が下にあるものとして判断する．

ここでは，図のように左側の物体が隠蔽されていると見た場合を考えて，隠蔽している物体の側面のエッジ x_{r1} と上端のエッジ y_{r1} を，

$$x_{r1} = V([x_2 - w_i - x_{tol}, x_2 - w_i + x_{tol}], [y_1, y_2], E_{same,thresh}) \quad (2.73)$$

$$y_{r1} = H([y_1 + 3, y_1 + 12], [x_{r1}, x_2], E_{same,thresh}) \quad (2.74)$$

として抽出する．ここでは，エッジ強度の閾値を小さい目 ($E_{same,thresh} = 15$) に設定する．これは，そもそも隠蔽の生じた物体同士の間に，はっきりとしたエッジが無かったために一つの領域になってしまったことから，エッジ探索の際の閾値をあまり厳しくすると，エッジが抽出できないからである．次に，隠蔽された物体の下端のエッジ y_{l2} と側面のエッジ x_{l2} を

$$y_{l2} = H([y_2 - 3, y_2 - 12], [x_1, x_{r1}], E_{same,thresh}) \quad (2.75)$$

$$x_{l2} = V([x_1 + w_i - x_{tol}, x_1 + w_i + x_{tol}], [y_1, y_{r1}], E_{same,thresh}) \quad (2.76)$$

として抽出する．

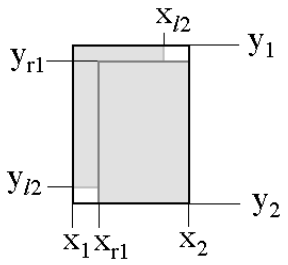


図 2.40: 同色物体が隠蔽された場合のモデル

図 2.41 に同色物体による隠蔽の例を示す．ここでは，下の棚にある午後の紅茶が重なっているものを認識することを考える．候補領域を抽出すると図 2.41(b) のように二つの物体が一つの大きな領域として抽出されるので，図 2.41(c) のような最終候補を提示することになる．そこで，ユーザから 2 つの物体が重なっていることを教えられると，図 2.41(b) の候補領域の下端を取り出して直線を当てはめて

みる (図 2.41(d) の黒線) と，右下がりになっていることが分かる．そこで，そのような位置関係での物体を分けるエッジを抽出すると，図 2.41(e) のような認識結果を得る．ここで，実線が隠蔽している物体で，破線が隠蔽されている物体を表す．

2.4.4　対話によるモデルの学習

2.4.1 節で述べたような認識の失敗のいくつかは，ある物体 A の特徴の色の範囲 (R_A とする) が推定された範囲外や別の物体と重なる範囲等の別の範囲 (R_I とする) にシフトしたことが原因である．ここで，シフトの原因として登録された物体の特徴の色そのものが誤っている場合と，色変換誤差による場合 (第 3 章を参照) の二通りがある．前者の場合は，周囲の物体と比較して，その物体だけが推定された色から外れている場合であり，後者は周囲の物体も全て同じように色がずれている場合である．

前者の場合には登録された特徴の色の範囲 R_A を修正する必要がある．そこで，システムは最初に R_I を含めるように R_A を修正して，R_A を含む区間全てをモデルから検索して修正する．次に，この修正によって区別の付かなくなる物体を検索して，あれば 2.2.3 で述べた方法と同様にして，特徴を追加する．

後者の場合には，現在観測されている冷蔵庫内での物体の色を使って，新たに色変換の推定パラメタや色変換誤差を計算し直す．その際，今回得られたデータの重みを設定する必要があるが，それは新たに計算された推定パラメタでうまく色が変換できるようにそれを設定する必要がある．具体的にどの様に設定するかは，今後の課題である．

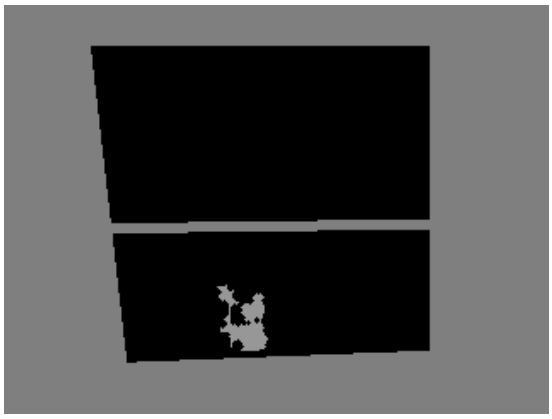
2.5　ユーザにとって選択しやすい認識結果表示法

ユーザにとって使いやすいインタラクティブビジョンシステムを構築する際には，認識結果をユーザに分かりやすく伝えることが重要になる．例えば，複数の候補からユーザが選択する場合に，その候補をどのように表示するとユーザにとって選択しやすいか，ということについて考える必要がある．特に，認識結果の物体候補が非常に多い場合には，最初に大まかな領域を選択してもらうことで，一回当たりに選択する候補数を減らす階層的な表示方法を用いる必要もある．本節では，最初に 2.5.1 でユーザにとって見やすい候補の表示方法について述べ，次に 2.5.2 で候補が多数存在するときに一回あたりの選択肢を減らすための階層的選択について述べる．

2.5.1　認識結果の候補の表示方法

認識結果の候補を表示する際には，以下の点について考慮する必要がある．

(a) 原画像



(b) 候補領域

(c) 候補物体の表示

(d) 下端のチェック

(e) 認識結果

図 2.41: 隠蔽された同色の物体の認識

1. 領域同士を区別するためのラベル付け
2. 領域の表示順序

領域同士を区別するためのラベル付け

システムが認識結果として表示している領域をユーザが区別して指定する際に，図 2.42(c) のような領域のラベル付けをしない単純な結果表示をした場合，ユーザは「下の棚の左から二番目のものを取って」や「柿の左上にある蜜柑を取って」というように領域を指定するためには場所の情報を与える必要がある．このような選択の仕方は，ユーザにとって負荷が大きく，かつ音声対話システムにとっても想定外の発話による誤認識の恐れがあるため，問題となる．そのような場合には領域を区別するためのラベルを付け，場所以外の情報でも領域を指定できるようにすることが望ましい．

ラベル付けの一つの方法としては，同図 (d) にあるように領域を色分け表示することが考えられる．この場合，ユーザに対して「何色で囲まれた領域ですか？」と尋ねることで,「青色」や「赤色」のような色名で選択してもらうことが可能になる．但し，色表示の場合は, 同じ色に対しても複数の指定方法が存在し，ユーザによっては色名の指定方法が異なる場合がある．例えば,「緑」のことを「青」と呼ぶ場合や,「水色」のことを「シアン」と呼ぶ場合などが考えられ，これらに対応するためには音声対話システムにあらかじめ様々な色の言い方を登録しておく必要がある．また表示候補数が増えてくると普段あまり使わない「薄紫色」や「深緑色」といったような色名での選択をユーザに要求することになり，ユーザにとっては負担が増加する．

別の方法としては，同図 (e) にあるように領域の近くの空き領域に番号を表示することが考えられる．この場合，ユーザに対して「何番の領域ですか？」と尋ねることで,「2 番」や「3 番」のように番号名で選択してもらうことが可能になる．色の場合とは異なり数字の指定方法については全ての人に共通であり，かつ候補数が増えても番号自体は普段使う数字の範囲内で収まるため，ユーザにとって指定しやすいものとなる．しかし，表示候補数が増加して密集しているような場合には，番号同士が重なってしまうことや番号と対応との関係が一見して分かり辛くなる (例えば，同図 (e) の 6 番の領域が左の楕円か左下の楕円のいずれを示しているかが分かり辛い) という問題がある．

これらに対して，同図 (f) のように領域を色分け表示した上で，その色と同じ色で番号を表示することで，領域と番号の対応付けを明確にするという方法がある．この場合は，色はあくまでも対応付けのための補助として用いられているため，色名指定の曖昧さ等の問題は避けることができる．しかし，表示する色によっては番号が見辛くなるという問題点がある．

よっていずれの方法にしても一長一短が存在する．

(a) 原画像

(b) 正規化画像

(c) 単純な結果表示

(d) 色分け表示

(e) 番号付け表示

(f) 色分けと番号付けによる表示

図 2.42: 表示領域同士を区別するためのラベル付け

領域の表示順序

先に図 2.42 で挙げたような認識結果をユーザに対して静的に表示すると，原画像の一部が領域の枠や番号によって隠されたままになるため，システムが表示している範囲が本当に正しいかどうかをユーザが判断することが困難になる．そこで，原画像 (図 2.42(a)) と認識結果画像を交互に表示して，ユーザの確認を補助することを考える．ここで，暗い照明条件での原画像を表示するとユーザによる確認が困難になるため，後に述べる色変換を用いて基準照明へ変換した正規化画像 (図 2.42(b)) を表示する．

このように画像を点滅させて表示させる場合，図 2.42 に示したように認識結果の候補を同時に表示する方法に加えて，各領域を順次表示する (図 2.43 の (a)〜(h) を順次表示する) という方法も考えられる．この場合は，先の番号付けの問題点で取り上げられた領域と番号の対応付けの問題点は無くなるという利点がある．但し，同時に全ての候補を表示する場合と比較して，全候補を表示し終わるまでの時間がかかるため，ユーザによっては煩わしく感じてしまうという欠点がある．この点を改善するために表示切り替えの間隔を短くするという方法も考えられるが，各候補が一瞬しか表示されないということになり，ユーザが一巡目の表示で候補を確認しきれずに，二順目以降で確認する場合があり，結局時間がかかってしまうということも考えられる．特に，体の不自由な年配の方が本システムを利用することを考えると，あまりに速く移り変わる表示はユーザが当惑する原因ともなる．

そのようなユーザのために，一つ一つ確実に確認できる方法としては，図 2.43(a) と同図 (b)〜(h) の候補の一つとを交互に表示して，ユーザに対して「これですか？」というように逐次確認するという方法がある．この方法であれば一つの候補に対してのみ注目してそれが正しいか正しくないかということのみを評価すれば良いため，一回あたりの選択にかかる労力は非常に小さい．但し，当然ながら同時表示，順次表示に比べて圧倒的に時間がかかるという点が問題となる．

表示方法に関する主観テスト

上記で述べた表示方法が，実際にユーザにとってどのように感じるかを確かめるために，図 2.42 の画像と選択候補を用いて，被験者 6 人に対して表 2.3 の 9 通りの表示方法に関する主観テストを行った．

また，各表示手法のテスト順序による依存性を軽減するために，被験者には予め 9 通り全ての表示方法を見てもらい，どのような表示方法があるかを知った上で，もう一度最初から順に各表示方法の主観テストを行い，評価を受けた．評価には表 2.4 に示す 11 段階の絶対評価を用いた．

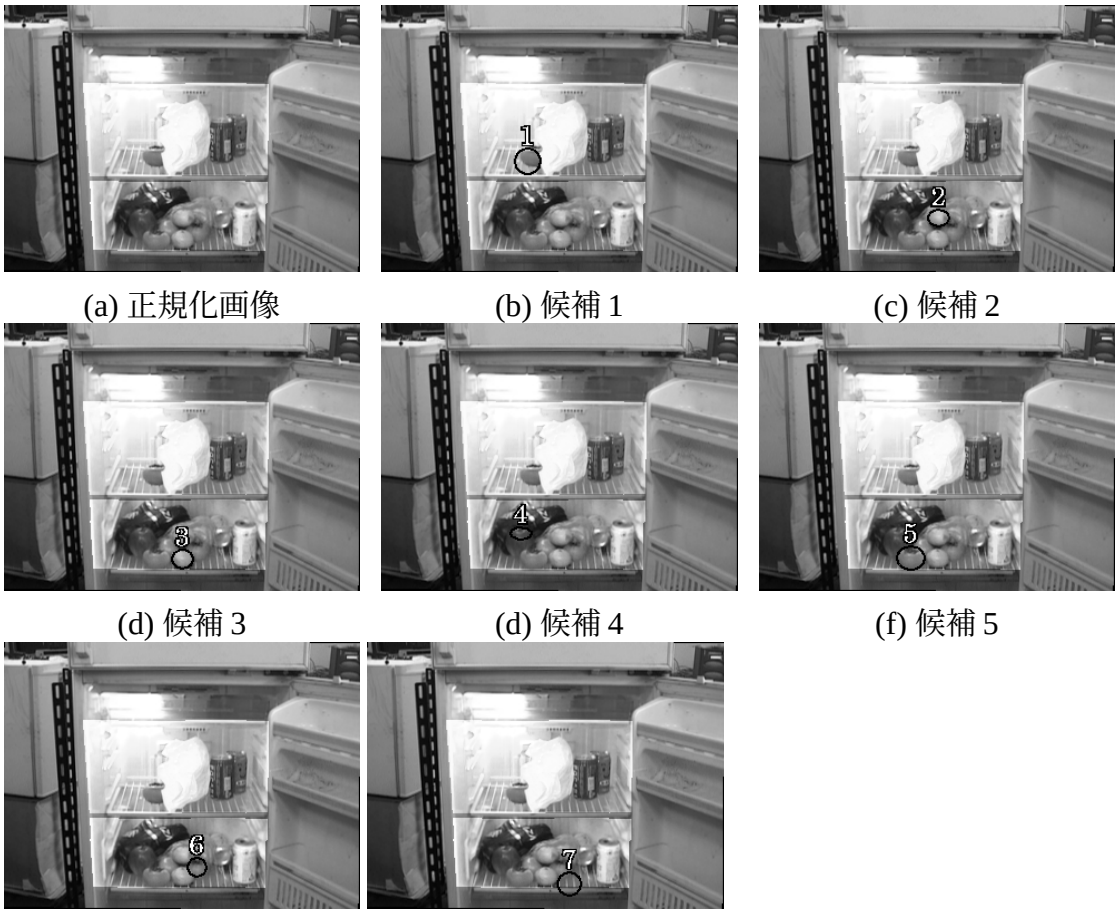
主観テストの結果を図 2.44 に示す．同図 (a) は各被験者の評価であり，同図 (b) はそれをまとめて最大値・最小値と平均を表したものである．被験者によって多少の個人差が存在するが，表示方法としては 4. 色分けと番号付けの両方をして同時

表 2.3: 主観テストに用いる表示方法

	ラベル付け	表示順序
1	無し	同時
2	色分け	同時
3	番号付け	同時
4	色分けと番号付け	同時
5	無し	順次
6	色分け	順次
7	番号付け	順次
8	色分けと番号付け	順次
9	無し	逐次確認

表 2.4: 主観テストに用いた 11 段階の絶対評価

評価値	評価内容
0	極めて選び難い
1	かなり選び難い
2	選びにくい
3	やや選び難い
4	どちらかといえば選び難い
5	普通
6	どちらかと言えば選び易い
7	やや選び易い
8	選び易い
9	非常に選び易い
10	極めて選び易い



(a) 正規化画像 (b) 候補 1 (c) 候補 2

(d) 候補 3 (d) 候補 4 (f) 候補 5

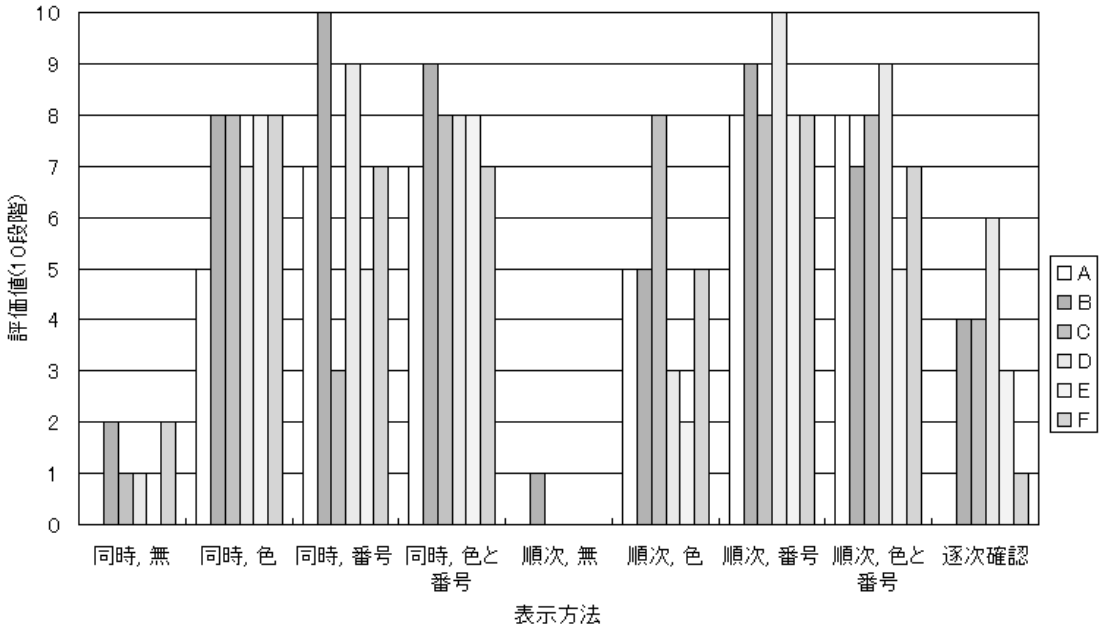
(g) 候補 6 (h) 候補 7

図 2.43: 候補の順次表示

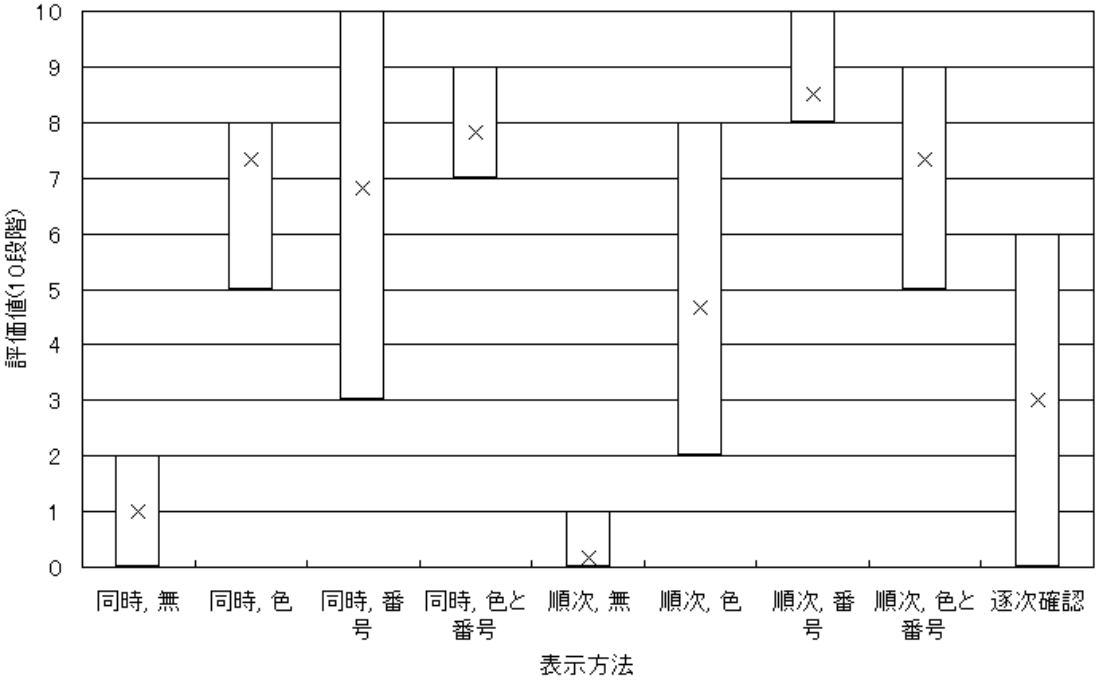
表示や 7. 番号付けして順次表示等が安定して高い評価を受けている．よって，システムとしてはこれらの高い評価を受けた表示方法をいくつか実装しておき，ユーザ毎の好みで表示方法を選択できるようにすることが望ましい．

2.5.2 認識結果の階層的選択

2.5.1 では，表示すべき候補が決まっている場合の表示方法について扱ったが，表示すべき候補数があまりに多い場合にはどのような表示手法を用いてもユーザの選択の負担は大きいものとなる．特に大きさや色にばらつきのある果物を認識する場合には，図 2.45(b) のように他の物体も含むような代表色領域を抽出することがあり，その結果として同図 (d) のように候補が多数になってしまうことが多い．このような場合には，最初は候補を大まかに絞るような選択をユーザにしてもらい，その後，絞られた候補をユーザに表示して選択してもらうという階層的選択を行う必要がある．以下では，どのように選択を階層化して，ユーザにとって選

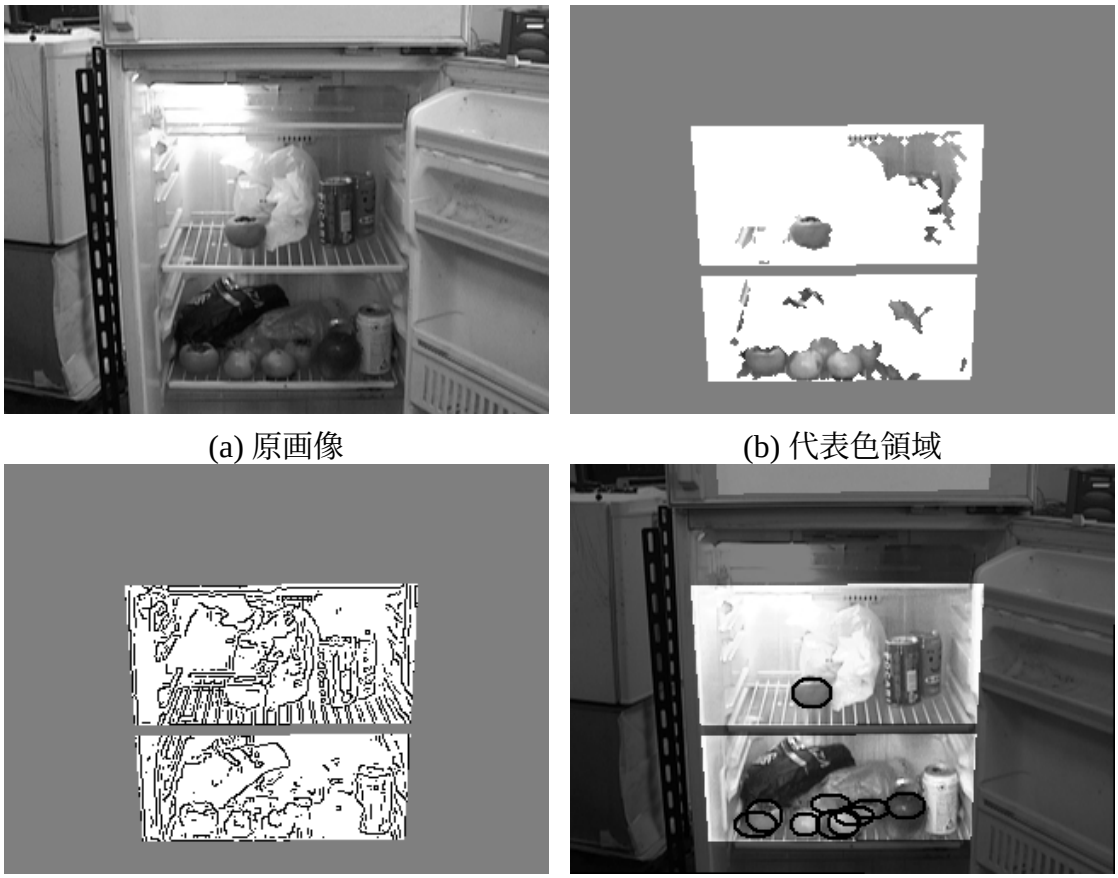


(a) 各被験者の評価



(b) 各表示方法の評価の最大値・最小値 (矩形の上端・下端) と平均 (×印)

図 2.44: 結果表示手法に関する主観テストの結果



(a) 原画像　(b) 代表色領域

(c) エッジ画像　(d) 果物の候補

図 2.45: 果物認識時の多数の候補領域

択しやすい候補表示を行うかについて述べる．

代表候補の限定表示

階層化の手法として，候補をいくつかのグループに振り分けておき，グループ毎に所属する候補を表示して，ユーザにその中に正しいもの (取って欲しいもの) があるかどうかを尋ねていくことで，一回当たりの選択候補を減らすというものがある．その際に選択効率を考えると，最初の方にユーザに表示するグループは，正しい物体が含まれている可能性の高い候補で構成されていることが望ましい．今回は，選択効率と候補表示の見やすさを考慮して，最初に表示するグループ (ここでは代表候補と呼ぶ) と二回目に表示するグループ (ここでは代替候補と呼ぶ) に分類する．ここで，図 2.45(d) における候補に着目すると，選択時の見辛さの主な原因となってのは互いに重なりあっている楕円であると考えられる．よって，代表候補は，重なっていない楕円と，互いに重なった候補群毎に楕円の信頼性 (2.3.2 節参照) が最も高いものを精選したものとで構成する．その結果を図 2.46(a) に示す．

これにより，ある程度の選択効率を保ったままで候補表示を見やすくすることができる．また，代表候補内に正しい物体が無かった場合には，同図 (b) の代替候補をユーザに表示して選択してもらう．

(a) 代表候補　　(b) 代替候補

図 2.46: 代表候補と代替候補

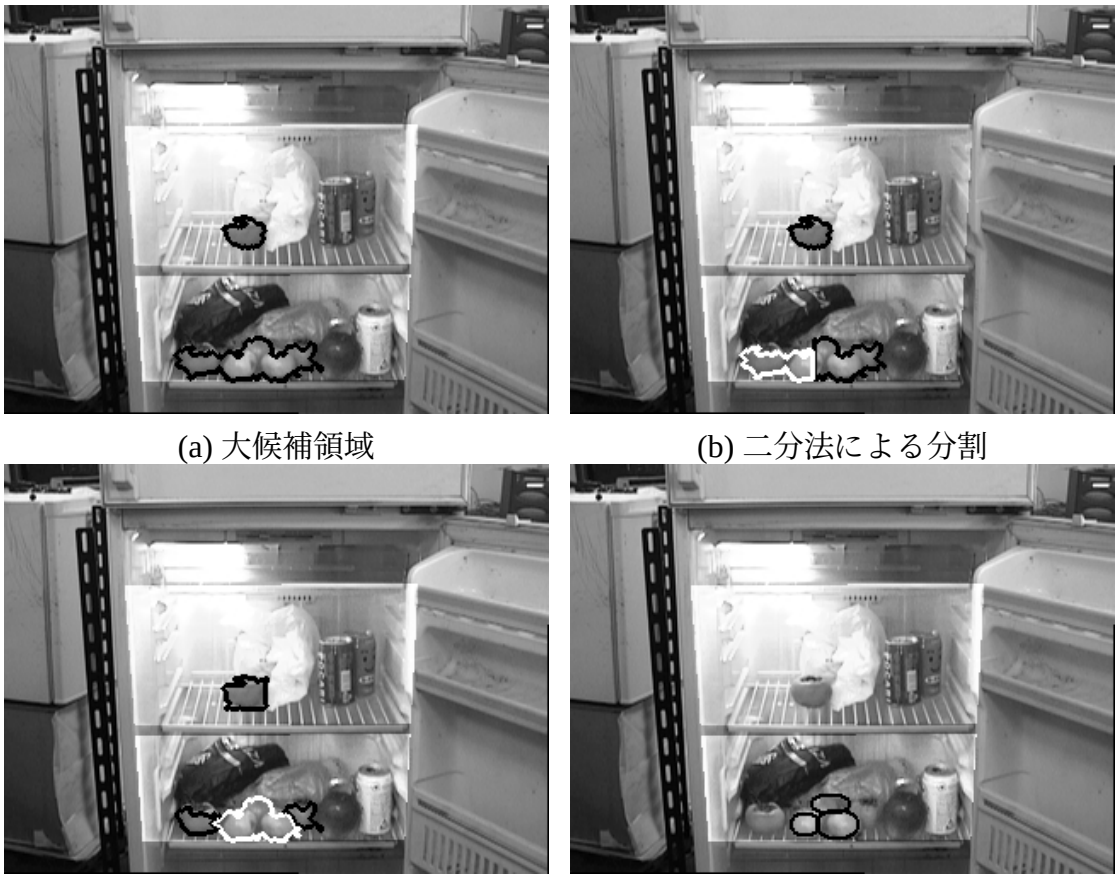
大まかな領域選択による階層化

他の階層化の方法としては，物体候補を表示する前に，大まかな候補領域をユーザに選択してもらい，選択された候補領域内でのみ物体検出を行うことで一回当たりの選択候補数を削減する方法がある．この際に図 2.45(a) のように物体が密集して置かれているような場合には，図 2.47(b) のように大きな一塊の候補領域として抽出される事がある．大きな候補領域のままでユーザに選択してもらっても，その後の候補数はそれほど減少しない．例えば，図 2.45(b) で下の棚の候補領域を選択しても，図 2.45(d) において上の棚の一つの候補が減少するだけである．そこで，大きな候補領域が存在する場合には，その領域をいくつかの小さな領域の分割しておき，その候補領域から選択してもらう事で，その後に表示する物体の表示候補を効率に減少させる．

このような大きな候補領域の分割方法としては，図 2.47(b) のような二分法による分割がある．しかし，様々な色の果物が集まってできた候補領域に対しては，図 2.47(c) のようの色に基づく分割を行った方が物体の種類毎に分割される可能性があるために望ましい．ここで，例えば図 2.47(c) の白色で囲まれた蜜柑の領域を選択した場合には，図 2.47(d) のように少数 (3 個) の物体の代表候補からの選択で済む．

但し，色による分割を行う場合には，図 2.48(a) のように他の果物の陰領域や背景領域の一部を含んでしまい，本来の物体の形状としてはあり得ないような細い領域部分が得られることがある．このような細い領域は図 2.48(b) のように削除して表示する方が，ユーザにとっては見やすいと考えられる．

具体的な削除方法は以下の通りである．



(a) 大候補領域　(b) 二分法による分割

(c) 色に基づく分割　(d) 選択された候補領域中の代表候補

図 2.47: 候補領域の分割方法

1. 認識対象の物体モデルから，最小の大きさを取得
2. 候補領域の各点を物体中心と見なして最小の物体モデルを重ね合わせ，図 2.49(a) のようにモデルの領域内における候補領域の割合が一定以上 (ここでは実験的に 70%とする) である場合には，その領域内の候補領域は正しい領域であると見て各点に一票ずつ投じ，図 2.49(b) のように一定割合を超えていない領域については細領域と見なして投票を行わない．
3. 候補領域中で投票数が 0 の領域を細領域と見なして削除する．

階層表示に関する主観テスト

上記で述べた階層表示が，実際にユーザにとってどのように感じるかを確かめるために，図 2.42 の画像を用いて，表 2.5 の階層表示に関する変更内容を行った場合に以前に比べてどの程度良くなったかという主観テストを，被験者 6 人に対して行った．

(a) 細領域削除前　　(b) 細領域削除後

図 2.48: 細領域の削除

(a) 正しい物体領域として見なされる候補領域　　(b) 細領域と見なされる候補領域

図 2.49: 細領域の削除モデル

評価尺度としては，表 2.4 に示す 11 段階の相対評価を用いた．

主観テストの結果を図 2.50 に示す．同図 (a) は各被験者の評価であり，同図 (b) はそれをまとめて最大値・最小値と平均を表したものである．

結果より，色に基づく領域分割と細領域の削除に関してはいずれの被験者に関しても「5. 変わらない」以上の評価を得ており，効果的な処理であると考えられる．一方，代表候補と代替候補に分けて表示する処理に関しては，多くの被験者が改善されたと感じているものの，一部の被験者で悪化したという評価を受けた．これは，選択回数が増えることによって煩わしさが増加したことが主な原因であ

表 2.5: 改善度合いを確かめる処理項目

	変更内容
1	全物体候補を表示する代わりに，代表候補と代替候補に分けて表示
2	大まかな候補領域を選択してもらって，領域内の物体候補のみを表示
3	大きな候補領域の分割を二分法から色に基づく分割に変更
4	細領域削除

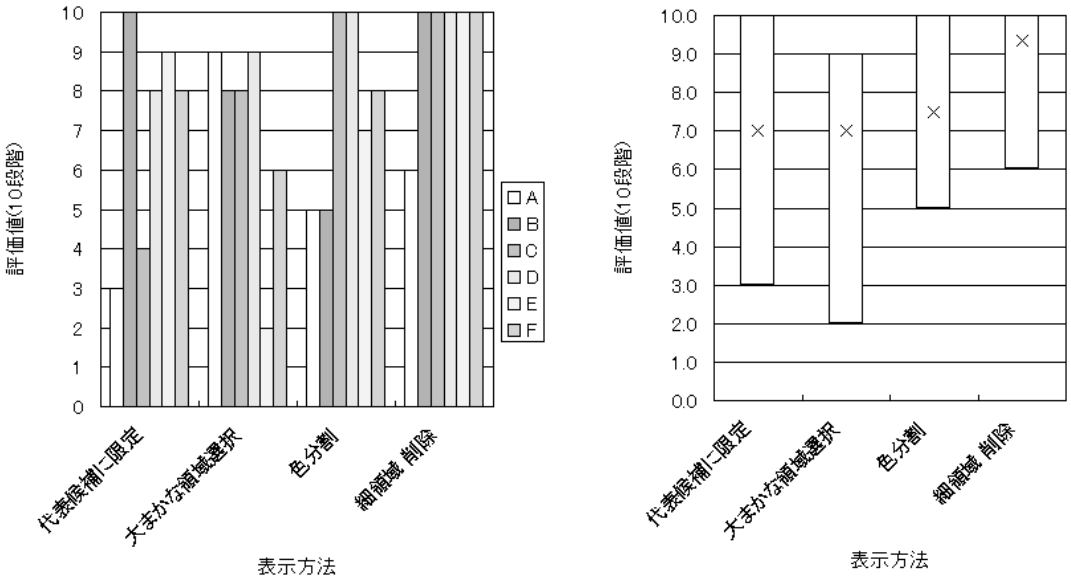
表 2.6: 主観テストに用いた 11 段階の相対評価

評価値	評価内容
0	著しく悪化
1	かなり悪化
2	悪化
3	やや悪化
4	どちらかといえば悪化
5	変わらない
6	どちらかと言えば改善
7	やや改善
8	改善
9	かなり改善
10	著しく改善

る．また，大まかな領域選択による候補数削減についても大多数の被験者が改善されたと評価したものの，一名の被験者は悪化したという評価を下している．これは，物体を取って欲しいと頼んだにも拘わらず，領域を表示されると違和感を受けたためである．また，改善されたと評価した被験者からも領域表示に関してはさらに改善すべきとの評価を受けた．これは，図 2.47(c) のような色分割によって得られた複雑な輪郭を表示するとユーザにとっては見辛いものとなるので，もう少し大まかな滑らかな領域輪郭を示した方が良いというものである．これは，細い領域削除と同様の方向の考え方であり，ユーザにとってより直感的に分かりやすい領域表示として必要な処理であると考えられる．

2.6 まとめ

本章では，まず可能な限り簡単に認識でき，ユーザに必要以上の対話の手間をかけさせないための自動物体認識手法について述べた．そのためには，物体がどの方向に向いていても自動物体認識が行えるような物体モデルをあらかじめ作成しておく必要がある．ここでは，物体の大きさ・代表色・2 次特徴を物体の特徴として，物体を多方向から撮影したものを重ね合わせて作成した投影画像を用いて，同一の特徴を持つ区間毎に登録した．これにより，多方向から見た画像から特徴を別々に登録しておく方法やテンプレートマッチングによる方法と比較して，登録しておく特徴の数が減り，認識が容易に行うことができる．これに加えて，多の物体との区別が付くだけの必要最低限の特徴を登録しておき，その特徴だけでは区別が不可能な物体が登録されて初めて 2 次特徴を登録するという方針により，



(a) 各被験者の評価　　(b) 各表示方法の評価の最大値・最小値 (矩形の上端・下端) と平均 (×印)

図 2.50: 結果表示手法に関する主観テストの結果

更に特徴の数を減らして，認識を容易にすることができた．

このようなモデルを基にした自動物体認識の手順は，大きく見て検出と照合に分かれる．検出は，指定された物体の代表色を用いて行い，その周囲に存在するエッジを抽出してその位置関係を調べることにより，缶・瓶・ペットボトル・果物の領域を検出する．そして，指定された物体に 2 次特徴が登録されていれば，その色に当てはまる領域を抽出してその位置・面積等の照合を行い，最終的な認識結果を得るものとした．

一連の自動物体認識が終えると，システムはその結果をユーザに提示する．ここで，指定物体が検出できなかったときや，認識結果が誤っている場合には，ユーザとの対話を行う．ここで，システムは現在の認識状況とユーザからの返答によって，各々に適したリカバリー処理を行う．例えば，ユーザから指定物体の隠蔽に関する情報が与えられると，隠蔽している物体の抽出を先に行うことで冷蔵庫内の位置を限定して，さらに隠蔽されていることから画像上での物体の大きさ (見え) が小さくなることを考慮して物体の領域を抽出することができる．ここでは，隠蔽による大きさの変化を考慮した認識や色の範囲を拡大した再抽出，ユーザの訂正による物体モデルの学習等を扱い，ユーザとの対話が物体認識に対する有効な補助となることが分かった．

また，多数の候補が検出された場合にどのようにその候補を提示するとユーザにとって選びやすいかということについて，領域に対するラベル付けや階層的な表示を導入して対応した．数名のユーザに対する主観テストを行うことで，こられの効果を確かめた．

第3章　グローバルな外部照明変化を考慮した色変換の学習

本章では，最初に 3.1 節で関連研究との関わりを述べた後，3.2 節で光の反射の物理モデルに基づく線形な色変換を定式化する．3.3 では一つの参照色を用いてグローバルな外部照明変化の下での色変換を推定する方法について述べる．3.4 節では，これらの色変換モデルのパラメタを求めるのに十分な学習セットが存在する場合の学習方法についてに述べ，3.5 節で本システムが新しい環境で使用され，十分な学習セットが存在しない場合における適応方法について述べる．また，3.6 節では色変換を考慮した物体認識について述べる．最後に，3.7 節で照明変化への対応に関する実験について述べる．

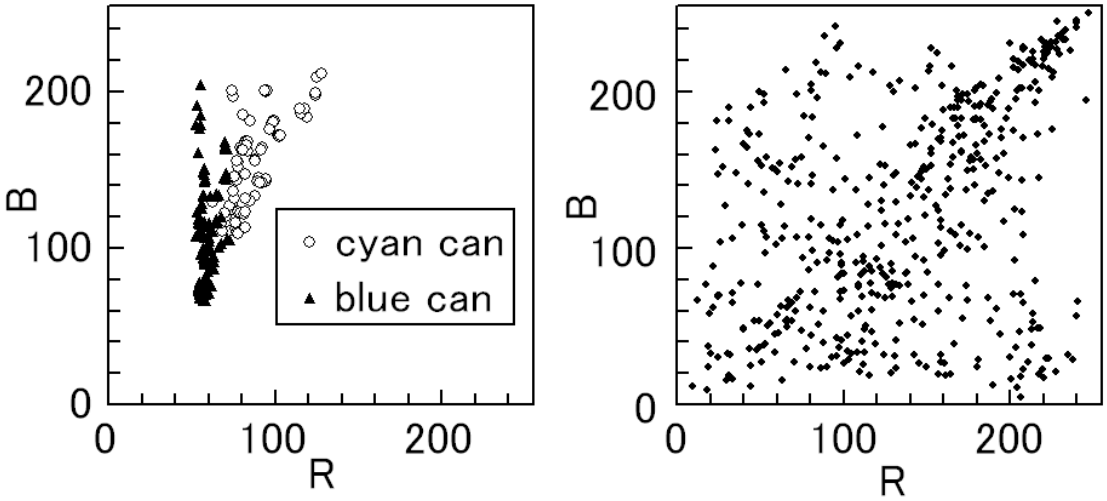
3.1　関連研究との関わり

本研究の物体認識システムにおいて，物体モデルの撮影時の照明条件 (図 3.1(a)) と認識時の冷蔵庫内の照明条件 (図 3.1(b)) とは異なるので，モデルをそのまま用いることはできない．また，序論でも述べたように，認識時の照明条件自体も時間によって変化する．実際に認識時に観測される物体の色が，照明変化によってどの程度変化するかを図 3.2(a) に示す．ここでは，水色の缶と青色の缶の R, B 成分を例に取っているが，いずれも観測される色が大きく変化しており，その分布が重なっていて区別できない部分が存在する．また，図 3.2(b) に示すように，認識対象の全物体モデルの色は色空間上で互いに近接して存在しており，似た色の物体が多数存在する．よって，これらを区別するためには認識時の照明条件を知って，物体の色がどのように変化するかを推定する必要がある．この推定精度が高いと区別可能な色が増加して，誤認識を訂正するためのユーザの対話を軽減できる．しかし，認識時の照明条件は未知であるため，どのようにして照明条件を推定するかが問題となる．

この問題に対しては，光の反射の物理法則に基づく色モデルを考えて，照明条件の変化によって物体の見えが変化しても，その表面反射率は変化しないというカラーコンスタンシー [50] に関する研究がある．その応用例としては影領域と元の領域の表面反射率が変化しないことを利用した移動物体の影検出の研究 [51] がある．人間の場合には優れた視覚システムによってこのカラーコンスタンシーを

(a) モデル撮影時の照明　　(b) 冷蔵庫内の照明

図 3.1: 照明変化の問題 ((b) の白枠内の物体は (a) の物体と同一)



(a) 照明変化による色分布　　(b) 物体モデルの色分布

図 3.2: 照明変化による色分布と物体モデルの色分布

実現している [49] が，コンピュータによる視覚システムにとってはカラーコンスタンシーの解決は容易ではない．

一方，カラーコンスタンシーの問題は，照明条件の変化に対する色の変換の問題として扱うこともできる．単純な方法としては，シーン中の全物体の平均表面反射率が灰色になるという仮定 [52] や最大の反射率を持つシーン中の点が最も明るい画素に対応するという仮定 [53] を用いて色の変換を行うものがある．しかし，これらの仮定はしばしば成り立たない場合がある．

シーンに対する緩い仮定を用いたものとしては，シーン中に存在し得る物体の表面反射率の範囲を知識として与えておき，現在観測されている色から色変換候補の範囲を制限することで色変換を求める手法 [58] や，通常考えられる光源の知識も利用して同様にして色変換を求める手法 [59] がある．しかしこれらは色度空

間による色変換を取り扱っているため，輝度に関するカラーコンスタンシーは実現していない．

これに対して，シーン中の物体の表面反射率に関する仮定を用いずに，色モデルを線形化して色変換を扱う研究もある [54]．その中でも，RGB の各成分を独立に変換することがよく行われている [55]．Drew ら [56] は，物体のモデルと認識対象の画像の間で特徴の対応を取り，その特徴同士の間での色変換を投票空間に投票することで，全体の色変換の候補を得る手法を提案している．

一方，統計に基づいて非線形な色変換を取り扱う研究もある．Miller ら [57] は異なる照明条件下で撮影した同じシーンの画像中の色の組から代表的な color flow を学習し，色変換を行う手法を提案している．[56] や [57] の手法では変換の推定に複数の色の組が必要となるが，サービスロボットの視覚システムにとって未知照明条件下で複数の参照色を得ることは困難である．

これに対して我々は，一つの参照色のみを用いて色の変換の推定をする手法を提案する．最初に，光の反射モデルをもとにした色モデルと色の変換式を導き，参照色の変化から色変換を推定する方法を定式化する．色変換を推定するためのパラメタについては，物体モデルの色と様々な照明条件下で観測した物体の色と参照色の組を用いて学習する．一方，実際にこのようなサービスロボットを購入して各家庭で用いる場合を考えると，新しい環境下での観測データが存在しないため，新たにパラメタを学習し直す必要がある．そこで，最初は適当に仮定したパラメタに基づいて色変換をして，物体認識を試みる．そのような色変換を用いると物体認識に失敗することがあるが，その際にはユーザとの対話によって物体認識を補助してもらい，正しい物体を選択してもらう．そうして得られた正しい対応の観測データを用いて色変換を推定するためのパラメタを更新する．物体を認識する毎にこのような更新をすることで，色変換の推定精度を向上させて新しい環境への適応を目指す．これにより自動物体認識の成功率が向上し，ユーザの対話の負担を軽減させることができる．

3.2 光の反射の物理モデルに基づく色変換のモデル

光の反射には拡散反射と鏡面反射の二つが存在する．しかし，本研究で扱う物体のほとんどは拡散反射が支配的な回転対称な物体であり，特定方向に生じる鏡面反射 (ハイライト) は物体領域の内の一部に留まるため，認識に与える影響は小さい．そこで簡単のために，モデル作成時にはハイライトを除去した投影画像を作成しておき，認識時には拡散反射のみが生じる部分，即ち Lambertian 表面と見なすことのできる部分を用いて認識する．その Lambertian 表面での光の反射の物理モデルより，照明条件 L のもとで観測される物体 O の色 $\mathbf{c}^{OL} = [c_1^{OL}, c_2^{OL}, c_3^{OL}]^T = [R^{OL}, G^{OL}, B^{OL}]^T$ は，以下で表される [65].

$$c_p^{OL} = \int F_p(\lambda) S^O(\lambda) I^L(\lambda) d\lambda \tag{3.1}$$

ここで，λ は光の波長，$F_p(\lambda)$ は色の p 成分に対するカメラレスポンス，$S^O(\lambda)$ は物体 O の波長に対する表面反射率分布，I^L は照明条件 L のもとで物体に到達する光の波長に対する強度分布を表している．これに，暗電流による影響を加えると，

$$c_p^{OL} = \int F_p(\lambda) S^O(\lambda) I^L(\lambda) d\lambda + c_{dark,p} \tag{3.2}$$

となる．ここで $c_{dark,p}$ は暗電流による p 成分への影響を表す．一方，観測色をこのような光波長の積分形のままで表現すると，有限次元の観測色 $\mathbf{c}^{OL}$ から連続値の分布 $F_p(\lambda), S^O(\lambda), I^L(\lambda)$ を復元することができない．そこで，これらの分布を有限次元で表現することを考える．

ここで，典型的な昼光に対しては光強度分布 $I^L(\lambda)$ がいくつかの基底関数の線形和で表現可能であること [66] や，表面反射率分布 $S^O(\lambda)$ が少数の基底関数の線形和で表現可能であること [67]が知られている．更にカメラレスポンスについても同様に基底関数の線形和で表現可能であるという有限次元線形モデルの仮定 [68] が良く知られており，その仮定を用いると各分布は以下のように表すことができる．

$$F_p(\lambda) = \sum_{j=1}^{n_F} f_{pj} F_{b,j}(\lambda) + f_{bias,p} \tag{3.3}$$

$$S^O(\lambda) = \sum_{k=1}^{n_S} s_k^O S_{b,k}(\lambda) + s_{bias} \tag{3.4}$$

$$I^O(\lambda) = \sum_{l=1}^{n_I} i_l^L I_{b,l}(\lambda) + i_{bias} \tag{3.5}$$

ここで，$F_{b,j}(\lambda)$ はカメラレスポンスに対する j 番目の基底関数で，f_{pj} は p 成分の j 番目の基底関数に対する係数であり，$f_{bias,p}$ は基底関数表現したときの p 成分のバイアス成分である．同様に，$S_{b,k}(\lambda)$ と $I_{b,l}(\lambda)$ は各々反射率分布の k 番目の基底関数と光強度分布の l 番目の基底関数で，s_k^O と i_l^L は各々基底関数 $S_{b,k}(\lambda)$ と $I_{b,l}(\lambda)$ に対する係数であり，s_{bias} と i_{bias} はバイアス成分である．また，n_F, n_S, n_I は各分布の基底関数の数であり，その具体的な数については，観測される色要素が RGB の三つであるので，それに合わせて $n_F = n_S = n_I = 3$ とする．

これらの有限個の基底関数は物体の種類，照明条件，カメラの種類に対して共通であり，基底関数に対する有限個の係数を変化させることにより様々な物体の表面反射率分布や様々な照明条件における入射光分布を表現する．逆に，基底関数に対する有限個の係数を決めると，基底関数を用いて元の分布を復元することができる．これにより，本来は波長に対する連続値である分布を有限個の係数で表現できることになる．

ここで，有限個の係数からの分布の復元を行うためには，これらの基底関数を予め求めておく必要がある．例えば物体の表面反射率分布に対する基底関数を求める場合には，以下のようにして求めることが多い．

1. 適当な波長間隔で物体反射率をサンプリング (N 個) するものとして，各波長に対するサンプリング値を並べた N 次元ベクトルを定義する．

2. 様々な物体に対して物体反射率の N 次元ベクトルを取得する．

3. 得られた N 次元ベクトルに対して主成分分析を行い，上位 n_S 個の固有ベクトルを取得して，それをスプライン近似する等して波長に対する連続分布に戻すことで，表面反射率分布に対する基底関数とする．

カメラレスポンスや入射光分布についても，様々なカメラや様々な照明条件に対してこのような処理を行うことで，基底関数を求めることができる．しかし，本研究ではこの基底関数の具体的な形を定めない．これは最終的に基底関数の積分項が色変換モデル中のパラメタにまとめられ，そのパラメタが学習によって決定されるためである．

ここで，有限次元線形モデルの表現の簡単化のため，バイアス項を以下のようにして総和の中に入れて表す．

$$
\begin{aligned}
F_p(\lambda) &= \sum_{j=1}^{4} f_{pj} F_{b,j}(\lambda) && (3.6)\\
S^O(\lambda) &= \sum_{k=1}^{4} s_k^O S_{b,k}(\lambda) && (3.7)\\
I^O(\lambda) &= \sum_{l=1}^{4} i_l^L I_{b,l}(\lambda) && (3.8)\\
f_{p4} &= 1\\
F_{b,4}(\lambda) &= f_{bias,p}\\
s_4^O &= 1\\
S_{b,4}(\lambda) &= s_{bias}\\
i_4^L &= 1\\
I_{b,4}(\lambda) &= i_{bias}
\end{aligned}
$$

これより式 (3.1) は以下のようになる．

$$
\begin{aligned}
c_p^{OL} &= \sum_{k=1}^{4}\sum_{l=1}^{4} f_{pkl} s_k^O i_l^L + c_{dark,p} && (3.9)\\
f_{pkl} &= \int \sum_{j=1}^{4} f_{pj} F_{b,j}(\lambda) S_{b,k}(\lambda) I_{b,l}(\lambda) d\lambda
\end{aligned}
$$

さらに上式を s_k^O について整頓すると，以下のように行列で表現することができる．

$$
\mathbf{c}^{OL} = M^L \mathbf{s}^O + \mathbf{c}_{dark} \tag{3.10}
$$

ここで M^L の pk 成分 m^L_{pk} は $\sum_{l=1}^4 f_{pkl} i^L_l$ である．ここで，$\mathbf{s}^O$ の第 4 成分が 1 であること利用して，暗電流の成分を以下のようにして行列にまとめる．

$$
\begin{aligned}
\mathbf{c}^{OL} &= M'^L \mathbf{s}^O \qquad (3.11)\\
M'^L &= \begin{bmatrix} m^L_{11} & m^L_{12} & m^L_{13} & m^L_{14}+c_{dark,1} \\ m^L_{21} & m^L_{22} & m^L_{23} & m^L_{24}+c_{dark,2} \\ m^L_{31} & m^L_{32} & m^L_{33} & m^L_{34}+c_{dark,3} \end{bmatrix} \\
\mathbf{s}^O &= [s^O_1, s^O_2, s^O_3, 1]^T
\end{aligned}
$$

さらに，後で行う色変換時の便宜性のために色ベクトル $\mathbf{c}^{OL}$ を同次座標 $[c^{OL}_1, c^{OL}_2, c^{OL}_3, 1]^T$ で再定義して，以下のようにする．

$$
\begin{aligned}
\mathbf{c}^{OL} &= A^L \mathbf{s}^O \qquad (3.12)\\
A^L &= \begin{bmatrix} m^L_{11} & m^L_{12} & m^L_{13} & m^L_{14}+c_{dark,1} \\ m^L_{21} & m^L_{22} & m^L_{23} & m^L_{24}+c_{dark,2} \\ m^L_{31} & m^L_{32} & m^L_{33} & m^L_{34}+c_{dark,3} \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{aligned}
$$

これと同様にして，式 (3.9) を i^L_l について整頓すると，以下のように行列で表現することができる．

$$
\begin{aligned}
\mathbf{c}^{OL} &= B^O \mathbf{i}^L \qquad (3.13)\\
B^O &= \begin{bmatrix} h^O_{11} & h^O_{12} & h^O_{13} & h^O_{14}+c_{dark,1} \\ h^O_{21} & h^O_{22} & h^O_{23} & h^O_{24}+c_{dark,2} \\ h^O_{31} & h^O_{32} & h^O_{33} & h^O_{34}+c_{dark,3} \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
h^O_{pl} &= \sum_{k=1}^{4} f_{pkl} s^O_k \\
\mathbf{i}^L &= [i^L_1, i^L_2, i^L_3, 1]^T \qquad (3.14)
\end{aligned}
$$

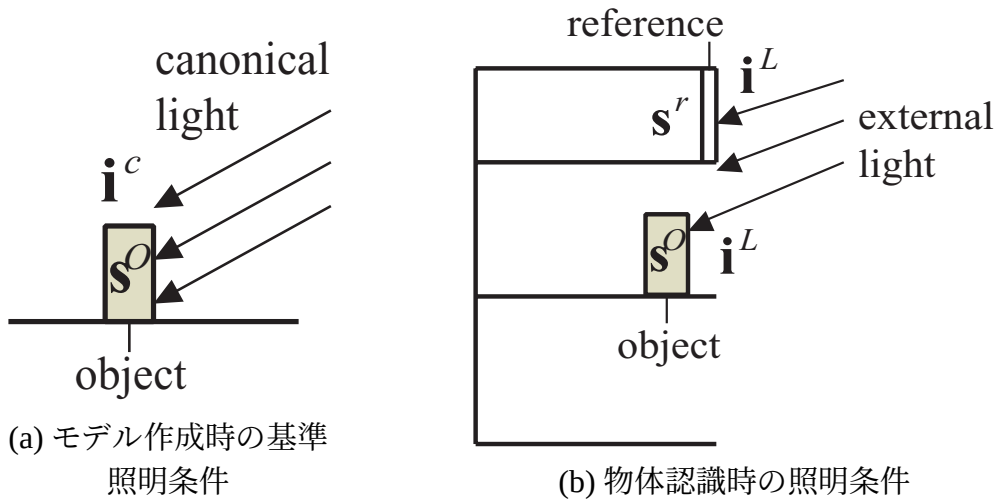
次に，色の変換について考える．異なる照明条件 L_1 と L_2 のもとで観測される物体 O の色を各々$\mathbf{c}^{OL_1}$，$\mathbf{c}^{OL_2}$ とすると，式 (3.10) より以下の式が得られる．

$$
\begin{aligned}
\mathbf{c}^{OL_1} &= A^{L_1} \mathbf{s}^O \qquad (3.15)\\
\mathbf{c}^{OL_2} &= A^{L_2} \mathbf{s}^O \qquad (3.16)
\end{aligned}
$$

これより線形な色変換が以下のように得られる．

$$
\begin{aligned}
\mathbf{c}^{OL_2} &= A^{L_2 L_1} \mathbf{c}^{OL_1} \qquad (3.17)\\
A^{L_2 L_1} &= A^{L_2} (A^{L_1})^{-1} \qquad (3.18)
\end{aligned}
$$

ここで，$A^{L_2 L_1}$ の第 4 行成分は $[0,0,0,1]$ となる．



(a) モデル作成時の基準照明条件

(b) 物体認識時の照明条件

図 3.3: 参照色を用いた色変換推定の問題設定

3.3 参照色を用いた色変換の推定

物体認識時の未知照明条件 L のもとで観測される物体 O の色 $\mathbf{c}^{OL}$ を推定するために，参照物体 r(ここでは冷蔵庫の扉や枠) の色を利用する．この問題設定を図 3.3 に示す．

最初に，モデル作成時の基準の照明条件 c のもとで観測される参照物体 r の色 $\mathbf{c}^{rc}$ と物体 O の色 $\mathbf{c}^{Oc}$ が与えられると，各物体の表面反射率 $\mathbf{s}$ が式 (3.10) を用いて以下のように表される．

$$\mathbf{s}^r = (A^c)^{-1}\mathbf{c}^{rc} \tag{3.19}$$

$$\mathbf{s}^O = (A^c)^{-1}\mathbf{c}^{Oc}. \tag{3.20}$$

次に，未知照明条件 L のもとで観測される参照物体 r の色 $\mathbf{c}^{rL}$ が与えられると，式 (3.9) より，

$$c_p^{rL} = \sum_k \sum_l f_{pkl} s_k^r i_l^L \tag{3.21}$$

となる．式 (3.19) から得られる $\mathbf{s}^r$ を上式に代入して整頓すると，未知照明条件 $\mathbf{i}^L$ は以下のようになる．

$$\mathbf{i}^L = (B^r)^{-1}\mathbf{c}^{rL} \tag{3.22}$$

ここで，行列 B^r の pl 成分は $(B^r)_{pl} = \sum_k \sum_j f_{pkl}((A^c)^{-1})_{kj} c_j^{rc}$ である．結局，式 (3.9), (3.20), (3.22) から求めるべき $\mathbf{c}^{OL}$ が以下のように得られる．

$$c_p^{OL} = \sum_j \sum_m \alpha_{pmj} c_m^{rL} c_j^{Oc} \tag{3.23}$$

ここで，各 α_{pmj} は色変換の推定パラメタであり，$\alpha_{pmj} = \sum_k \sum_l f_{pkl}((A^c)^{-1})_{kj}((B^r)^{-1})_{lm}$ である．表記の簡単化のために基準の照明条件下での物体モデルの色 $\mathbf{c}^{Oc}$, 未知照

明条件下で観測される物体の色$\mathbf{c}^{OL}$, 参照物体の色 $\mathbf{c}^{rL}$ を各々$\mathbf{c}^m, \mathbf{c}^o, \mathbf{c}^r$ として定義して，式 (3.23) を以下の二つの式で表現する．

$$\mathbf{c}^o = T(\mathbf{c}^r;\alpha)\,\mathbf{c}^m \tag{3.24}$$

$$T_{ij}(\mathbf{c}^r;\alpha) = \sum_k \alpha_{ijk} c^r_k \tag{3.25}$$

前者は物体の色が変換行列 $T(\mathbf{c}^r;\alpha)$ を用いて線形に変換されることを意味し，後者はその変換行列 $T(\mathbf{c}^r;\alpha)$ の各要素 $T_{ij}(\mathbf{c}^r;\alpha)$ が観測される参照物体の色 $\mathbf{c}^r$ の各成分の線形結合で表されることを意味している．

3.4 学習セットが十分にある場合の色変換の学習

まず最初に，求めるべきパラメタの数について考える．推定すべき色変換の行列 T は 4×4 となるが，先にも述べたように 4 行目の要素については，$[0,0,0,1]$ で固定であるので，パラメタの自由度は実質 12 である．よって，推定パラメタ α の実質的な個数は，12×4=48 となる．

次に，n 回の観測を行ったときに得られる方程式の数について考える．物体認識が 1 回成功する毎に，物体モデルと認識した物体の領域との正しい対応関係が得られる．ここで，認識時の対象物体と参照物体の観測色 $\mathbf{c}^o, \mathbf{c}^r$ として，抽出された領域の平均色を用いるので，1 回の観測で一組の $\mathbf{c}^m, \mathbf{c}^o, \mathbf{c}^r$ が得られ，式 (4.7) から 3 本の方程式が求まる[1]．よって，n 回の観測で得られる方程式の数は $3n$ である．

よって，この色変換の推定パラメタ α を求めるには，16 組の $\mathbf{c}^m, \mathbf{c}^o, \mathbf{c}^r$ が必要である．しかし，実際には線形化の仮定等により，この色変換には誤差が含まれているので，16 以上の色の組を用いて最小自乗法によって求める．物体認識時の照明条件下での誤差を用いると，最小自乗法の評価式 S_o は以下のようになる．

$$S_o = \sum_{j=1}^{n} \delta_j^T \delta_j \rightarrow min \tag{3.26}$$

$$where\ \ \delta_j = T(\mathbf{c}^r_j;\alpha)\,\mathbf{c}^m_j - \mathbf{c}^o_j \tag{3.27}$$

ここで n は観測数，δ_j は j 番目の観測に対する認識時の照明条件下での色変換誤差である．しかし，この評価式では，暗い照明条件下での誤差の重みが明るい照明条件下での誤差の重みに比べて小さくなるので，暗い照明条件への変換を軽視するような推定パラメタが求まる可能性がある．様々な照明条件下での誤差を均等に扱うために，各照明条件下での誤差をモデル作成時の基準の照明条件下に逆変換した誤差

$$\Delta_j = T(\mathbf{c}^r_j;\alpha)^{-1}\delta_j = \mathbf{c}^m_j - T(\mathbf{c}^r_j;\alpha)^{-1}\mathbf{c}^o_j. \tag{3.28}$$

[1]色変換式 (4.7) の 4 行目は両辺とも常に 1 となるので，観測毎に実質 3 本の方程式が得られることになる．

を用いる．さらに，反射率の低い物体の誤差は反射率の高い物体の誤差よりも小さくなるので，物体毎に変換誤差の共分散を求めてその共分散によって正規化したマハラノビス距離を用いる．結局，推定パラメタ α は以下の評価式 S を最小にするものとして求められる．

$$S = \sum_{j=1}^{n} \Delta_j^T \Sigma_{O(j)}^{-1} \Delta_j \rightarrow min \tag{3.29}$$

$$\Sigma_{O(j)} = \frac{1}{n_{O(j)}} \sum_{k \in I_{O(j)}} \Delta_k \Delta_k^T \tag{3.30}$$

ここで，$O(j)$ は j 番目の観測の対象物体であり，$\Sigma_{O(j)}$ は，物体 $O(j)$ の色変換誤差の共分散行列である．式 (3.30) の $n_{O(j)}$ は物体 $O(j)$ の観測数で，$I_{O(j)}$ は物体 $O(j)$ の観測番号の集合 $\{k|O(k) = O(j)\}$ である．

式 (3.29) は α について 2 次形式ではないため，共役勾配法によって最小化の問題を解く．初期値は，α について 2 次形式になっている式 (3.26) の S_o を最小化することで求める．

3.4.1 色変換の学習結果

学習には図 3.4 のように，様々な照明条件で撮影されたサンプル画像を用いた．また，物体としては様々な色の 16 種類の缶を用いた．ここでは，114 枚サンプル画像より学習用の色の組として 912 の観測値を取得して，学習を行った．

色変換の推定パラメタを学習した結果の学習データ全体の色変換誤差を図 3.5 に示す．この図は，学習に用いた冷蔵庫内の物体の観測色を基準照明条件下での色に変換して示し，物体毎の色変換誤差の共分散行列を楕円で表示したものである．色変換誤差が無ければすべての点はモデルの色を表す楕円の中心点に集まり，色変換誤差が小さい程この楕円も小さくなる．同図より，色変換誤差がある程度小さくなっており，色変換の推定精度が良いことが分かる．

加えて，図 3.6 のように，冷蔵庫の扉のシーンに対しても様々な照明条件で撮影されたサンプル画像を用いて，学習を行った．物体としては様々な色の 16 種類の缶を用い，40 枚サンプル画像より学習用の色の組として 320 の観測値を取得して，学習を行った．

色変換の推定パラメタを学習した結果の学習データ全体の色変換誤差を図 3.7 に示す．冷蔵庫内のシーンと同様，色変換誤差がある程度小さくなっており，色変換の推定精度が良いことが分かる．

次に，冷蔵庫内のシーンに対して学習した推定パラメタを用いて，ある物体モデルの色を変換した例を図 3.8 に示す．元の物体モデル M は，参照色である冷蔵庫の扉の色 (黒枠) を観測して，暗い照明条件 L_1 のもとでのモデル M_1 と，明るい照明条件 L_2 のもとでのモデル M_2 に変換され，各照明条件下での物体 (白丸) と同じ色になっている．

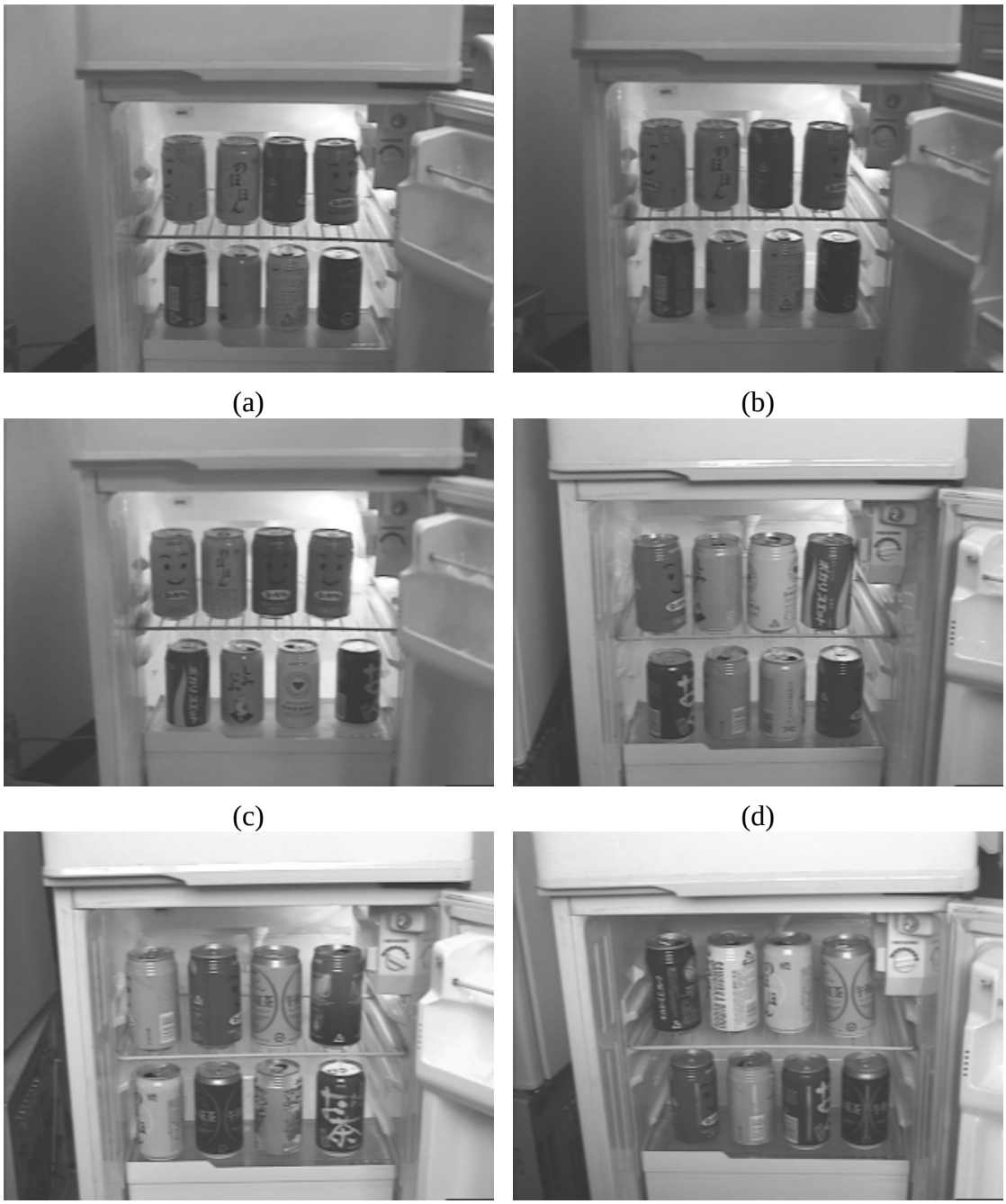
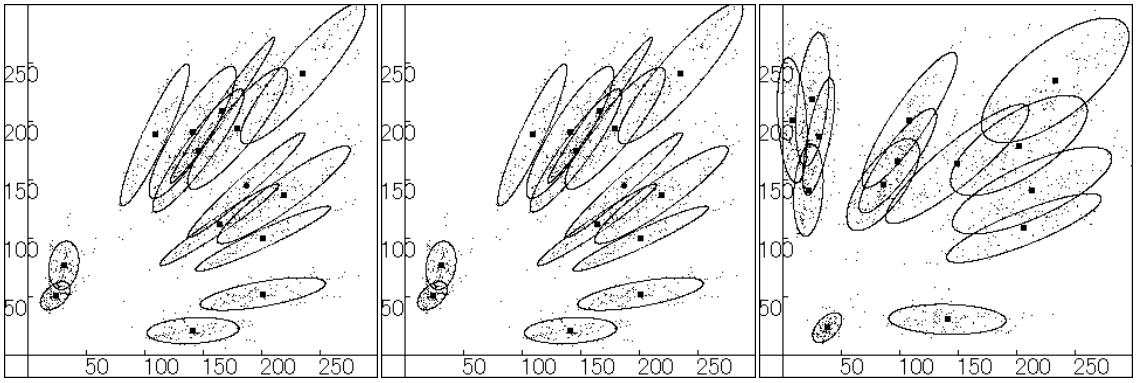


図 3.4: 学習サンプル画像



(a) R-G (b) G-B (c) B-R

プロット: 基準照明に変換した物体の観測色，楕円: 物体毎の色変換誤差の共分散，黒丸: 物体モデルの色

図 3.5: 基準照明下での色変換誤差

様々な照明条件に対して十分に学習した色変換パラメタを用いて物体モデルを変換したものを図 3.9 に示す．この図では，左が様々な照明条件での原画像を表しており，右の上列に並んでいる物体モデルは登録時のものであり，各原画像の冷蔵庫内での物体配置と対応している．また，右の下列に並んでいる物体モデルは登録時の物体モデルを原画像の照明条件に合わせて変換してものである．これより，同図 (a) のような明るめの照明条件や同図 (d) のようなかなり暗めの照明条件に対しても適切にモデルを変換できていることが分かる．

3.5 新しい環境での色変換の学習

新しい環境下で初めて認識を行う場合には，その環境中の参照物体に関する情報が全く無いので，モデル作成時の照明条件下での参照物体の色を白色 [255,255,255] と仮定して明度補正のみを行って認識を試みる．ここで，色変換の推定が間違っている場合には，2.4 で述べたようにユーザとの対話を通して再度認識を試みる．このようにして認識を成功させると，1 組の物体の色と参照色の組が得られる．但し，全ての推定パラメタを求めるためには色の組が足りないので，推定パラメタの中で重要なパラメタのみを求めて，残りのパラメタを 0 に拘束しておく．2 回目以降はその一部を拘束した推定パラメタを用いて色変換をして認識を行い，そこで得られる色の組を新たに加えて重要な推定パラメタから順番に拘束を解いて，パラメタを更新する．

ここで推定パラメタの拘束をどのように解いて学習していくかが問題となる．この問題は大きく分けると，以下の二つの問題になる．

1. パラメタの拘束をどの順番で解いていくか

図 3.6: 冷蔵庫の扉に対する学習サンプル画像

2. パラメタの拘束をいつ解くか (いくつの色の組が得られた時点で解くか)

(1) については，全てのパラメタを個別に扱うのは煩雑なので，まず，推定パラメタを式 (3.24) の変換行列 T の対角項 T_{ii} や定数項 T_{i4} に用いられるパラメタのように性質の似たパラメタ毎にグループ化する．具体的には，以下のグループ数 $n_p = 9$ のパラメタグループ G_i に分けた．

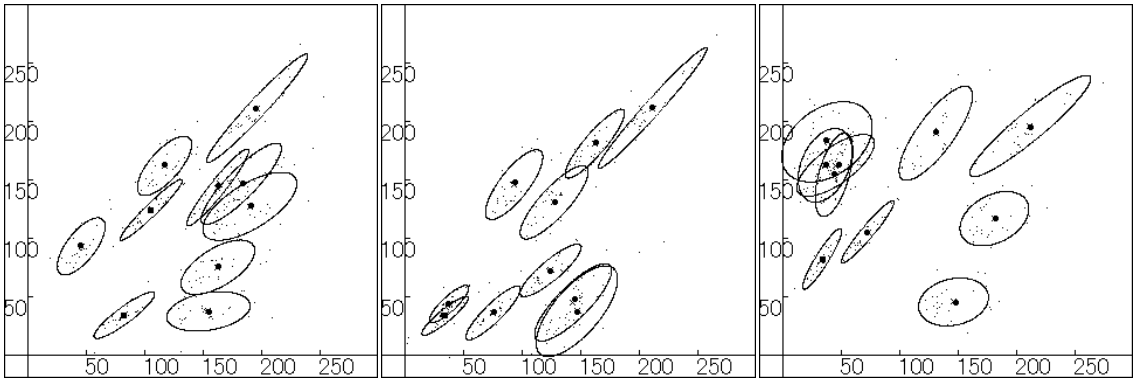
$$G_1\{\alpha_{iii}\}, G_2\{\alpha_{ii4}\}, G_3\{\alpha_{i44}\}, G_4\{\alpha_{i4i}\}, G_5\{\alpha_{iji}\}, G_6\{\alpha_{iij}\}, G_7\{\alpha_{ij4}\}, G_8\{\alpha_{i4j}\}, G_9\{\alpha_{ijk}\} \tag{3.31}$$

ここで，$1 \leq i, j, k \leq 3, j \neq i, k \neq i$ である．

次に，これらのパラメタグループの拘束を解く順序を決定する．ここでは，3.7.2 節で述べる適応実験によって，実験的に認識率の向上に効率的に寄与するパラメタグループの順番を決めておき，その順序に従って拘束を解くものとする．

まず，色変換の対角項に直接的に影響を及ぼす最も重要な G_1 に対して，各々グループ G_i の拘束を解いたものを α'_i とする．この時の適応実験結果を図 3.10 に示す．

この図より，G_1 の次には G_2 の拘束を解くと効率的であることが分かる．次に，G_1, G_2 に加えて残りの G_i を加えたものに対して適応実験を行い，次に拘束を解く



(a) R-G (b) G-B (c) B-R

プロット: 基準照明に変換した物体の観測色，楕円: 物体毎の色変換誤差の共分散，黒丸: 物体モデルの色

図 3.7: 冷蔵庫の扉に対する基準照明下での色変換誤差

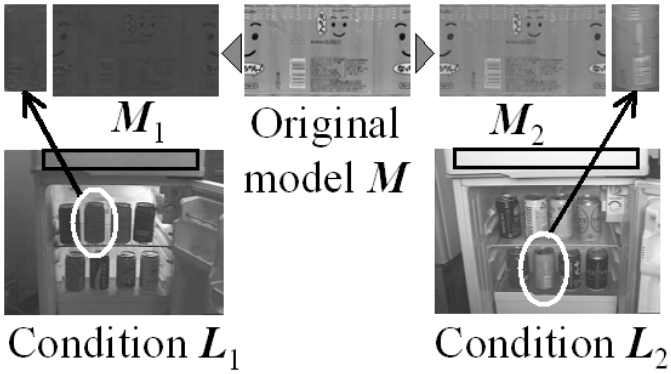


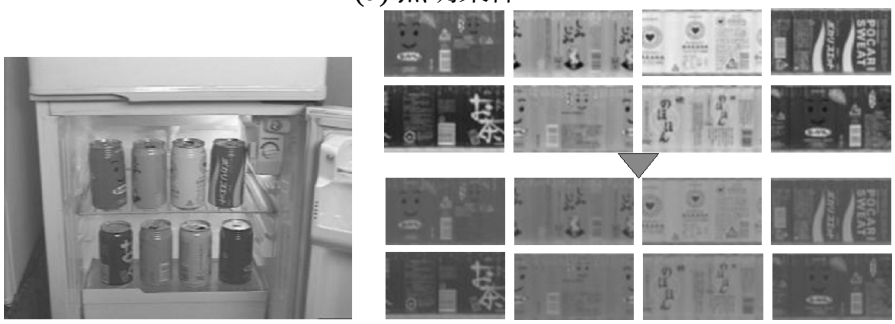
図 3.8: 色変換の結果

グループを決定する．これを最後のグループまで繰り返していくと，結局，先に決めたグループ G_i の順序と同じになることが分かった．ここで，明度補正のみのパラメタを α_0 に対して，各グループ G_i の拘束を解いていったものを α_i として定義する．ちなみに，α_9 は全ての拘束を解いた推定パラメタである．

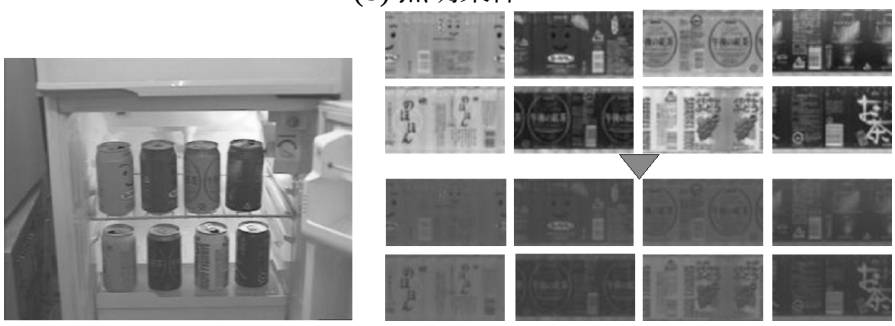
(2) については，各推定パラメタグループを求めるのに必要最低限度の色の組が得られた時点ですぐに拘束を解くと，過学習により不適切な推定パラメタが算出されて認識率が低下する場合があるので，パラメタ数に対してある程度の色の組が得られてから拘束を解く．一方，少ないパラメタ数のままでは様々な照明条件下での色の組が得られるにつれて，変換を十分に表すことができなくなるので，十分な色の組が得られた場合には次のパラメタの拘束を解く．

但し，パラメタグループの拘束を解くのに最適な点は明確に決まっているものではなく，パラメタグループの拘束を解くと，その時点で推定パラメタが急激に変化して認識率の変化が起こることがある (図 3.11 の灰線)．そこで，各段階での

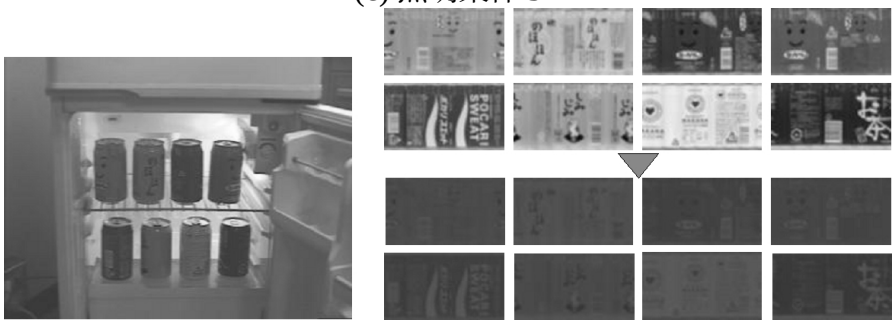
(a) 照明条件 1



(b) 照明条件 2



(c) 照明条件 3



(d) 照明条件 4

左: 原画像, 右上列: 物体モデル，右下列: 原画像の照明条件に合わせて変換した物体モデル

図 3.9: 物体モデルの色変換例

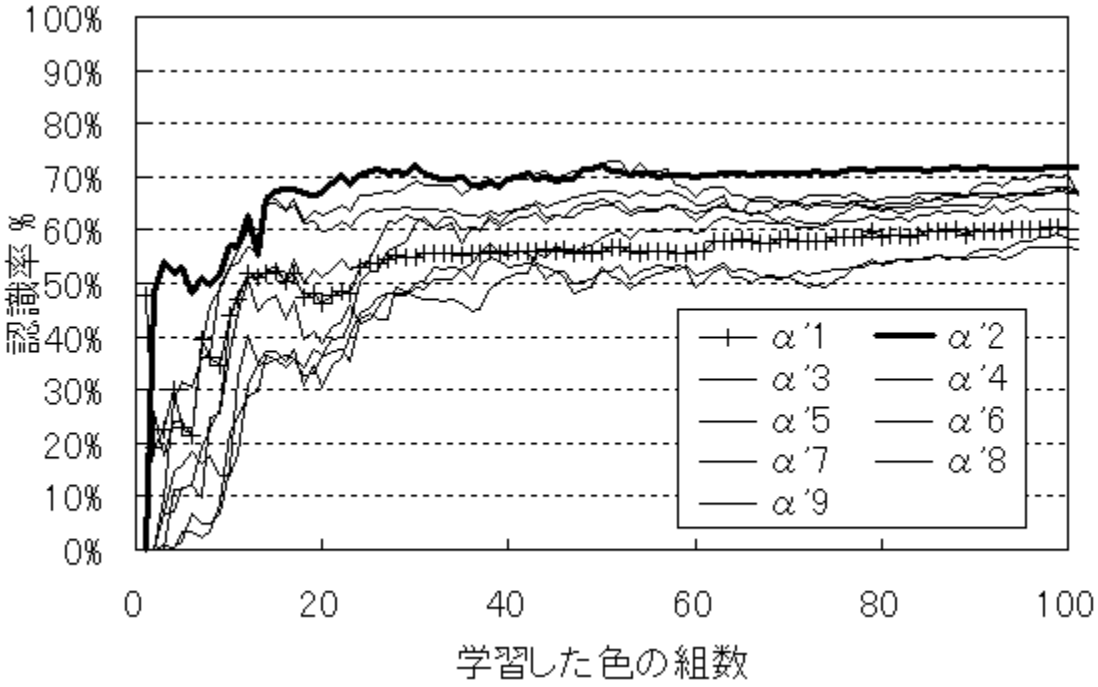


図 3.10: グループ順序決定のための適応実験

推定パラメタ $\alpha_i (i = 0, \ldots, n_p)$ に重みを与えて，以下のような重み付き線形和によって実際に用いる推定パラメタ α を求めることで，急激な変化を軽減する (図 3.11 の黒線).

$$\alpha = \sum_{i=0}^{n_p} w_i \alpha_i, \quad w_i = \frac{d_i}{\sum_{i=0}^{n_p} d_i} \tag{3.32}$$

$$d_i = \begin{cases} \delta_{i0} & (n < 2n_i) \\ 1 - (1 - \delta_{i0}) \frac{4n_i - n}{2n_i} & (2n_i \leq n < 4n_i) \\ 1 - (1 - \delta_{in_p}) \frac{n - 4n_i}{2n_i} & (4n_i \leq n < 6n_i) \\ \delta_{in_p} & (n \geq 6n_i) \end{cases} \tag{3.33}$$

ここで，w_i は α_i に対する重み，n_i は α_i の中で拘束を解かれたパラメタ数，δ_{ij} はクロネッカーのデルタである．また，n はこれまでに蓄えている色の組から得られる方程式の数で，n_c 組の色が得られたときには $n = 3n_c$ となる．上式は，観測した色の組によって得られる方程式の数が $2n_i$ より小さいときには，過学習を防ぐために重みを 0 にしておき，$4n_i$ に至るまでは徐々に重みを大きくしていき，そこから $6n_i$ に向けては，パラメタの数が足りずに十分に変換を表すことができなくなることを防ぐために重みを小さくしていき，それ以上では重みを 0 とすることを意味している．これにより，推定パラメタを緩やかに移行させて，急激な認識率の変化を防ぐことができる．

推定パラメタの算出方法は前節で述べた方法と同様であるが，その際には式(3.29)

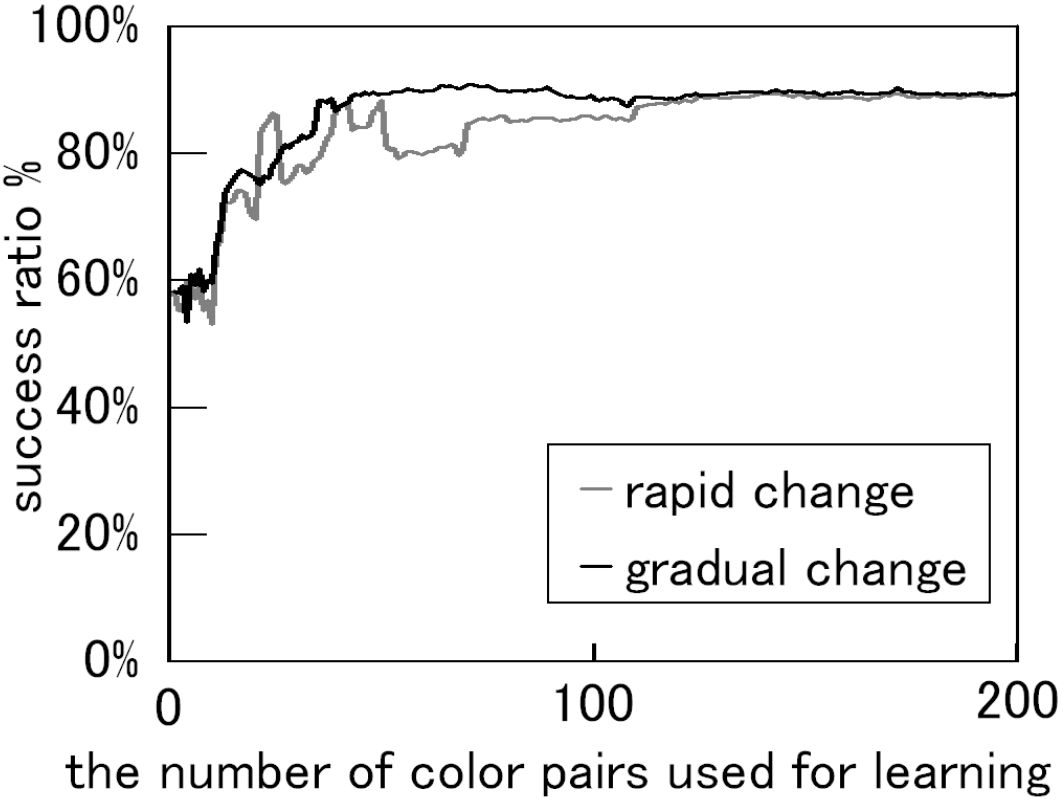


図 3.11: パラメタ拘束解除による認識率の急激な変化

の変換誤差の共分散行列 Σ_i が必要となる．しかし，初期段階においてはそれを計算するのに十分な色の組が存在しない．ここでは，あらかじめ実験用の環境で学習しておいた変換誤差の共分散行列を代用して，十分な色の組が得られたときに新しい環境での共分散行列に切り替えるものとする．

3.6 照明変化を考慮した物体認識

照明変化を考慮した物体認識を行う際には，以下の手順で行う．

1. 冷蔵庫領域を抽出して計算した参照色 $\mathbf{c}^r$ によって色変換行列 T を推定
2. 色変換行列の逆行列 T^{-1} で原画像をモデル登録時の基準照明条件へ逆変換した正規化画像を作成
3. 正規化画像で物体認識

手順の最後の物体認識は，大きく分けると物体の検出と識別からなる．物体の検出時には，正規化画像上での各画素値 $\mathbf{c}$ に対して，物体 O の代表色 $\mathbf{c}_O^m$ に該当す

るかどうかを判定して，該当する画素の集合として候補領域を抽出する．そのため，色変換の学習時に考慮した領域の平均色の色変換誤差に加えて，領域内での色のばらつきについても考える必要がある．この色のばらつきは，代表色 $\mathbf{c}_O^m$ を求める際に用いた投影画像上の一様色領域内での色の共分散行列 $\tilde{\Sigma}_O$ で表現する．結局，この二つのばらつきを併せた共分散行列 $\Sigma_O + \tilde{\Sigma}_O$ を用いて，正規化画像の画素値 $\mathbf{c}$ とモデル $\mathbf{c}_O^m$ の誤差を正規化した以下のマハラノビス距離を評価に用いる．

$$d_d = \sqrt{(\mathbf{c} - \mathbf{c}_O^m)^T (\Sigma_O + \tilde{\Sigma}_O)^{-1} (\mathbf{c} - \mathbf{c}_O^m)} \tag{3.34}$$

最終的には，このマハラノビス距離が，3 次元正規分布の 90%信頼区間に対応する閾値 d_t 以下であれば，代表色 $\mathbf{c}_O^m$ に該当すると判定する．

一方，検出した物体領域の識別を行うときには，色変換の学習時と同様に，画素値 $\mathbf{c}$ の代わりに領域内の画素値の平均値 $\overline{\mathbf{c}}$ を用いて，以下のマハラノビス距離 d_r を評価に用いる．

$$d_r = \sqrt{(\overline{\mathbf{c}} - \mathbf{c}_O^m)^T {\Sigma_O}^{-1} (\overline{\mathbf{c}} - \mathbf{c}_O^m)} \tag{3.35}$$

各領域をこのマハラノビス距離が最も小さい物体として識別し，2 位以降の物体を代替候補としておく．

3.7 実験

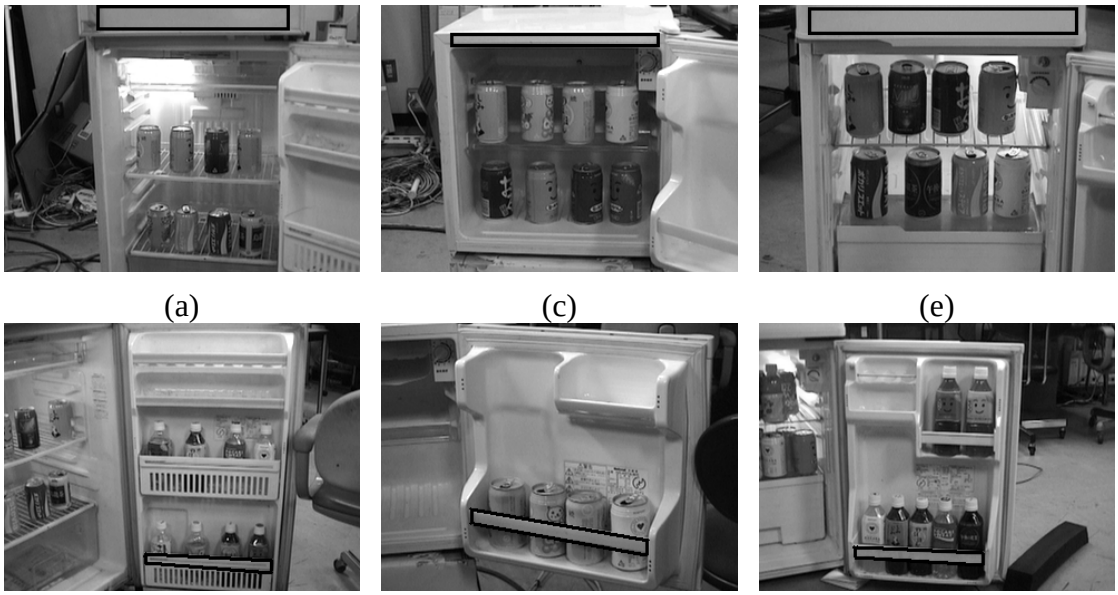
本節では，最初に様々な照明条件下での候補領域抽出実験について述べ，次に新しい環境への適応実験について述べる．

3.7.1 候補領域抽出実験

まず，十分に推定パラメタを学習した場合の色変換推定の性能を調べるための実験を行った．この実験には様々な照明条件で撮影した 114 枚の画像から，3.12(e) の冷蔵庫に対して，参照物体 (黒枠で囲われた冷蔵庫の扉) と物体の色の 456 組を学習セットとしてあらかじめ抽出しておき，それとは別の 456 組を実験セットとして用いて，候補領域の抽出実験を行った．比較のために対角項のみを用いた色変換 [55] についても同様の実験を行った．

実験結果を表 3.1 に示す．いずれの手法においても対話を用いた物体認識も含めると殆どの物体が抽出されている．しかし，自動物体認識に関しては我々の手法の方が高い認識率を示しているため，ユーザの負担を効果的に減らしていることが分かる．また，式 (3.29) で定義された S を観測総数で割った，正規化された色変換の自乗誤差の平均 $S_m = S/n$ についても，我々の手法の方が小さい値を示しており，対角項のみの変換よりも色変換の精度が高いことが分かる．

また，図 3.12(e) 以外の冷蔵庫のポケットのシーンや他の冷蔵庫を用いて同様の実験を行い，表 3.2 のような結果を得た．これより，いずれの冷蔵庫に対しても



(a) (c) (e)

(b) (d) (f)

(a),(c),(e): 冷蔵庫内部のシーン, (b),(d),(f): 冷蔵庫のポケットのシーン, 黒枠: 参照物体領域

図 3.12: 実験に用いた冷蔵庫内部とポケット

我々の手法がうまくいっていることが分かる．特に，参照色が薄緑色の冷蔵庫 (a) や薄黄色の冷蔵庫 (d) に対してもうまくいっていることから，ある程度明るい色であれば様々な参照色に対しても本手法がうまく対応していることが分かる．

3.7.2　新しい環境への適応実験

ここでは，初期段階における色変換の学習に関する実験について述べる．まず，予備実験として，変換誤差の共分散行列 Σ_i の学習に用いた冷蔵庫のシーンに対す

表 3.1: 各変換方法を用いたときの候補領域抽出の成功率 (%) と変換誤差

(a) 学習セット

手法	自動	色拡張	その他	S_m
対角	58.1	37.7	4.2	8.16
本手法	91.4	8.6	0.0	2.89

(b) テストセット

手法	自動	色拡張	その他	S_m
対角	58.3	37.3	4.4	8.21
本手法	91.2	8.8	0.0	3.01

自動: 自動物体認識で成功, 色拡張: 一度色拡張することによって成功, その他: 色拡張以外の対話によって成功または認識失敗

表 3.2: 他の冷蔵庫に対する候補領域抽出の成功率 (%)

冷蔵庫	手法	学習セット			テストセット		
		自動	色拡張	その他	自動	色拡張	その他
(a)	対角	48.7	43.5	8.8	45.9	40.8	13.3
	本手法	94.7	5.3	0.0	92.3	7.7	0.0
(b)	対角	83.1	16.9	0.0	82.0	18.0	0.0
	本手法	95.9	4.1	0.0	93.2	6.8	0.0
(c)	対角	46.9	47.7	5.4	46.9	47.5	5.6
	本手法	94.1	5.9	0.0	92.0	8.0	0.0
(d)	対角	32.5	36.2	31.3	30.9	39.7	29.4
	本手法	97.5	2.5	0.0	95.3	4.4	0.3
(f)	対角	65.0	34.3	3.2	69.6	30.0	0.4
	本手法	95.0	5.0	0.0	90.7	8.9	0.4

自動: 自動物体認識で成功, 色拡張: 一度色拡張することによって成功, その他: 色拡張以外の対話によって成功または認識失敗

る実験を行った．実験手順としては，最初に先の実験で用いた 114 枚の画像から抽出した 912 組の色の組を学習セット 456 組と実験セット 456 組にランダムに分けておく．次に，学習セットの色の組が順番に認識されていくものと想定して，色の組を追加する毎に推定パラメタを学習していき，その推定パラメタを実験セットに適用して認識を行う．

この予備実験を様々なセットに対して 10 回繰り返したときの認識率の平均を図 3.13 に示す．ここでは，例として $\alpha_0, \alpha_1, \alpha_2, \alpha_5, \alpha_8, \alpha_9, \alpha$ について，学習セットの内の 100 組まで学習したときの実験セットに対する認識率を示した．この図より拘束を解いたパラメタ数の少ない推定パラメタ (α_0〜α_2) では，最初からある程度の認識率を示しているものの，それから先の認識率は向上していない．これは，少ないパラメタでは色の変換を十分に表すことができないためである．これに対して，拘束を解いたパラメタ数の多い推定パラメタ ($\alpha_5, \alpha_8, \alpha_9$) では，最初の内は過学習のために低い認識率を示しているものの，学習する色の組が増加するにつれて認識率が向上していき，十分に学習すると少ないパラメタの認識率を上回っている．実際に用いる推定パラメタ α は学習セットの数の少ない間はパラメタ数の少ないパラメタを用いておき，そこから徐々にパラメタ数の多いセットへと遷移していくため，各段階で取り得る最大の認識率を示している．

ここで，十分に学習しても認識率が 90%程度にとどまっているのは，2.3.1 で述べたようにマッチングの基準を正規分布の 90%信頼区間で定義したためである．残りの 10%については冷蔵庫内の物体の位置による照明の当たり具合の変化による影響を受けたものである．これら認識の失敗については，ユーザとの対話によって回復する．

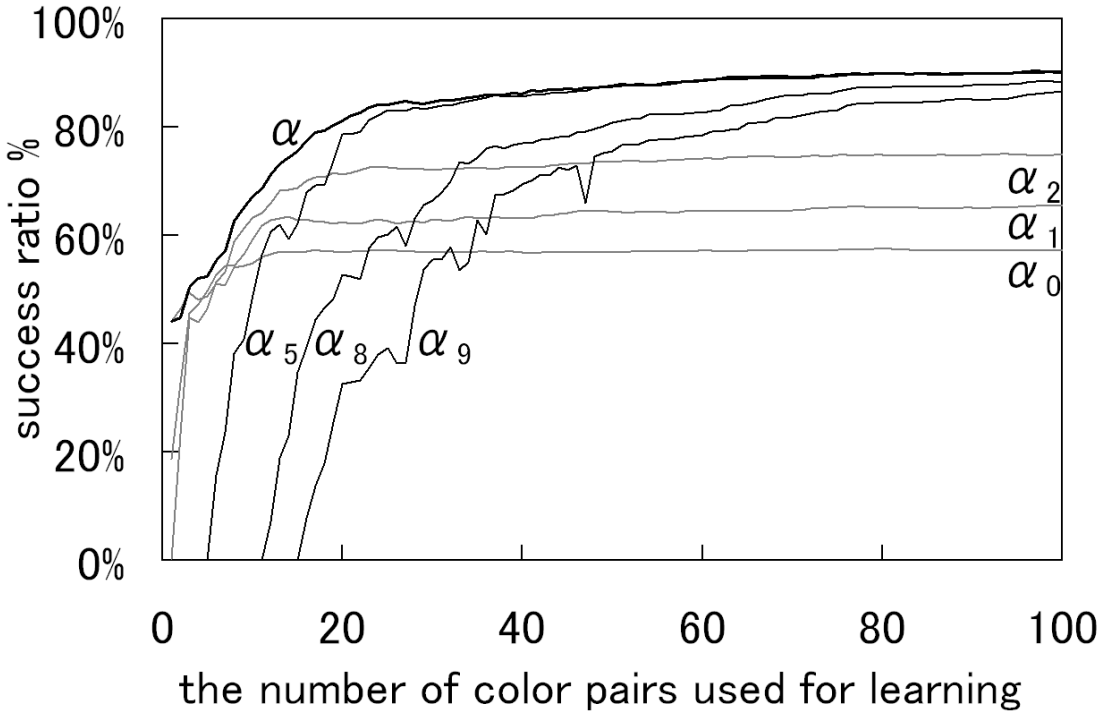
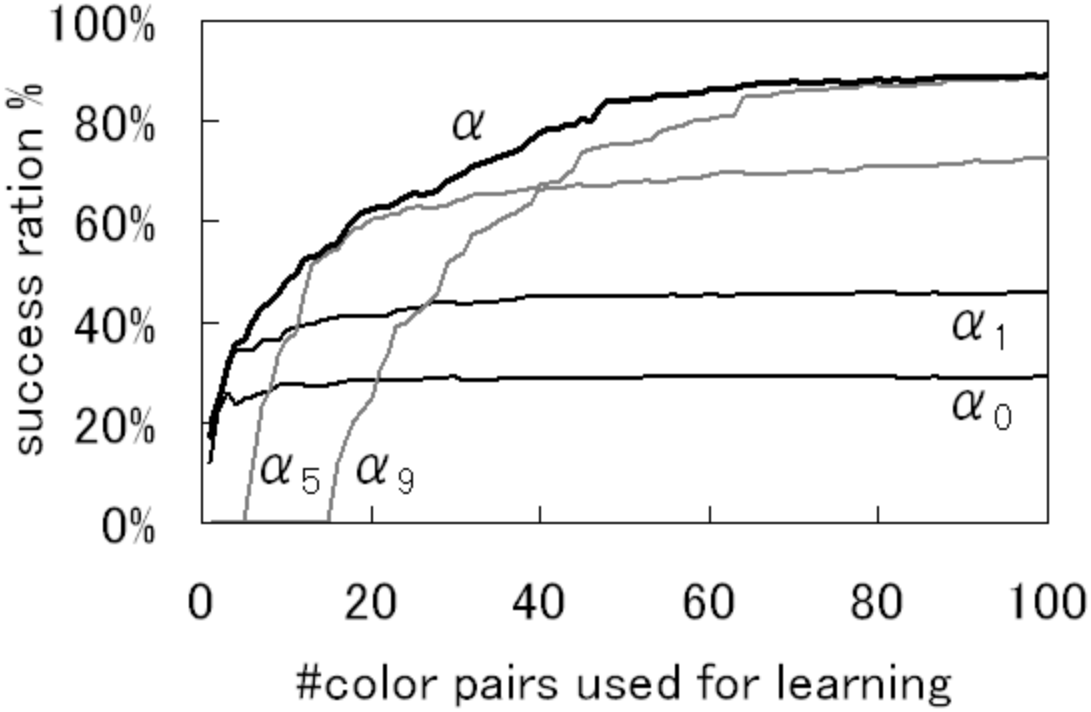


図 3.13: 初期段階における推定パラメタ学習の予備実験

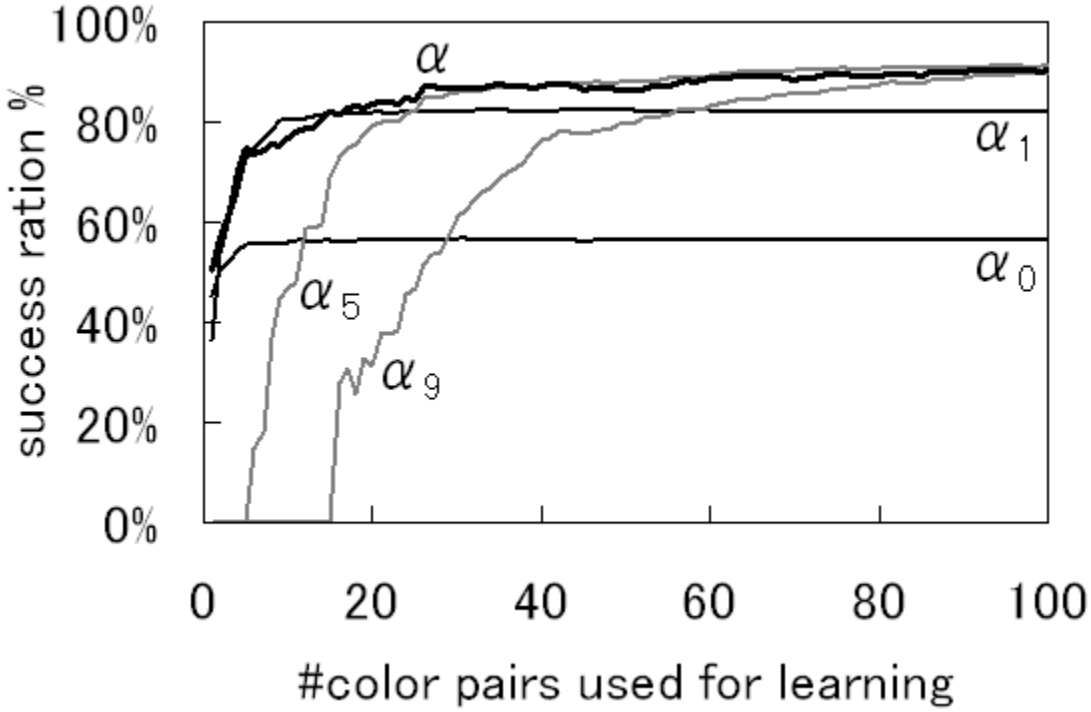
また，色変換誤差の共分散行列 Σ_i を学習していない環境として，予備実験に用いた冷蔵庫のポケットのシーンとそれとは別の冷蔵庫のシーンについて同様の実験を行った．冷蔵庫のポケットのシーンについては，物体を収納するための枠の部分を参照物体とした．これらの環境での参照色は予備実験に用いたほぼ白色の冷蔵庫の扉と違い，それぞれ薄い黄色と薄い緑色になっている．図 3.14 は各シーンに対する推定パラメタの学習結果について示している．予備実験の時と同様に学習の各段階で良い認識率を示しており，新しい環境にも適応できていることが分かる．

3.8 まとめ

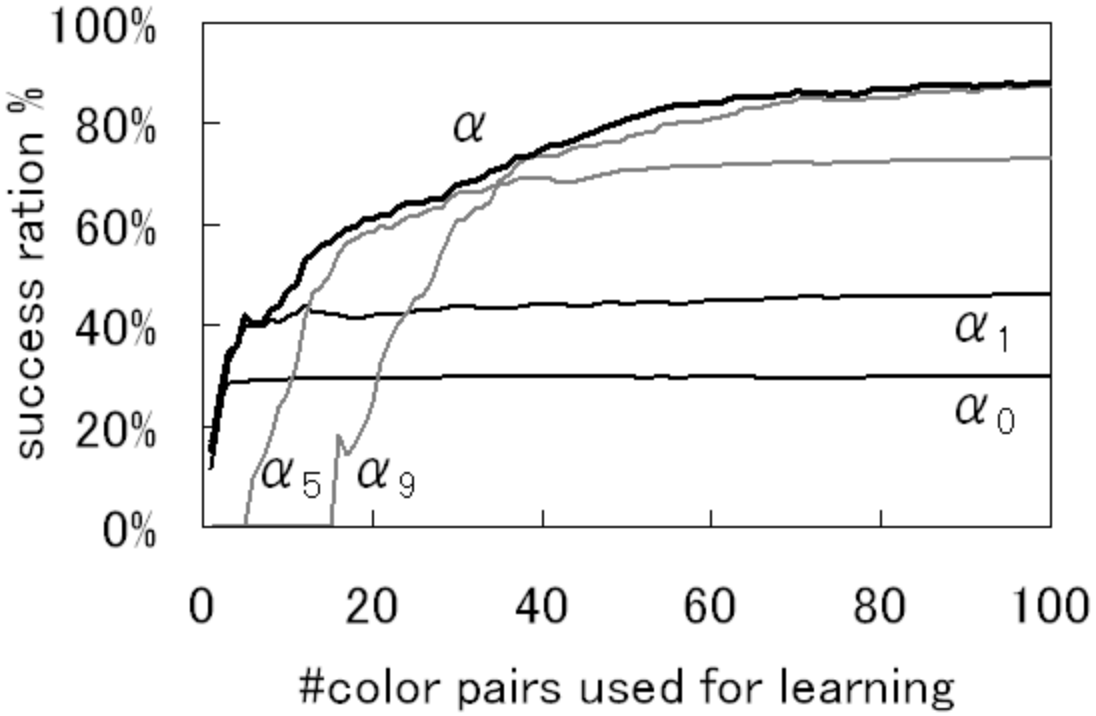
本章では，グローバルな外部照明変化を考慮した場合に，一つの参照物体の色を観測することで物体認識時の色変換を推定する手法を提案した．色の変換の推定パラメタを算出する際には，様々な照明条件下での誤差を平等に扱うために，モデル作成時の照明条件での誤差を物体毎の変換誤差の共分散行列で正規化したマハラノビス距離で評価した．また，学習用の色の組が十分に無い初期段階において，推定パラメタを徐々に追加して効率良く学習していく方法について述べた．そして，実験によって，十分に色の組があるときと十分な色の組が無い初期段階の両方について，提案した色変換の推定パラメタが有効であることを確認した．


100%
80%
60%
40%
20%
0%
success ration %
α
α 1
α 0
α 5
α 9
0
20
40
60
80
100
#color pairs used for learning

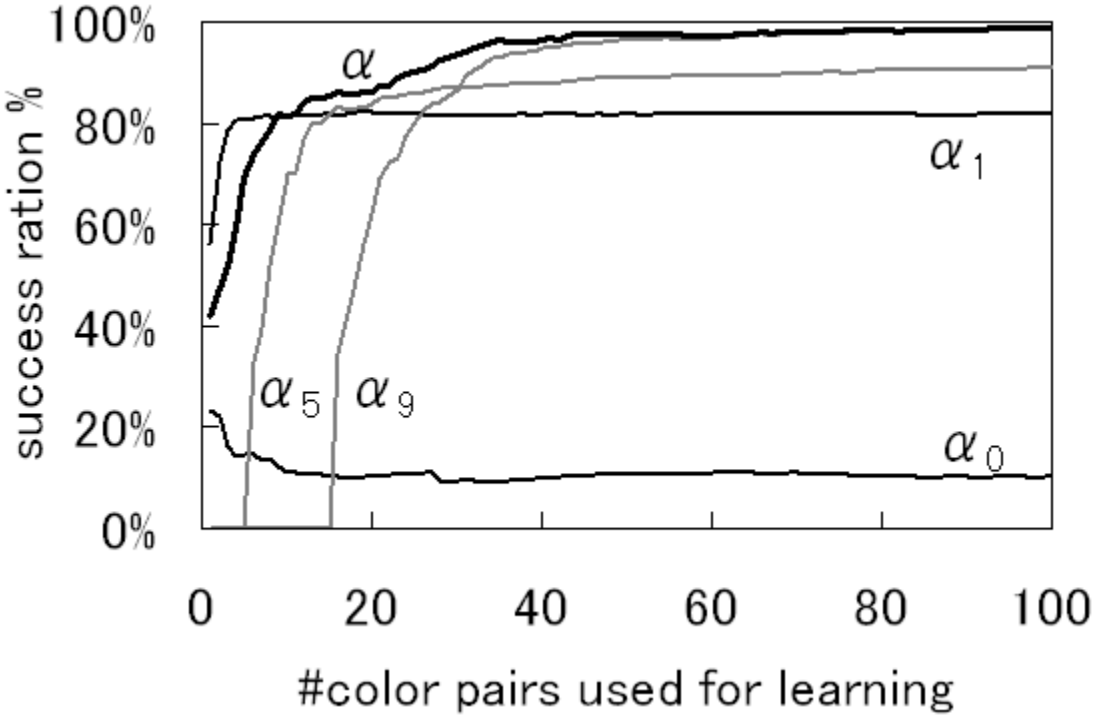
(a)


100%
80%
60%
40%
20%
0%
success ration %
α
α 1
α 0
α 5
α 9
0
20
40
60
80
100
#color pairs used for learning

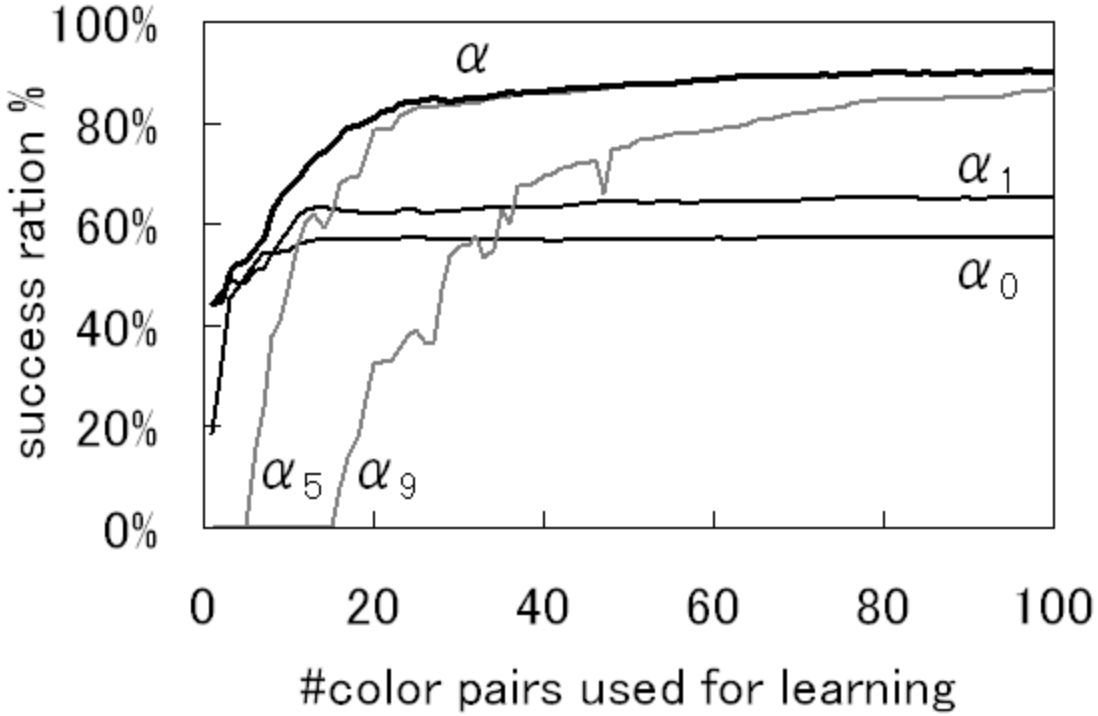
(b)


100%
80%
60%
40%
20%
0%
success ration %
α
α 1
α 0
α 5
α 9
0
20
40
60
80
100
#color pairs used for learning

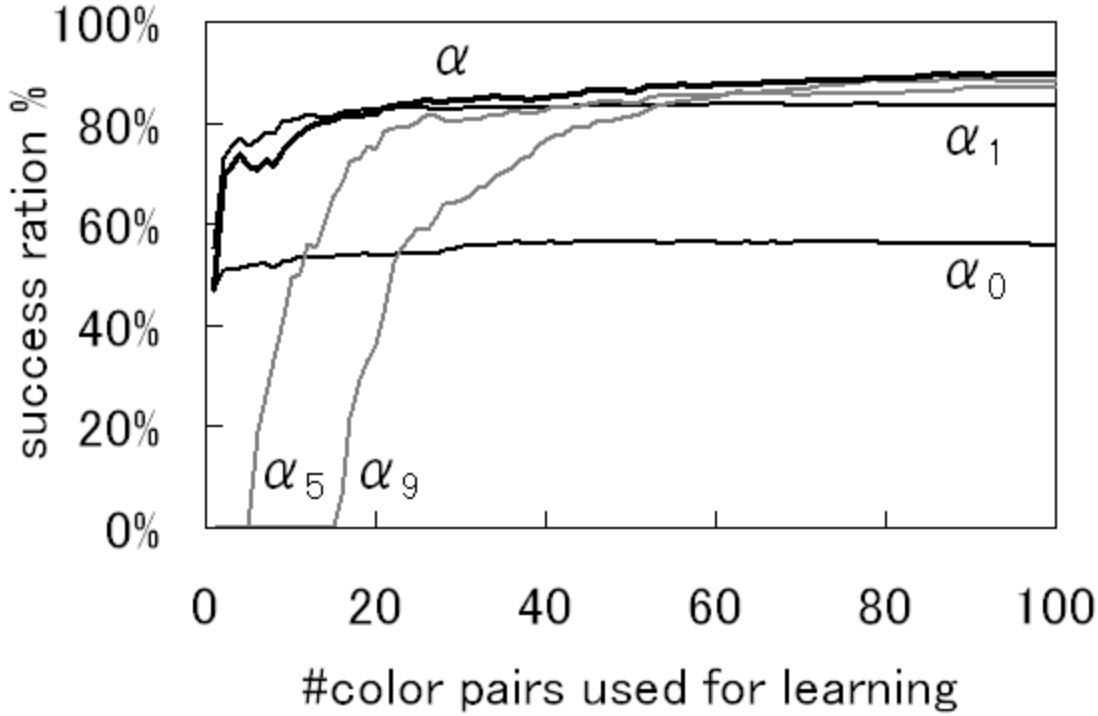
(c)


100%
80%
60%
40%
20%
0%
success ration %
α
α 1
α 5
α 9
α 0
0
20
40
60
80
100
#color pairs used for learning

(d)



(e)



(f)

図 3.14: 初期段階における推定パラメタの性能実験

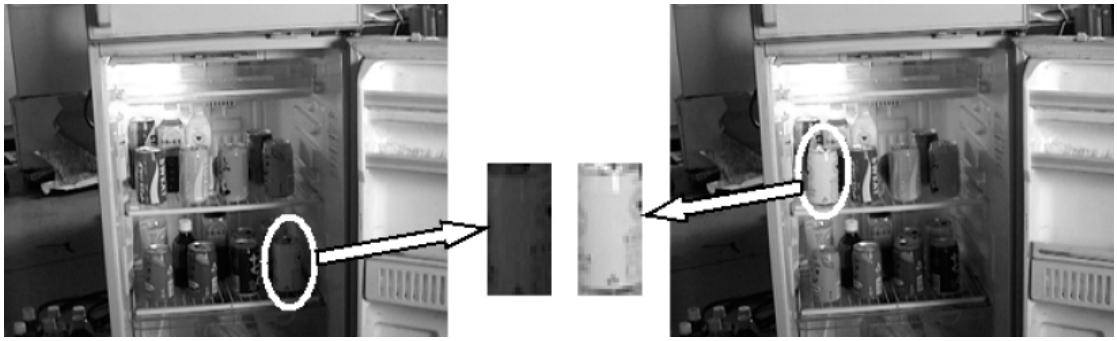
第4章　ローカルな照度変化を考慮した色変換の学習

本章では，第 3 章の手法をローカルな照度変化を考慮した場合に拡張する．最初に 4.1 節で関連研究との関わりを述べた後，4.2 節で，外部照明と内部照明のローカルな照明条件の変化を考慮した場合の色変換の推定方法について述べ，その学習方法を 4.3 で述べる．次に，4.4 節でローカルな照明条件を考慮した場合の物体認識手順について述べ，最後に 4.5 節で実験について述べる．

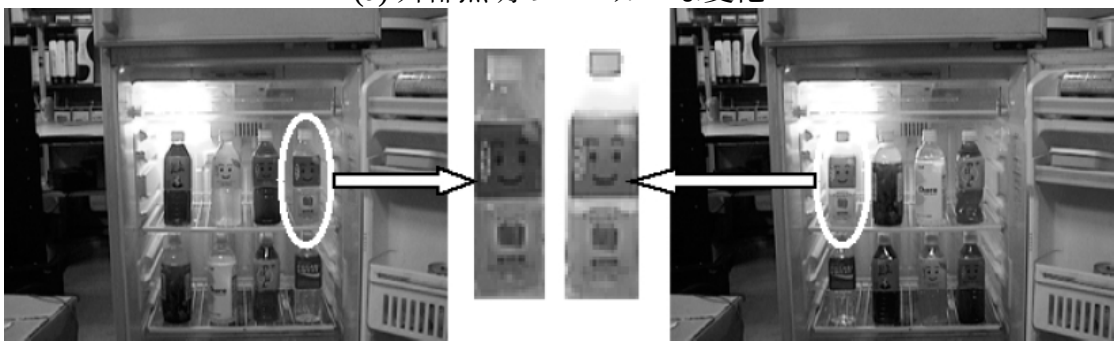
4.1　関連研究との関わり

シーン内でのローカルな照明変動を考慮したカラーコンスタンシーの研究としては，シーンを一様な反射率を持つ表面に領域分割して，各領域内での色変化から照明変動を推定する手法 [60] や，照明変動を利用して基準照明への色変換候補範囲を限定して，可能な範囲の重心を解とする研究 [61][62] がある．しかし，これらの手法は，対象シーンがモンドリアン [63] のような面積の大きな同一表面で構成されている必要があり，オクルージョンエッジがあるようなシーンには対応できない．

本研究では，このようなシーンに対する仮定を用いずにローカルな照明変動に対応することを目的とする．ここで扱う光源は，物体を前から照らす部屋の電灯や外からの光などの外部照明と，物体を後ろから照らす冷蔵庫内のランプによる内部照明の 2 種類とする．これらの照度は，図 4.1 のように同じ光源の下でも，シーン内の場所によって大きく変化する．これらの照明の影響による色変換を，物体モデルの色と様々な照明条件下で観測したその物体の色と参照物体の色の組を用いて学習する．まず，冷蔵庫内全体のグローバルな外部照明については，認識時に観測される参照物体の色を用いて推定する．冷蔵庫中での相対的な外部照明の照度変化については，光源との位置関係による影響が強いため，冷蔵庫内の位置との関係を学習する．一方，内部照明については，光源が全ての観測に対して共通であるため，その色成分は全ての観測に対して不変である．しかし，内部照明の照度については，光源との位置関係に加えて，物体自身の光の透過率や照明を遮る物体の有無による影響があり，冷蔵庫内の位置・物体の種類・物体の配置が変化するとその値が変化する．そこで，内部照明の色成分と照度を分けて扱い，色



(a) 外部照明のローカルな変化



(b) 内部照明のローカルな変化

図 4.1: 場所による照度変化

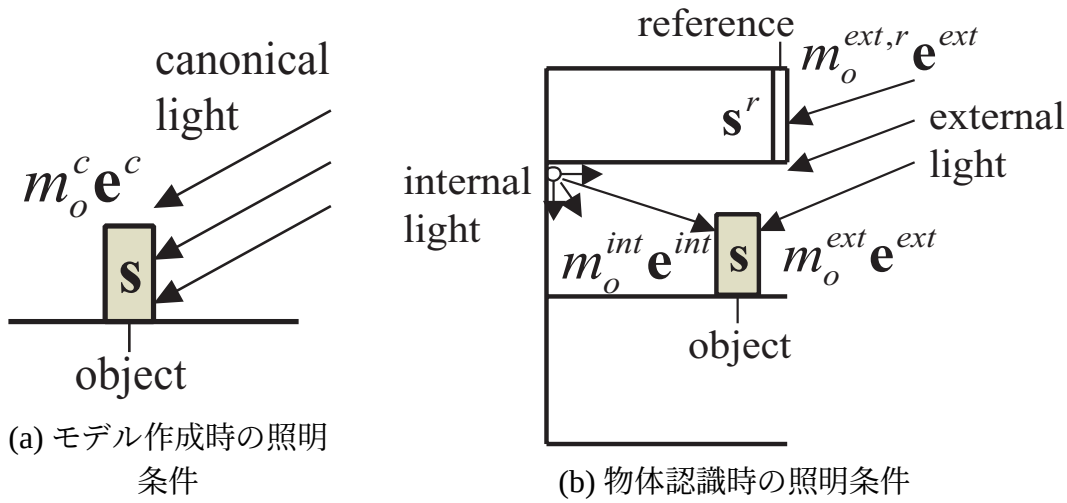
成分についてはすべての観測データに対して最も合うものを学習する．また，内部照明照度については，外部照明照度のように冷蔵庫内の位置との関係だけでは学習できないので，学習時には観測毎に最も合う内部照明照度を当てはめておき，物体認識時には，徐々に照度を変化させながら認識を試みて，認識が成功したときの照度をその場所の物体への内部照明照度とする．これにより，シーン内での照度変化に対応した正確な色変換を実現する．

4.2　ローカルな照度変化を考慮した色変換の推定

ここでは，シーン内での照度変化がある場合の色変換について定式化する．まず，色変換の問題設定について簡単に述べた後，半透明物体の色がどのように観測されるかをモデリングする．次に，そのような物体を含むシーンでの外部照明と内部照明の照度変化を考慮した色変換を定式化する．

4.2.1　色変換の問題設定

認識対象の物体モデルは図 4.2(a) のように，基準照明 (canonical light) 条件下で作成される．これに対して，物体認識時には図 4.2(b) のように部屋の照明の点灯・



(a) モデル作成時の照明条件　　(b) 物体認識時の照明条件

図 4.2: ローカルな照明変動を考慮した問題設定

消灯や時刻によってシーン毎に変化する外部照明 (external light) が物体を前方から照らし，冷蔵庫内のランプによる内部照明 (internal light) が物体を後方から照らす．グローバルな外部照明の変化については，参照物体 (reference) の色を観測することによって推定する．物体を後方から照らす内部照明については，ペットボトルのような半透明物体に対してのみ影響を与えるものと考える．

ここで，入射光 $\mathbf{i}$ を照度 $m_o = \|\mathbf{i}\|$ と色度 $\mathbf{e} = \mathbf{i}/\|\mathbf{i}\|$ の積 $\mathbf{i} = m_o\mathbf{e}$ に分解して考える．

外部照明の入射光 $m_o^{ext}\mathbf{e}^{ext}$ は，部屋の照明や窓の外からの光が光源として考えられるため，その照度 m_o^{ext} と色度 $\mathbf{e}^{ext}$ は時間によって変化する．また，これらの光源は冷蔵庫から比較的遠いので，冷蔵庫内の位置による色度 $\mathbf{e}^{ext}$ の変化は無いと見なすが，照度 m_o^{ext} については冷蔵庫の扉の影などの影響を受けて変化する．

一方，内部照明の入射光 $m_o^{int}\mathbf{e}^{int}$ は，時間変化しない冷蔵庫内のランプが光源なので，色度 $\mathbf{e}^{int}$ は時間によって変化しない．しかし，照度 m_o^{int} は内部照明を遮る物体の有無によって影響を受けるので，物体配置の変化を含む時間 (状況) の変化によって変化する．これらの入射光の照度と色度の性質を表 4.1 にまとめる．

結局問題は，認識時の照明条件下での外部照明と内部照明の照度と色度を推定して，認識時に撮影された画像をモデル登録時の基準照明条件下で得られるはずの画像に変換することとなる

4.2.2　半透明物体の色のモデリング

ペットボトルのラベルのように半透明な物体表面に入射した光は図 4.3 のように (a) 鏡面反射光，(b) 拡散反射光，(c) 鏡面透過光，(d) 拡散透過光に分かれる．(a) と (c) は，ある特定の方向に対して進み，表面の色の影響は受けずに光源色そのものを反映する．(b) と (d) は入射光が物体内部で散乱されて，あらゆる方向に拡散し

表 4.1: 入射光の照度と色度の性質 (○依存，×非依存)

入射光の要素	入射光の要素の依存要因	
	冷蔵庫内の位置	時間 (状況変化)
外部照明の照度 m_o^{ext}	○	○
外部照明の色度 $\mathbf{e}^{ext}$	×	○
内部照明の照度 m_o^{int}	○	○
内部照明の色度 $\mathbf{e}^{int}$	×	×

時間 (状況変化): 部屋の照明の点灯・消灯，時刻による外からの照明の変化や物体の配置の変化などの状況変化を含む

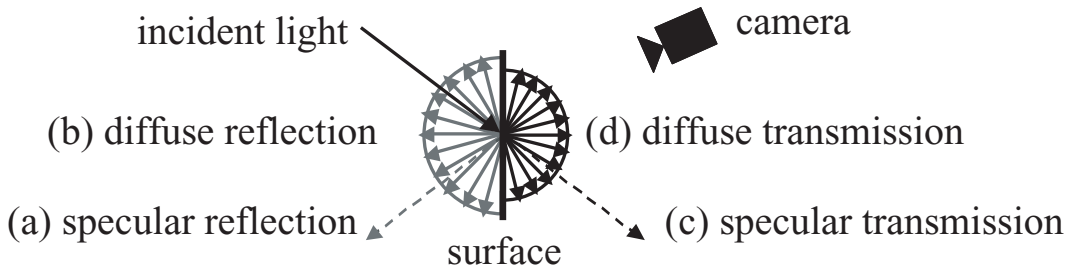


図 4.3: 半透明物体の色のモデリング

た光であり，光源の色に物体固有の色の影響を加えた光となる．その内，前方に散乱したものが (b) となり，後方に散乱したものが (d) となる．また，(b) と (d) については，前方・後方のそれぞれで，いずれの方向に対しても均等な拡散 (図 4.3 の左右の半円) が生じるものとする．

このように物理光学現象に基づいて半透明物体の色を推定する研究としては [69] があるが，透過光として，ガラス等の吸収性フィルターに対する鏡面透過光のみを考慮しており，拡散透過光については考慮していない．

本研究では，図 4.2(b) のように物体を後ろから照らすような内部照明による透過光を考える．本システムでは，冷蔵庫内を少し上から見下ろすような位置にあるカメラによってシーンを撮影するため，鏡面透過光に寄与する光はカメラと物体を結ぶ光軸上の冷蔵庫の壁や他の物体での反射光のみで，冷蔵庫のランプから物体に直接入射する光と光軸上以外の壁や物体での反射光は拡散透過光として寄与するため，カメラで観測される透過光としては拡散透過光が支配的になる．

このことを確かめるために，冷蔵庫の壁に黒画用紙を貼って鏡面透過光をカットした場合 (次頁の図 4.4(a)) と通常の場合 (図 4.4(b)) とでそれぞれ観測されるラベルの色を比較する予備実験を行った．図 4.5 は，実験に用いた物体 (O_1〜O_4) に対して，(a) と (b) のそれぞれの場合に対してペットボトルのラベル画像を抽出して，冷蔵庫のランプからの距離の順 (内部照明強度の強い順) で並べたものである．い

(a) 鏡面透過成分をカットした場合

(b) 鏡面透過成分をカットしなかった場合

図 4.4: 予備実験に用いた画像

ずれの場合にもランプから遠くなるにしたがって暗くなっているが，鏡面透過成分の有無に対してあまり色が変わっていないことが分かる．また，定量的な評価として，図 4.5 のラベルの代表色を基準照明条件に変換した色の比較のグラフを図 4.6 に示す．この図からも，鏡面透過成分の有無にかかわらずあまり色が変わっていないことが分かる．

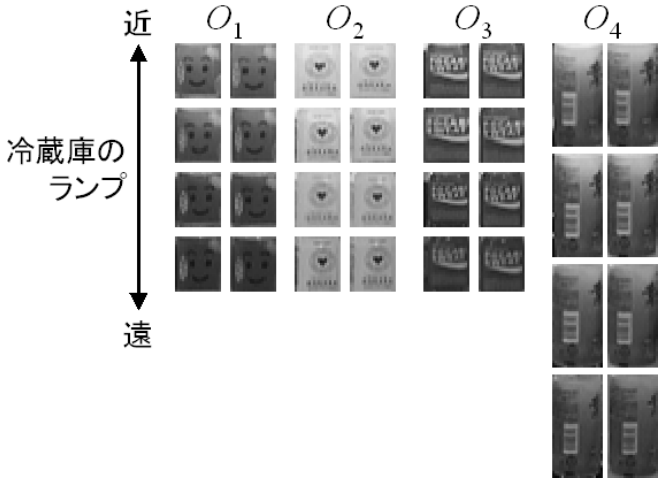
これより，ペットボトルのラベルを通して観測される透過光は拡散透過光であると近似することができる．また，拡散透過光と拡散反射光は同じ散乱現象によって同じ物体固有の色の影響を受けるため，波長に対する拡散透過率分布が拡散反射率分布の定数倍で表現できると仮定する．よって，拡散透過率分布を有限次元線形モデルで表現したときの基底関数の係数ベクトル $\mathbf{s}^{trans}$ を，拡散反射率の係数ベクトル $\mathbf{s}$ と定数倍の係数 t を用いて，$\mathbf{s}^{trans} = t\mathbf{s}$ として表す．結局，拡散透過光の色 $\mathbf{c}^t$ は式 (3.9) と同様にして以下で表される．

$$c_p^t = \sum_{k=1}^{4}\sum_{l=1}^{4} f_{pkl} t s_k i_l \tag{4.1}$$

4.2.3 照度変化を含む未知照明下での物体の色推定

冷蔵庫内に置かれた物体は，図 4.2(b) のように，前方からの外部照明 $m_o^{ext}\mathbf{e}^{ext}$ と後方からの内部照明 $m_o^{int}\mathbf{e}^{int}$ に照らされる．拡散反射率が $\mathbf{s}$ で拡散透過率が $t\mathbf{s}$ の物体がこのような照明に照らされたときにカメラで観測される物体の色 $\mathbf{c}^o$ は，外部照明による色 $\mathbf{c}^{ext,o}$ と内部照明による色 $\mathbf{c}^{int,o}$ の和として以下のように表される．

$$\begin{aligned} c_p^o &= c_p^{ext,o} + c_p^{int,o} \\ &= \sum_{k=1}^{4}\sum_{l=1}^{4} f_{pkl}(s_k m_o^{ext} e_l^{ext} + t s_k m_o^{int} e_l^{int}) \end{aligned} \tag{4.2}$$



左列: (a) 鏡面透過成分をカットした場合，右列: (b) 鏡面透過成分をカットしなかった場合

図 4.5: ペットボトルのラベル画像の比較

一方，その物体のモデル登録は，図 4.2(a) のように基準照明 $m_o^c \mathbf{e}^c$ の下で行うので，モデルの色 $\mathbf{c}^m$ は，式 (3.10) より，以下のようになる．

$$\mathbf{c}^m = m_o^c A^c \mathbf{s},\ where\ (A^c)_{pk} = \sum_{l=1}^{4} f_{pkl} e_l^c \tag{4.3}$$

また，認識時において，表面反射率 $\mathbf{s}^r$ を持つ参照物体は図 4.2(b) のように外部照明 $m_o^{ext,r} \mathbf{e}^{ext}$ のみを受けるので，観測される参照色 $\mathbf{c}^r$ は，式 (3.10) より，

$$\mathbf{c}^r = m_o^{ext,r} B^r \mathbf{e}^{ext},\ where\ (B^r)_{pl} = \sum_{k=1}^{4} f_{pkl} s_k^r. \tag{4.4}$$

式 (4.3), (4.4) より，物体の拡散反射率 $\mathbf{s}$ と外部照明の入射光の色度 $\mathbf{e}^{ext}$ は以下のようになる．

$$\mathbf{s} = (A^c)^{-1} \mathbf{c}^m / m_o^c \tag{4.5}$$

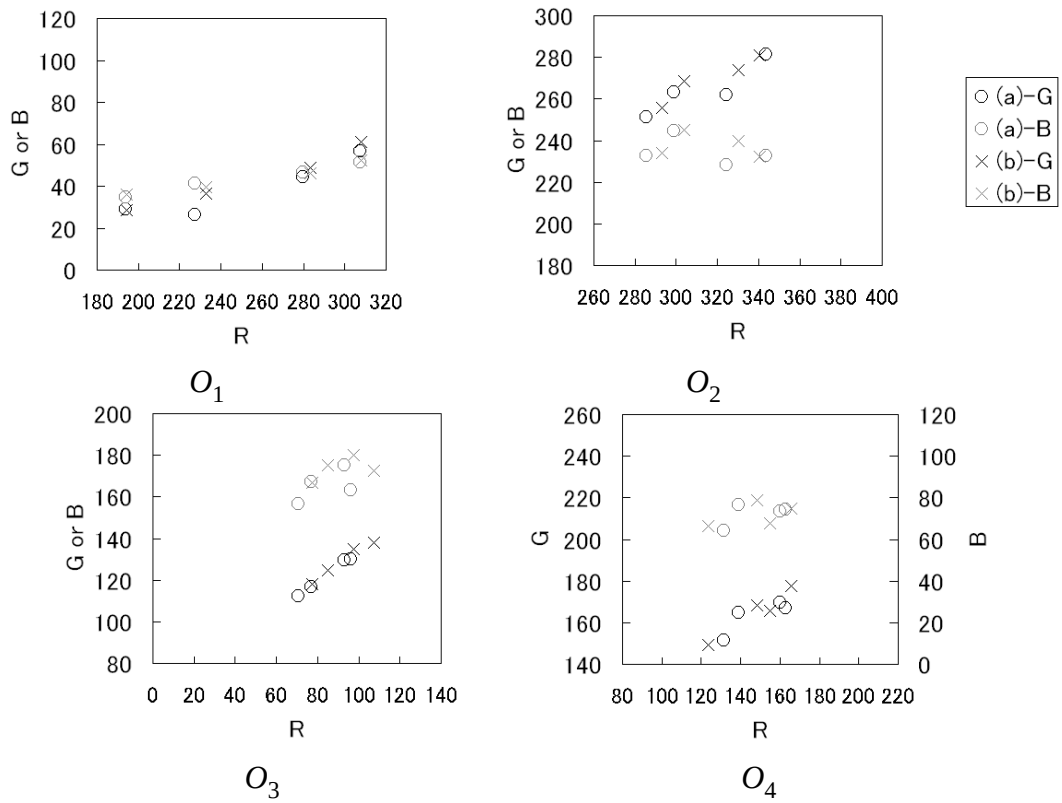
$$\mathbf{e}^{ext} = (B^r)^{-1} \mathbf{c}^r / m_o^{ext,r} \tag{4.6}$$

式 (4.2) に式 (4.5), (4.6) を代入すると，

$$\mathbf{c}^o = T\,\mathbf{c}^m \tag{4.7}$$

$$T = m^{ext} T^{ext} + m^{int} T^{int} + T^h \tag{4.8}$$

$$(T^{ext})_{pu} = (1 - \delta_{p4}) \sum_{v=1}^{4} \alpha_{puv} c_v^r \tag{4.9}$$



○: (a) 鏡面透過成分をカットした場合，×: (b) 鏡面透過成分をカットしなかった場合

図 4.6: 基準照明条件下での物体の代表色の比較

$$(T^{int})_{pu} = (1-\delta_{p4})\sum_{k=1}^{4}\sum_{l=1}^{4} f_{pkl}((A^c)^{-1})_{ku} e_l^{int}) \tag{4.10}$$

$$T^h = diag[0,0,0,1] \tag{4.11}$$

$$\alpha_{puv} = \sum_{k=1}^{4}\sum_{l=1}^{4} f_{pkl}((A^c)^{-1})_{ku}((B^r)^{-1})_{lv}/m_o^c \tag{4.12}$$

$$m^{ext} = m_o^{ext}/m_o^{ext,r} \tag{4.13}$$

$$m^{int} = tm_o^{int}/m_o^c. \tag{4.14}$$

ここで，T はモデルの色 $\mathbf{c}^m$ を認識時の物体の色 $\mathbf{c}^o$ に変換するための色変換行列であり，外部照明と内部照明に関する色変換行列 T^{ext}, T^{int} の重みつき和と T^h の和で表される．δ_{ij} はクロネッカーのデルタであり，T^{ext}, T^{int} の第 4 行成分はいずれも 0 になる．色変換行列の第 4 行成分をこのように表現したのは，モデルの色 $\mathbf{c}^m = [c_1^m, c_2^m, c_3^m, 1]^T$ を変換した $\mathbf{c}^o$ の第 4 行目の成分が常に 1 になるようにするためである．また，α は外部照明による色変換行列 T^{ext} を認識時の参照色 $\mathbf{c}^r$ から推定するためのパラメタである．m^{ext} は参照物体に対して認識対象物体がどの程度

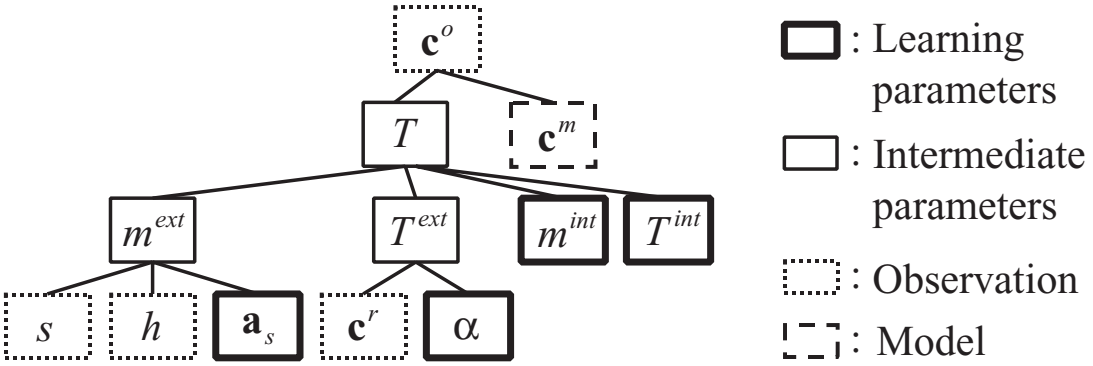


図 4.7: 色変換要素の関係

の外部照明を受けるかを相対的に表したものであり，m^{int} はモデル登録時の物体への入射光に比べて，どの程度の内部照明の影響を受けるかを相対的に表したものである．ただし，缶のように光が透過しない物体については，その内部照明重みを $m^{int} = 0$ としておく．これらの重み m^{ext}, m^{int} によって，冷蔵庫内の場所による照度変化の影響を表すことができる．

ただし，外部照明重み m^{ext} の変化については，冷蔵庫内の場所による影響が最も大きいと考えられるので，冷蔵庫の棚毎に左端を 0，右端を 1 とする横方向の相対位置 h への依存性を学習することができる．ここでは，外部照明重み m^{ext} を実験的に以下のような h の 2 次式で近似する．

$$m^{ext} = \sum_{j=0}^{2} a_{s,j} h^j \tag{4.15}$$

ここで，$a_{s,j}$ は冷蔵庫の s 番目の棚の横方向の相対位置 h の j 次の係数である．また，冷蔵内での奥行きによっても外部照明重み m^{ext} は変化するが，学習時には奥行き一定を仮定した本手法のモデルを用いてパラメタを求め，その仮定によって生じた色変換誤差の分散を元にして，認識時の奥行き変化による誤差を許容できるようにする (詳細は次章以降を参照).

これにより，パラメタ $\mathbf{a}_s, \alpha, m^{int}, T^{int}$ が求まっていれば，式 (4.8), (4.9), (4.15) より認識時に観測される参照色 $\mathbf{c}^r$ と冷蔵庫内の棚 s と位置 h から色変換行列 T を定めて，式 (4.7) より物体モデルの色 $\mathbf{c}^m$ を認識時の物体の色 $\mathbf{c}^o$ に変換することができる．これらの色変換要素の関係を図 4.7 にまとめる．また，これらの色変換要素の，冷蔵庫内の位置，物体の種類，時間に対する依存性を表 4.2 にまとめる．

ここで，色変換行列 T には，パラメタの積の項 $\mathbf{a}_s\alpha$ と $m^{int}T^{int}$ が存在して，しかも各パラメタに制約が無いために，パラメタが一意に決まらない．そこで，α, T^{int} に制約 $\|\alpha\| = 1,\ \|T^{int}\| = 1$ を加えることで，パラメタの組を一意に決定する．

表 4.2: 色変換要素の依存性 (○依存，×非依存)

要素	位置	物体	時間	要素	位置	物体	時間
$\mathbf{a}_s$	×	×	×	m^{int}	○	○	○
$\mathbf{c}^m$	×	○	×	s	○	×	×
$\mathbf{c}^o$	○	○	○	T	○	○	○
$\mathbf{c}^r$	×	×	○	T^{ext}	×	×	○
h	○	×	×	T^{int}	×	×	×
m^{ext}	○	×	×	α	×	×	×

位置: 冷蔵庫内の位置，物体: 物体の種類，時間: 状況変化

4.3 ローカルな照度変化を考慮した色変換の学習

最初に 4.3.1 節で色変換の学習方法について述べ，次に 4.3.2 節で学習した色変換の結果について述べる．

4.3.1 色変換の学習方法

まず求めるべき色変換パラメタ $\mathbf{a}_s, \alpha, m^{int}, T^{int}$ の個数について考える．α は従来法で扱ったグローバルな外部照明を推定するパラメタであり，表 4.2 にあるように位置・物体・時間の異なる観測に対して共通である．よって，その個数は先の場合と同様に 48 である．$\mathbf{a}_s$ は冷蔵庫内の位置と外部照明重み m^{ext} の関係を表すパラメタであり，表 4.2 より位置・物体・時間の異なる観測に対して共通である．ただし，式 (4.15) より冷蔵庫の棚毎にパラメタが異なるので，冷蔵庫の棚の数を n_s とすると，パラメタ $\mathbf{a}_s$ の個数は $3n_s$ となる．T^{int} は内部照明による色変換を表すパラメタであり，表 4.2 より位置・物体・時間の異なる観測に対して共通である．よって，その個数は 12 である[1]．m^{int} は内部照明重みを表すパラメタであり，表 4.2 に示すように位置・物体・時間の異なる観測に対して変化するため，その個数は観測回数 n と等しくなる．結局，n 回の観測が得られたときに求めるべきすべてのパラメタ数は，$60+3n_s+n$ である．

これより，パラメタを算出するためには，観測数が $n_c = 30+(3/2)n_s$ だけ必要である．実際にパラメタを計算する際には n_c 以上の観測値を用いて，前節と同様にして基準照明での色変換誤差を物体毎の色変換誤差の共分散行列で正規化したマハラノビス距離の総和を評価値に用いる．結局，最小化すべき評価式 S は以下

[1] T^{int} は T^{ext} と同様に，第 4 行目の成分については 0 で固定なので，実質 $3\times 4 = 12$ である．

のようになる．

$$
\begin{aligned}
S &= \sum_{j=1}^{n} \Delta_j^T \Sigma_{O(j)}^{-1} \Delta_j \rightarrow min \qquad (4.16)\\
under &\quad \|\alpha\| = 1,\ \|T^{int}\| = 1,\ m_j^{ext} \geq 0,\ m_j^{int} \geq 0\\
where &\quad \Delta_j = \mathbf{c}_j^m - T_j^{-1}\mathbf{c}_j^o\\
&\quad \Sigma_{O(j)} = \frac{1}{n_{O(j)}} \sum_{k \in I_{O(j)}} \Delta_k \Delta_k^T \qquad (4.17)
\end{aligned}
$$

ここで，式(4.16)の最小化問題について注目すると，式(4.17)の共分散行列 $\Sigma_{O(j)}$ を計算するためには S の最小化により色変換パラメタが求まっている必要がある．一方，その最小化を行うためには共分散行列 $\Sigma_{O(j)}$ が必要となる．そこで，以下の手順で S の最小化を解く．

1. 共分散行列 $\Sigma_{O(j)}$ を単位行列 I_4 に初期化
2. $\mathbf{a}_s, m_j^{int}, T^{int}$ を固定値に拘束して，α を S の最小化により算出
3. すべての色変換パラメタ $\mathbf{a}_s, \alpha, m_j^{int}, T^{int}$ を S の最小化 (共役勾配法) により算出
4. 共分散行列 $\Sigma_{O(j)}$ を算出
5. 色変換パラメタと共分散行列の値が収束するまで，手順 (3)，(4) を反復

手順 (2) は，冷蔵庫内の位置で変化しない外部照明のみを考慮した色変換パラメタを算出したことに相当し，手順 (3) の共役勾配法による最小化の初期値を与えるためのものである．

また，本システムを新しい環境 (冷蔵庫) に適応させる場合は，3.5 節で述べたグローバルな外部照明変化の推定パラメタ α に加えて，外部照明重みの推定パラメタ $\mathbf{a}_s$ や内部照明のパラメタ T^{int}, m_j^{int} についても考慮する必要がある．そこで，最初は冷蔵庫内の位置で変化しない外部照明のみを考慮して，$\mathbf{a}_s, T^{int}, m_j^{int}$ を固定値に拘束した上で，α のみを更新する．この適応過程で，(A) 外部照明照度のローカルな変化の影響を受けた観測や (B) 内部照明の影響を受けた観測が得られると，α のみでは色変換誤差が大きくなる．そこで，誤差が大きくなったときには，その誤差の主な要因が (A)，(B) のいずれによるものかを調べて，より大きな影響を与える要因に対応するパラメタ ((A) に対しては $\mathbf{a}_s$，(B) に対しては m_j^{int}, T^{int}) の拘束を解いて，そのパラメタについても学習する．具体的には，$\mathbf{a}_s$ の拘束を解いた場合と，m_j^{int}, T^{int} の拘束を解いた場合で，誤差がより小さくなる方の拘束を解くものとする．また，それ以降で色変換誤差が大きくなることがあれば，残りのパラメタの拘束を解いて，そのパラメタについても学習する．

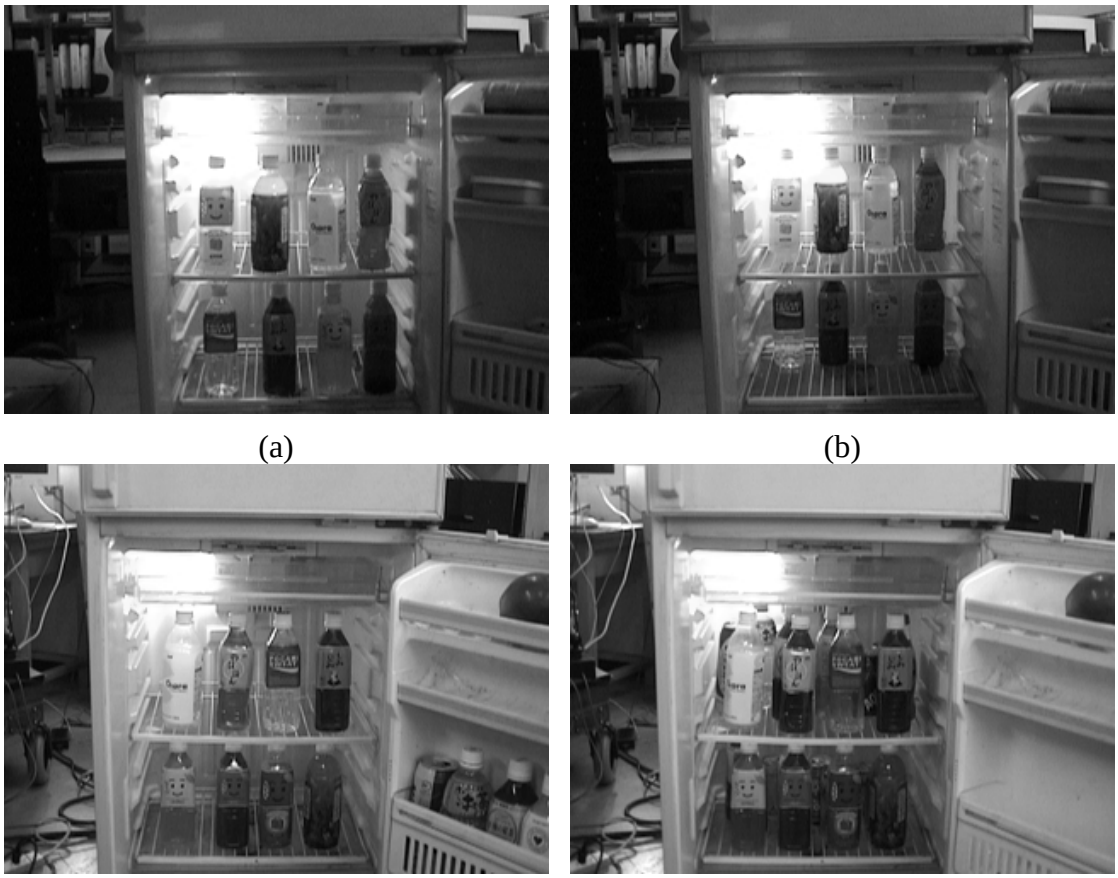
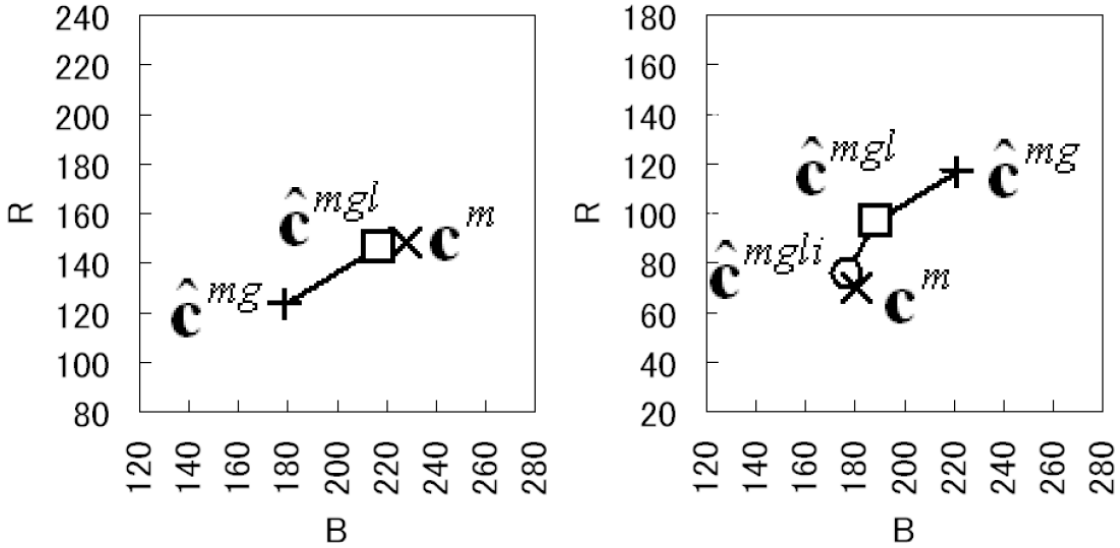


図 4.8: 学習サンプル画像

4.3.2　色変換の学習結果

学習には図 4.8 のように，様々な照明条件で撮影されたサンプル画像を用いた．また，物体としては缶・ペットボトル毎に様々な色のものを 8 種類ずつを用いた．ここでは，160 枚サンプル画像より学習用に各々640 の観測値を取得して，学習を行った．

まず，外部・内部照明重みの変化を考慮することによる色推定への影響例を図 4.9 に示す．ここで，$\mathbf{c}^m$ は物体モデルの色，$\hat{\mathbf{c}}^{mg}$ は冷蔵庫内の位置によって変化しない外部照明のみを考慮した場合 (以下，Global と表記) に観測色を基準照明へ変換した色，$\hat{\mathbf{c}}^{mgl}$ は冷蔵庫内の位置による外部照明重み変化を考慮した場合 (以下，Global + Local と表記) の色，$\hat{\mathbf{c}}^{mgli}$ はペットボトルに対して観測毎に最も良い内部照明重みを与えた場合 (以下，Global + Local + Int. light と表記) の色を表す．図 4.9(a) は缶が扉の影で暗くなる位置に置かれていたために，$\hat{\mathbf{c}}^{mg}$ がモデル $\mathbf{c}^m$ よりも暗くなってしまうのに対して，場所による外部照明重みを学習した $\hat{\mathbf{c}}^{mgl}$ は，モデルの色 $\mathbf{c}^m$ に近くなっていることが分かる．また，図 4.9(b) は比較的外部照明の良く当たる場所にあるペットボトルが，後ろから内部照明によって照らされてい



(a) 缶　　(b) ペットボトル

図 4.9: 外部・内部照明重み変化を考慮するによる色推定への影響

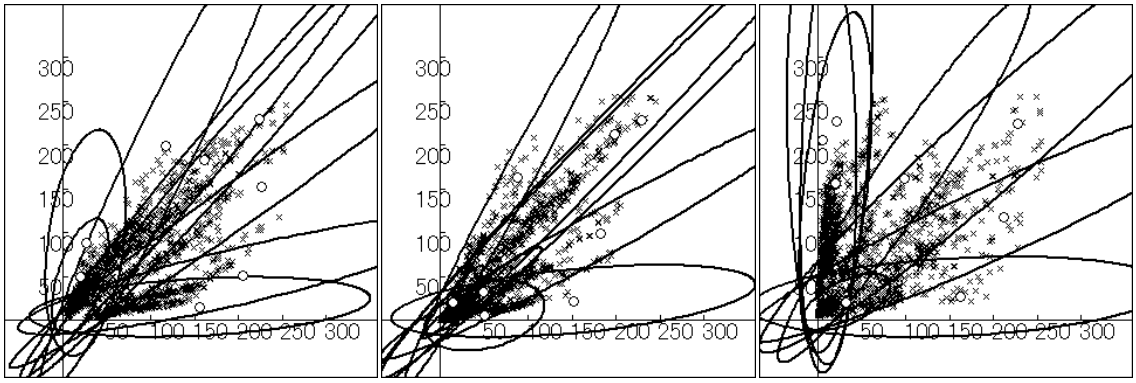
るため，従来法で推定した $\hat{\mathbf{c}}^{mg}$ はモデル $\mathbf{c}^{m}$ に比べてかなり明るくなっている．それに対して，場所による外部照明重みを学習した $\hat{\mathbf{c}}^{mgl}$ は多少モデル $\mathbf{c}^{m}$ に近づき，さらに内部照明の適応を行った $\hat{\mathbf{c}}^{mgli}$ はモデルの色 $\mathbf{c}^{m}$ に非常に近い値になっていることが分かる．

また，缶とペットボトルの各々について，学習データ全体の色変換誤差の比較を図 4.10，4.11 に示す．この図は，学習に用いた冷蔵庫内の物体の観測色を，各手法によって基準照明条件下での色に変換して示し，物体毎の色変換誤差の共分散行列を楕円で表示したものである．色変換誤差が無ければすべての点はモデルの色を表す楕円の中心点に集まり，色変換誤差が小さい程この楕円も小さくなる．

缶について注目してみると，色変換を行わない場合には図 4.10(a) のように照明条件の変化によって観測色が様々に変化するために，物体モデルとの色の差は非常に大きくなる．これに対して，グローバルな外部照明を考慮して色変換を行うと，同図 (b) のように物体毎に色がまとまり色変換誤差が小さくなっていることが分かる．さらに，ローカルな外部照明変化についても考慮すると，同図 (c) のように色変換誤差がより小さくなっていることが分かる．

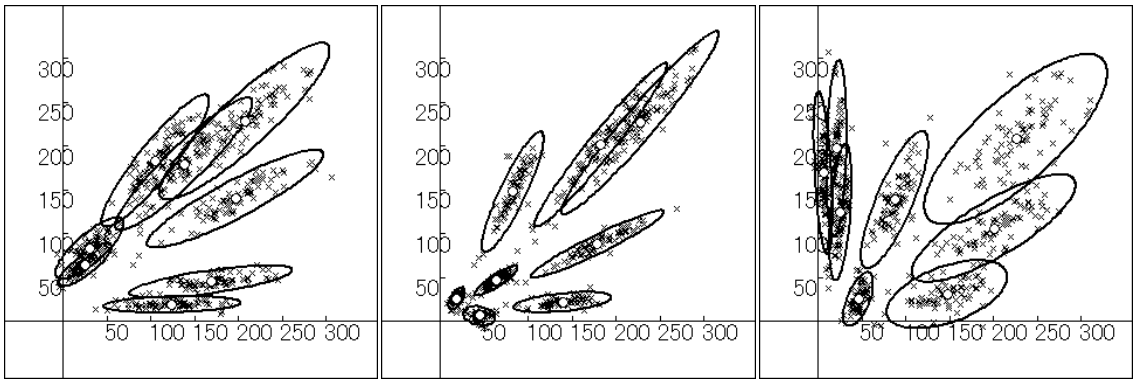
一方，ペットボトルについて注目すると，缶の場合と同様に図 4.11(a) のように，物体モデルの色との差が大きくなっていることが分かる．これに対して，グローバルな外部照明，グローバルとローカルな外部照明による影響を考慮したものが各々同図 (b),(c) であり，色変換誤差が小さくなっている．しかし，ペットボトルの場合は内部照明の影響を受けるため，この段階ではまだ大きな色変換誤差が残っている．ここで，さらに内部照明についても考慮すると同図 (d) のように，色変換誤差が非常に小さくなっていることが分かる．

これより，ローカルな外部照明と内部照明を考慮することにより，色変換誤差



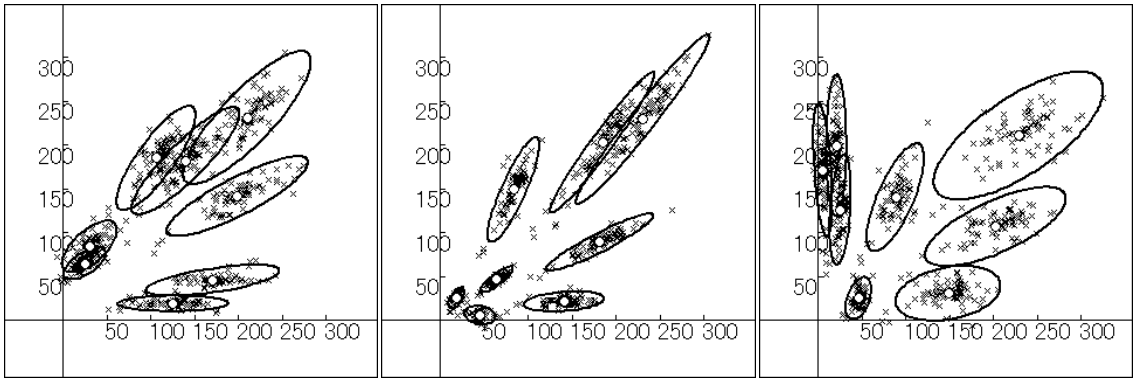
R-G　　G-B　　B-R

(a) 色変換無し



R-G　　G-B　　B-R

(b) Global

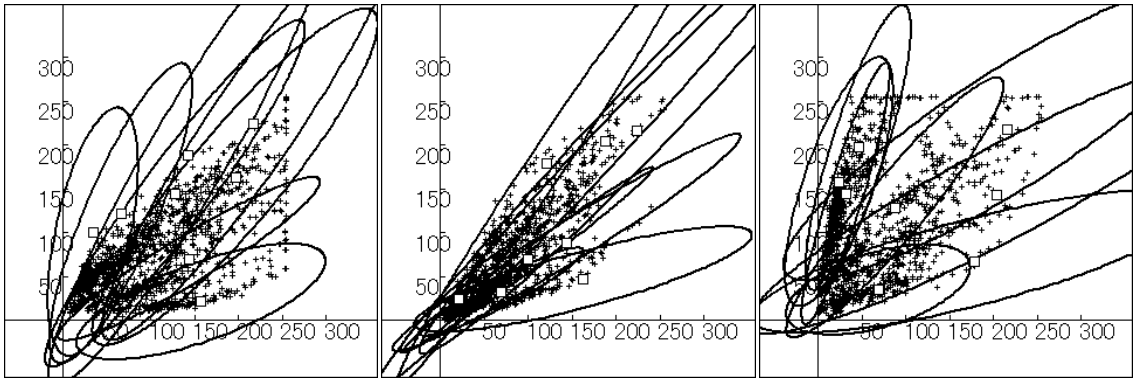


R-G　　G-B　　B-R

(c) Global + Local

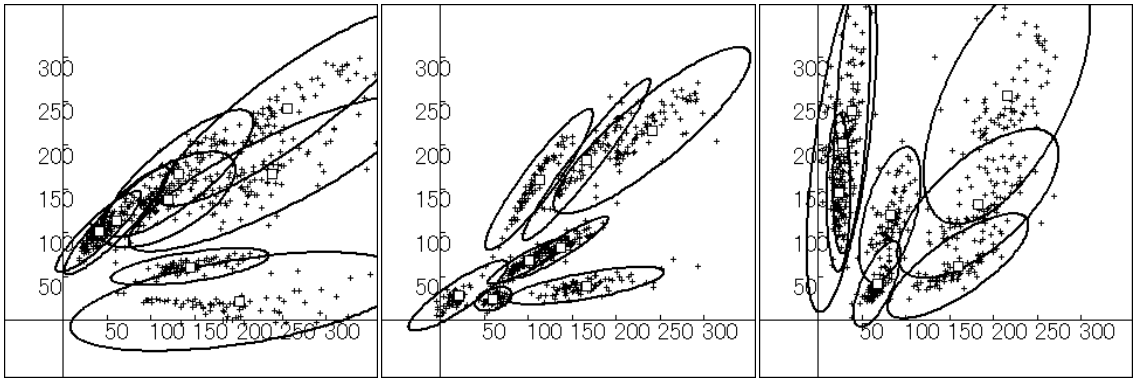
プロット: 基準照明に変換した物体の観測色，楕円: 物体毎の色変換誤差の共分散，白丸: 物体モデルの色

図 4.10: 基準照明下での缶の色変換誤差



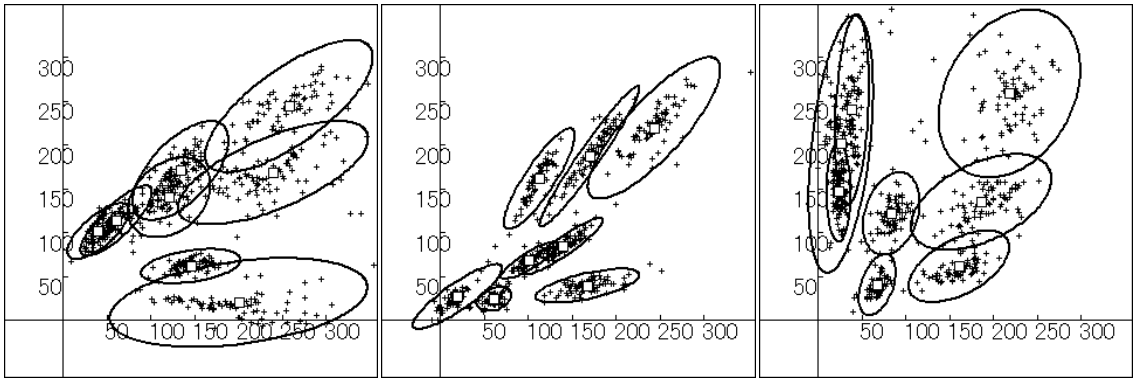
R-G　　G-B　　B-R

(a) 色変換無し



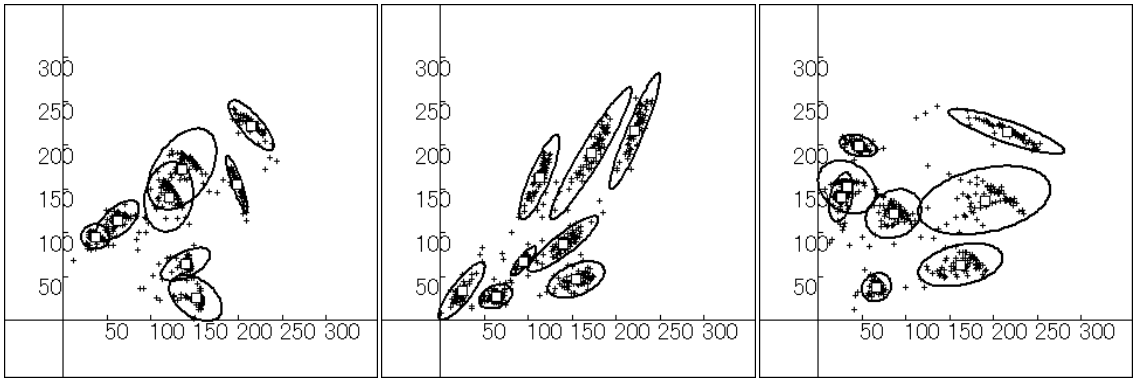
R-G　　G-B　　B-R

(b) Global



R-G　　G-B　　B-R

(c) Global + Local



R-G　　G-B　　B-R

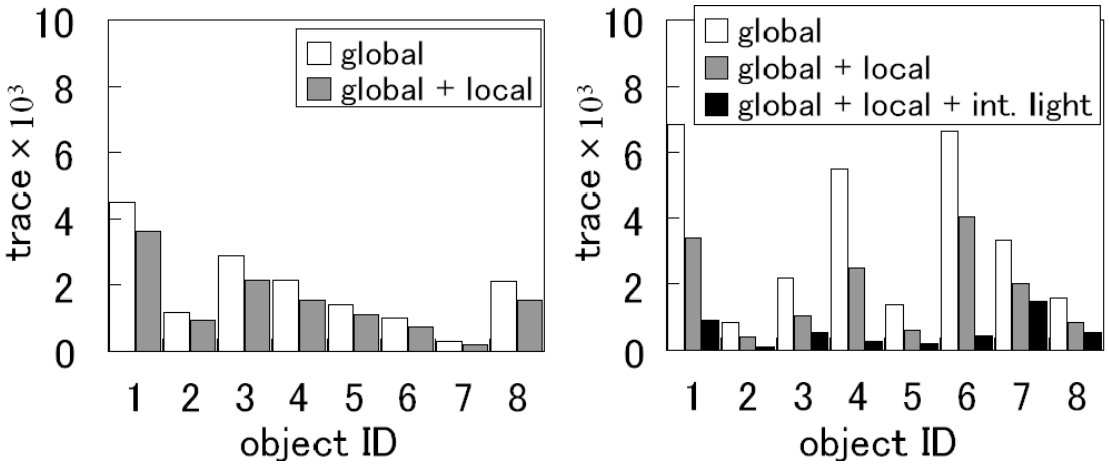
(d) Global + Local + Int. light

プロット: 基準照明に変換した物体の観測色，楕円: 物体毎の色変換誤差の共分散，白矩形: 物体モデルの色

図 4.11: 基準照明下でのペットボトルの色変換誤差

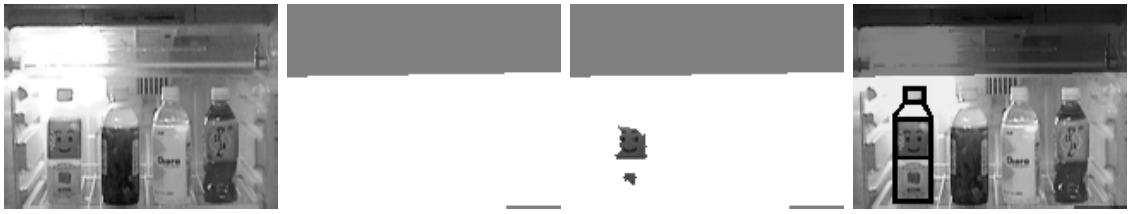
を小さくすることができ，正確な色変換を実現することができる．

また，定量的評価として，図 4.12 に色変換誤差の共分散行列のトレースを示す．これより，定量的にも本手法によって色変換誤差を小さくすることができることが分かる．



(a) 缶　　(b) ペットボトル

図 4.12: 色変換誤差の共分散行列のトレース



(a) 原画像　(b) 代表色領域 (m^{int} = 0.0)　(c) 代表色領域 (m^{int} = 0.7)　(d) 認識結果

図 4.13: 物体認識例 (拡大図)

4.4　ローカルな照度変化を考慮した物体認識

これまでに述べたように内部照明重み m^{int} は観測毎に求める必要があるために，物体認識時には様々な値を与えてみて，対象物体をうまく認識できたときの値を採用する．その際の認識手順は以下のようになる．

1. 式 (4.15) より，外部照明重み m^{ext} を冷蔵庫内領域の各点で算出
2. 内部照明重み m^{int} を 0 に初期化
3. 式 (4.8), (4.9) より，冷蔵庫内領域の各点で色変換行列 T を算出
4. 原画像を色変換行列の逆行列 T^{-1} でモデル登録時の基準照明条件へ逆変換して，正規化画像を作成
5. 正規化画像で物体認識を行い，成功したら終了
6. 失敗時は，缶であれば認識失敗として終了して[2]，ペットボトルであれば m^{int} を 1.0 になるまで 0.1 だけ増加させながら手順 (3)〜(5) を繰り返し，それでも成功しなければ認識失敗として終了

実際の認識例を図 4.13 に示す．この例は，ユーザが左のペットボトルを取って欲しいと依頼した場合である．システムはまず内部照明の影響が無いものと見て認識を試みるが，実際には物体が冷蔵庫の光源の直前に存在して，内部照明の影響を大きく受けるため，認識に失敗している (図 4.13(b))．そこで，システムは徐々に内部照明重み m^{int} を増加させていき，$m^{int} = 0.7$ となったところで候補領域の抽出に成功し (図 4.13(c))，最終的に認識に成功する (図 4.13(d)).

4.5　実験

ここでは，学習した色変換パラメタを用いた物体の抽出実験と識別実験について述べる

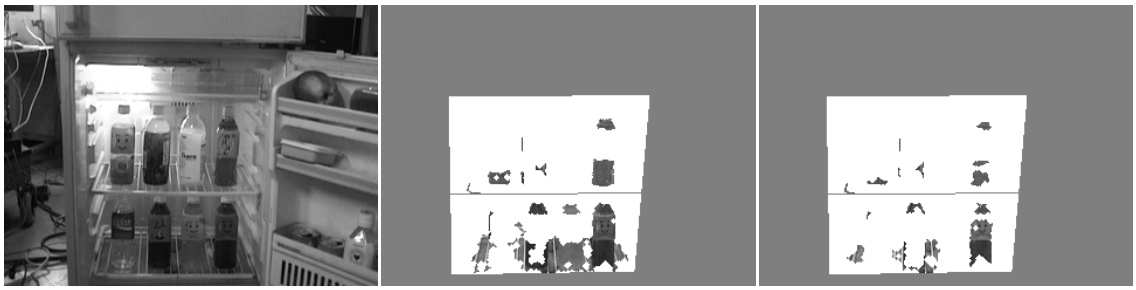
[2]缶は内部照明の影響が無いため

4.5.1　物体の検出実験

物体の検出時には色変換誤差を考慮して候補領域を抽出する．その際，色変換誤差が大きいと，対象物体の領域と背景や他の物体の領域が結合して抽出 (以下オーバーラップと呼ぶ) されてしまい，物体の検出に失敗することがある．このような場合には，ユーザとの対話によって認識を補助してもらうが，可能な限りこのような手間は省くことが望ましい．そこで，検出対象の物体の隣接領域が背景の場合と他の物体の場合とで，領域のオーバーラップ率 $r_{overlap} = n_{overlap}/n$ について検討する．ここで，n は候補領域抽出の実験回数，$n_{overlap}$ はその中で候補領域がオーバーラップした回数である．以下では，このオーバーラップ率 $r_{overlap}$ が低い程，物体の検出性能が高いものと判断する．

物体と背景がオーバーラップする場合

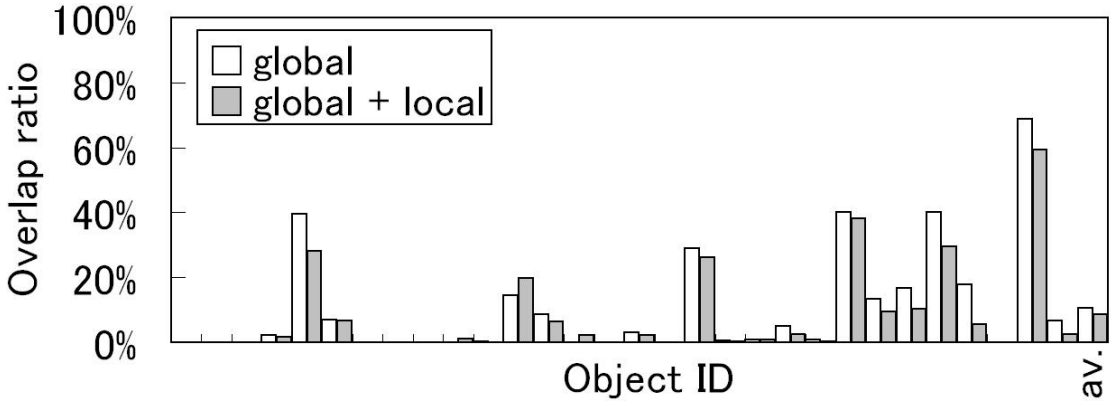
物体の色が橙や黄のように冷蔵庫内の光源の色に近いと，図 4.14(b) のように，物体と背景色の領域がオーバーラップすることがある．しかし，本手法の色変換を用いると色変換誤差が小さくなるので，抽出すべき色の範囲をより小さく見積もることができ，図 4.14(c) のように，背景から分離してうまく抽出することができる．



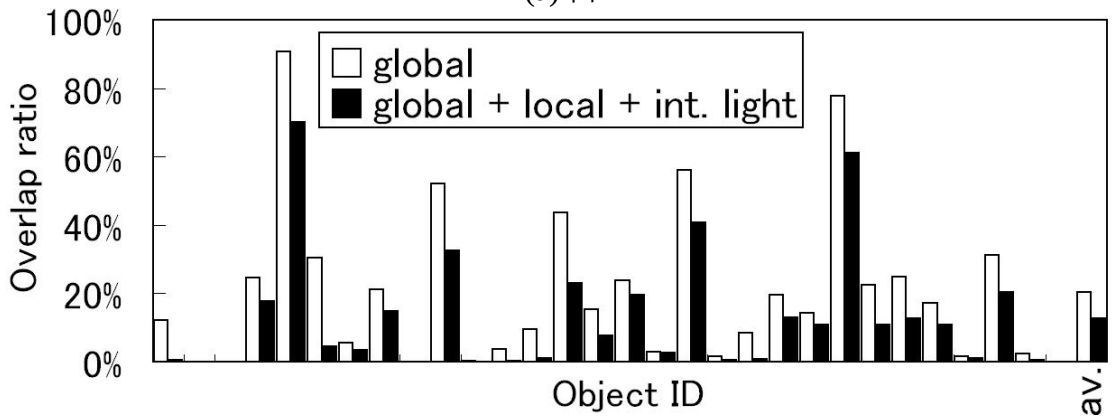
(a) 原画像　(b) 候補領域 (Global)　(c) 候補領域 (Global + Local + Int. light)

図 4.14: 物体と背景のオーバーラップ例

実験では，缶・ペットボトルについて各々30 種類の物体について，物体と背景のオーバーラップ率について調べ，その結果を図 4.15 に示す．缶については，内部照明の影響がほとんど無いため，一様な外部照明を仮定していた従来法でも，それ程オーバーラップ率は高くないが，外部照明照度の場所による変化を考慮した提案手法によって若干オーバーラップ率が下がっている．これに対して，ペットボトルについては，内部照明の影響を大きく受けるため，従来法では色変換誤差が非常に大きくなり，オーバーラップ率が非常に高くなるが，提案手法では内部照明の影響も考慮しているため，オーバーラップ率が低く抑えられている．



(a) 缶



(b) ペットボトル

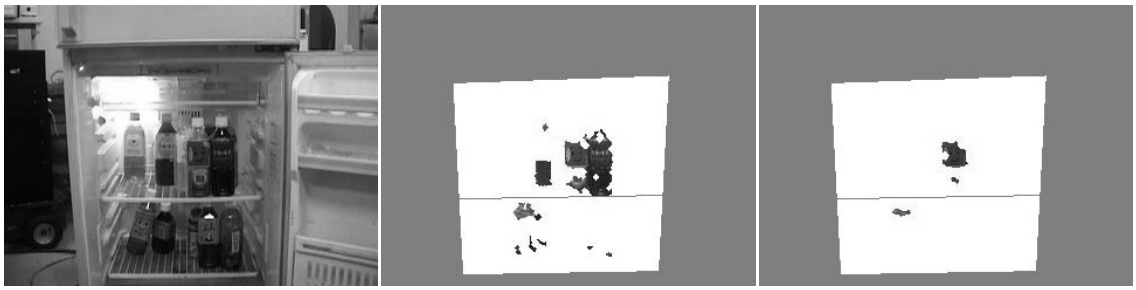
図 4.15: 物体と背景のオーバーラップ率

物体同士がオーバーラップする場合

背景とのオーバーラップの場合と同様，比較的色の似た物体が隣接すると，従来手法では図 4.16(b) のように，物体の領域がオーバーラップすることが多い．しかし，提案手法では図 4.16(c) のように，物体同士を分離して抽出する可能性が高くなる．実験では先と同様に，缶・ペットボトルについて各々30 種類の物体について，物体同士のオーバーラップ率について調べ，ペットボトルについての結果を図 4.17 に示す．同図では結果の対称性から左上半分の表示を省略してある．ここでもやはり内部照明の影響を考慮した提案手法においてオーバーラップ率が低くなっていることが分かる．

4.5.2 物体の識別実験

ここでは，物体の領域が検出できたとして，その領域がどの物体であるかを識別する実験を行う．実験に用いる観測値として，160 枚サンプル画像から缶・ペットボトルのそれぞれ 640 組抽出しておき，それを学習セットとテストセットに分



(a) 原画像　(b) 候補領域 (Global)　(c) 候補領域 (Global + Local + Int. light)

図 4.16: 物体同士のオーバーラップ例

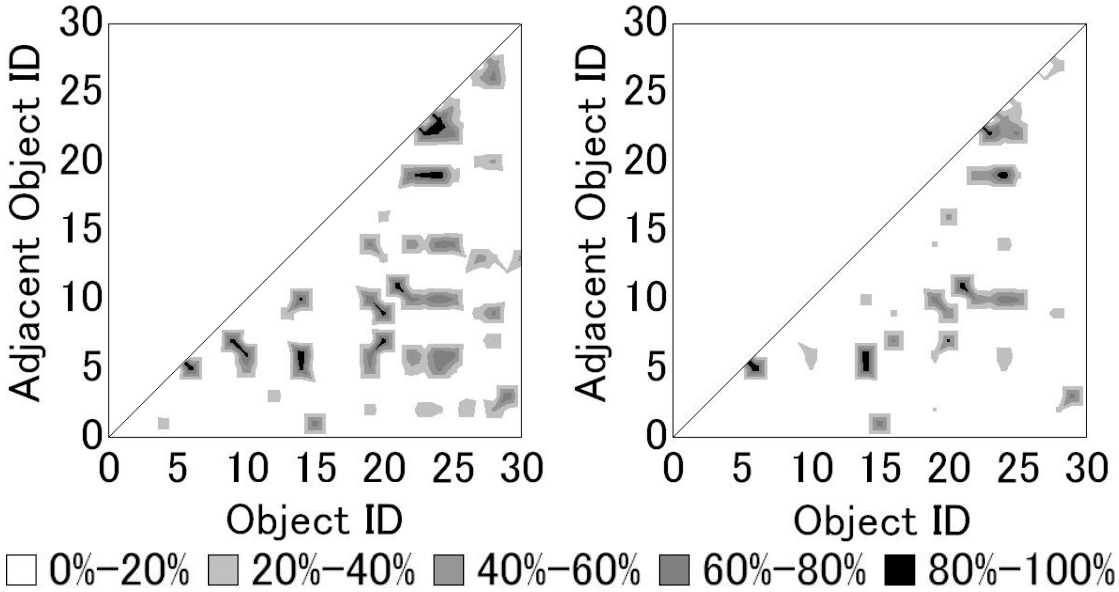
けて，学習セットによって色変換パラメタを学習しておく．そして，学習した色変換パラメタを用いて，学習セットとテストセットの各々に対して行った識別実験の結果を図 4.18 に示す．図 4.18 の横軸は，正しい物体がその順位までに入っているかどうかを示している．この結果より，識別率に関しては，従来法・提案手法による差が余り無いことが分かる．これは，学習セットの中に内部照明の影響を受けた観測値が存在し，従来手法でもそれを学習するため，内部照明方向への変化に鈍感になるために，内部照明の影響を受けた物体と受けていない物体共に同じ物体として見なすことができたためと考えられる．

4.5.3　総合評価

結局，物体の識別率については従来手法・提案手法共に余り変わらないものの，物体の検出率については提案手法が従来手法よりも上回るため，物体の検出・識別を含めた全体の認識率については，提案手法の方が優れていることが分かる．特に，内部照明の影響を受けるペットボトルについては，この傾向が顕著になっている．

4.6　まとめ

本章では，第 3 章で述べた手法を拡張して，外部照明と内部照明の 2 種類の光源がある冷蔵庫内のシーンを対象として，シーン内での照度変化を考慮した色変換の学習方法について提案した．外部照明照度の変化については，主に光源と物体との位置関係の影響が強いため，冷蔵庫の場所との依存関係を学習した．内部照明照度については，物体自身の光の透過率や内部照明を遮る物体の有無によって変化するため，観測データ毎に適切な値を算出して，物体認識時には，徐々に内部照明重みを変化させることで認識を試みた．これにより，従来のシーン内で一様な照度を仮定していたときと比較して，より精度の高い色変換が推定できた

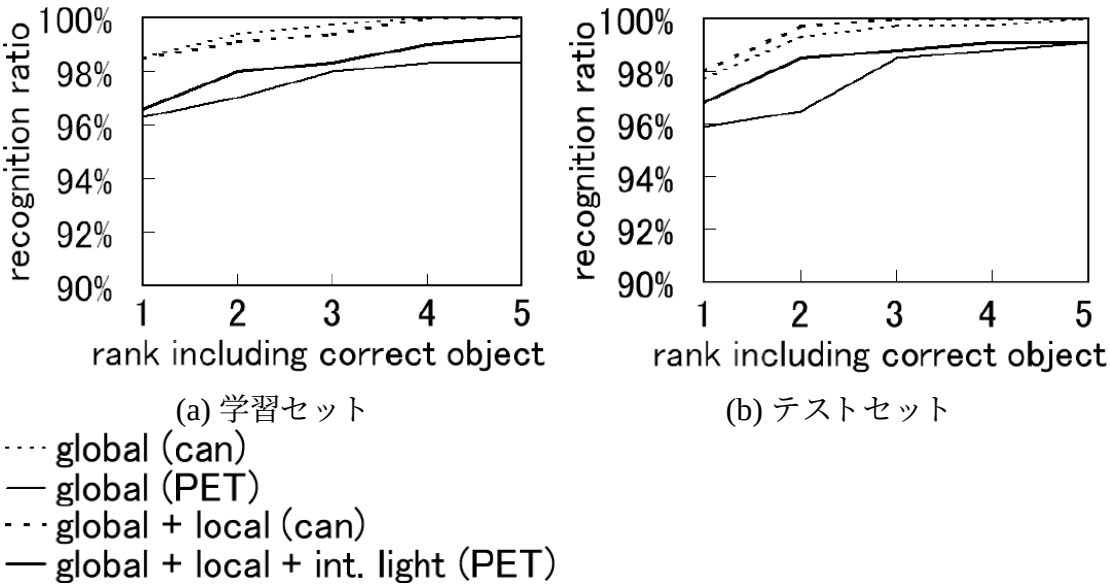


(a) Global (av. 9.9%)

(b) Global + Local + Int. light (av. 5.2%)

図 4.17: 物体同士のオーバーラップ率 (ペットボトル)

ことを実験により示した．また，実験により，拡張前のグローバルな外部照明変化を考慮した場合と比較して色変換誤差が小さくなり，物体検出時に背景や他の物体とのオーバーラップが抑制され，物体認識全体としての性能が上がることを確認した．



(a) 学習セット　　(b) テストセット

図 4.18: 物体の識別率

第5章　結論

本論文では，サービスロボットにとって必要なユーザとの対話を利用した物体認識とその際のユーザの対話の負担を可能な限り減らすための照明変化への対応手法に述べた．物体認識時には画像上の物体の色，画像上の物体の大きさ，画像上の位置といった変数が物体の向きや照明条件の影響を受けてばらつきを持つ．そこで，これらのばらつきを考慮して，ユーザに必要以上の対話の手間をかけさせないために，可能な限り自動的に物体認識を行うことを目的とした．

物体モデル作成時には，物体認識時の物体の向きの変化に対応して，かつ簡単に物体を認識できるように，投影画像を用いた方向毎の特徴抽出と他の物体との区別の付く必要最低限の特徴登録 (2 次特徴の階層的追加) を行った．

ここで，画像上の物体の色や隠蔽による画像上での物体の大きさの変化が生じると，自動認識に失敗することがある．このような場合に対して，現在の認識状況とユーザからの返答に応じて，各々適したリカバリー処理を考案した．ここでは，隠蔽された物体の認識，複数物体からの選択や色の範囲を拡大した再抽出，ユーザの訂正による物体モデルの学習等を扱い，ユーザとの対話が物体認識に対する有効な補助となることを確かめた．

また，多数の候補を提示する際に，領域に対するラベル付けや階層的な表示を導入することでユーザにとって選択しやすい認識結果表示を行い，その有効性や問題点を数名のユーザに対する主観テストにより確認した．

一方，照明条件変化は画像上の物体の色を大きく変化させ，自動物体認識を困難にする大きな要因となる．そこで，本論文ではこのような照明変化へ対応するための，色変換の学習手法について提案した．最初に，グローバルな照明を考慮した場合に，一つの参照物体の色を観測することで物体認識時の色変換を推定する手法を定式化した．加えて，本システムを新しい環境に適応した場合に，ユーザとの対話を利用しながら認識を成功させることで色変換の推定精度を徐々に高めていくことで，ユーザの対話の負担を減らす学習方法について述べた．

次にこれを拡張して，外部照明と内部照明の 2 種類の光源がある冷蔵庫内のシーンを対象として，シーン内でのローカルな照度変化を考慮した場合に拡張した．これにより，様々な照明条件下での精度の高い色変換推定を実現して，実験により自動物体認識の性能が上がることを確認した．

本論文で提案した照明変化への対応手法は，冷蔵庫環境に限らず，参照色として利用できるものが一つでもあるシーン，例えば，枠の部分を参照色とした食器棚のシーンやホワイトボードを参照色とした会議室のシーンに対して応用すること

が可能と考えられる．また，今回提案した対話を用いた物体認識の枠組みは，ユーザと対話的に画像検索するシステム等，ユーザが側に居ながら使うようなシステムへの応用が可能と考えられる．

最後に，今後の課題としては以下のようなもの残っている．

- 候補選択にかかる時間を最小化する認識結果表示法

  2.5 節で述べた階層表示の場合，一回あたりについて見ると，選択肢が減り選択時間が短縮されるものの，階層化による選択回数の増加という欠点がある．よって，候補選択時間にかかる時間を最小化するためには，これらのトレードオフの問題を取り扱う必要がある．その際には，実際にユーザが音声で答えることを想定して，音声認識率も考慮に入れて考える必要がある．例えば，逐次確認のように「はい」「いいえ」のみで答えるような選択の場合は音声認識も容易であるため高い確率で選択が成功する．これに対して，色名で答える場合には 2.5.1 節でも述べたように，ユーザの想定外の発話が生じる可能性があり，その結果，選択の成功率が下がる．最終的には，様々な認識結果表示法の候補を作成して，候補毎に音声認識率も考慮した場合の選択時間の期待値を計算して，それを最小にするような認識結果表示法を採用することが考えられる．

- 認識対象物体の拡張

  現在は，認識対象として缶・瓶・ペットボトル・果物について取り扱っているが，実際の冷蔵庫にはタッパウェアや袋等の多くの種類の物体が入っている．そのような物体については，ユーザが取って欲しい場合では無くても，缶やペットボトル等の取って欲しい物体を隠蔽していたり，またそれらをマニピュレータで掴む際に邪魔になったりするような場合には，一旦退けるなどの措置が必要である．このようなことを行うためには，結局，物体の領域を認識する必要があるため，タッパウェアや袋への認識対象物体の拡張が課題となる．

- 音声対話システム・移動マニピュレータとの統合

  本論文ではサービスロボットにおける画像処理による物体認識手法について述べたが，実際に音声対話システムや移動マニピュレータと統合する際には，いくつかの課題がある．音声対話システムとの関係としては，現在，物体認識の失敗を音声対話によって補うことを行っているが，逆に音声認識に失敗した場合に，画像処理によって得られている認識状況を利用して，音声認識の失敗を成功に導くことが考えられる．また，移動マニピュレータとの関係としては，ユーザから持ってきて欲しいと頼まれた物体が，他の物体に隠蔽されているような状況で，現在のマニピュレータと物体の位置関係がどのようになっているかを画像処理によって理解して，マニピュレータの軌道計画の補助とすること等が挙げられる．

関連図書

[1] 小山猛, 山藤和男, 田中孝之, ”介護用装着型ヒューマン・アシスト装置に関する研究 (第 1 報，コンセプト，システム設計と実機の開発)”, 日本機械学会論文集 C 編, Vol. 66, No. 651, pp. 155-160, 2000.

[2] 竹内郁雄，柄川索, ”歩行支援機器の開発”, 日本ロボット学会誌, Vol. 21, No. 4, pp. 51-55, May 2003.

[3] 石井純夫，”食事支援ロボット「マイスプーン」”, 日本ロボット学会誌, Vol. 21, No. 4, pp. 44-47, May 2003.

[4] Y. Takahashi，T. Komeda, T. Uchida，M. Miyagi, and H. Koyama, ”Development of the mobile robot system to aid the daily life for physically handicapped”, Proc. of Int. Conf. on Machine Automation 2000, pp. 549-554, Osaka, Japan, Sep. 2000.

[5] 村上佳史，久野義徳，島田伸敬，白井良明, ”知的車椅子のための歩行者の観察に基づく衝突回避”, 日本ロボット学会誌, Vol. 20, No. 2, pp. 206-213, 2002.

[6] 石松隆和，茂地徹，扇谷保彦，杉山和一, ”斜面市街地の移動手段の開発”, 日本ロボット学会誌, Vol. 21, No. 4, pp. 40-43, May 2003.

[7] 谷腰欣司, ”磁気センサとその使い方-ホール素子を使いこなすための実用技術-”, 日刊工業新聞社, 1986.

[8] ”RF タグの市場動向”, シーエムシー出版, 2004.

[9] 奥野忠一，久米均，芳賀敏郎，吉澤正, ”多変量解析法”, 日科技連出版社, 1981.

[10] P.V.C. Hough, ”Method and means for recognizing complex patterns”, U.S. Patent No. 3069654, 1962.

[11] R.O. Duda and P.E. Hart, ”Use of the Hough transformation to detect lines and curves in pictures”, Comm. ACM, Vol. 15, No. 1, pp. 11-15, 1972.

[12] 谷内田正彦, ”多重画像を利用したシーンの理解”, 情報処理学会誌, Vol. 24, No. 12, p. 6, 1983.

[13] M. Yachida, M. Asada, and S. Tsuji, ”Automatic Analysis of Moving Images”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-3, No. 1, pp. 12-20, 1981.

[14] B.K.P. Horn and B.G. Schunck, ”Determining Optical Flow”, Artificial Intelligence, Vol. 17, No. 9, pp. 55-60, 1987.

[15] W.E.L. Grimson, ”From Images to Surfaces”, MIT Press, Cambridge, 1981.

[16] Y. Shirai and M. Suwa, ”Recognition of polyhedrons with a range finder”, Proc. of 2nd Int. Joint Conf. on Artificial Intelligence, pp. 80-87.

[17] G.W. Zhao and S. Yuta, ”Obstacle Detection by Vision System for an Autonomous Vehicle”, Proc. of Intelligent Vehicles ’93 Symposium, pp. 31-36, 1993.

[18] 熊野仁子, 大矢晃久, 油田信一, ”障害物回避のための移動ロボット用ステレオ視覚センサの開発”, 第 17 回日本ロボット学会学術講演会予稿集, pp.797-798, 1999.

[19] 文仁赫, 三浦純, 白井良明, ”不確かさを考慮した観測位置と移動のオンライン計画手法”, 日本ロボット学会誌, Vol. 17, No. 8, pp. 1107-1113, 1999.

[20] 安江, 白井, ”物体認識のための両眼立体視”, 電子技術総合研究所彙報, Vol. 37, pp. 1101-1119, 1973.

[21] 角保志, 富田文明, ”ステレオビジョンによる 3 次元物体の認識”, 電子情報通信学会論文誌, Vol. J80-D-II, No. 5, pp. 1105-1112, 1997.

[22] 佐川立昌, 岡田慧, 加賀美聡, 稲葉雅幸, 井上博允, ”複数の距離画像からの表面形状モデリングとその階層的なマッチングによる物体認識”, ロボティクス・メカトロニクス講演会予稿集, 1P1-68-100, 2000.

[23] 株式会社ビュープラス, ”Bumblebee”, http://www.viewplus.co.jp/products/bumblebee/bumblebee_main.htm

[24] 株式会社ビュープラス, ”Digiclops”, http://www.viewplus.co.jp/products/digiclops/digiclops_main.html

[25] R.O. Duda, D. Nitzan, and P. Barrett, ”Use of range and reflectance data to find planar surface regions”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 1. No. 3, pp. 259-271, 1979.

[26] D. Zuk and M. Dellave, ”3D sensor for the autonomous land vehicle”, Operation Manual, ERIM, 1985.

[27] M. Rioux, F. Blais, J.A. Beraldin, and P. Boulanger, ”Range image sensors development at NRC laboratories”, Proc. of the Workshop on Interpretation of 3D scenes, pp. 27-29, Austin, TX, Nov. 1989.

[28] C. Thorpe ed., ”Vision and Navigation”, The Carnegie-Mellon Narvlab, Kluwer Academic Publishers, 1990.

[29] 安藤吉伸, 坪内孝司, 油田信一, ”レーザレンジセンサを用いた移動ロボットの壁沿い走行”, ロボティクス・メカトロニクス講演会’98 講演論文集, 1AII1-12, 1998.

[30] Y. Adachi, H. Saito, Y. Matsumoto, and T. Ogasawara, ”Memory-Based Navigation using Data Sequence of Laser Range Finder”, Proc. of 10th IEEE Int. Workshop on Symposium on Computational Intelligence in Robotics and Automation, 2003.

[31] 朝倉弘崇, 山下淳, 金子透, ”レーザレンジファインダを搭載した移動ロボットによる 3 次元形状計測”, 映像情報メディア学会技術報告, Vol. 28, No. 45, pp. 25-26, 2004.

[32] N. Yokoya and M.D. Levine, ”Range Image Segmentation Based on Differential Geometry: A Hybrid Approach”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-11, No. 6, pp. 643-649, 1989.

[33] K. Ikeuchi and B.K.P. Horn, ”Numerical shape from shading and occluding boundaries”, Artificial Intelligence, Vol. 17, No. 1-3, pp. 17-45, 1981.

[34] J. Kender, ”Shape from texture: An aggregation transformation that maps a class of textures into surface orientation”, Proc. of Computer Vision and Pattern Recognition’86, p. 578, 1986.

[35] Y. Shirai, ”Three-Dimensional Computer Vision”, Springer-Verlag, pp. 62-68, 1987.

[36] A. Okamoto, Y. Shirai, and M. Asada, ”Integration of Color and Range Data for Three-Dimensional Scene Description”, IEICE Trans. Inf. And Syst., Vol. E76-D, No. 4, pp. 501-506, 1993.

[37] 日本色彩学会，”カラーハンドブック”，東京大学出版, 1989.

[38] 浅田尚紀，”カメラキャリブレーション”，コンピュータビジョン技術評論と将来展望, pp. 37-53, 新技術コミュニケーションズ, 1998.

[39] R.Y. Tsai, ”An efficient and accurate camera calibration technique for 3D machine vision”, Proc. of Computer Vision and Pattern Recognition’86, pp. 364-374, 1986.

[40] A. Rosenfeld and A.C. Kak, ”Digital Picture Processing”, Academic Press, New York, 1976.

[41] G. Nagy and Jie Zou, ”Interactive Visual Pattern Recognition”, Proc. of 16th Int. Conf. on Pattern Recognition, Vol. 2, pp. 478-481, 2002.

[42] Y.P. Hung and Y.P. Tsai, ”Trail-Dependent Intelligent Scissors Based on Multi-scale Image Segmentation”, Proc. of 5th Asian Conf. on Computer Vision, Vol. 2, pp. 539-544, 2002.

[43] 藤井憲作，杉山和弘, ”歩行者ナビゲーション支援のための場所案内文生成手法”, 電子情報通信学会論文誌, Vol. J82-DII, No. 11, pp. 2026-2034, 1999.

[44] 岩田満，鬼沢武久，”絵のつながりを考慮した絵情報の言語的表現”, 電子情報通信学会論文誌, Vol. J84-D-II, No. 2, pp. 227-250, 2001.

[45] 渡辺靖彦，長尾真，岡田至弘, ”画像の内容を説明するテキストを利用した画像解析”, 画像の認識・理解シンポジウム (MIRU’96) 論文集, Vol. 2, pp. 271-276, 1996.

[46] S. Wachsmuth and G. Sagerer, ”Connecting Concepts from Vision and Speech Processing”, Workshop on Integration of Speech and Image Understanding, 1999.

[47] U. Ahlrichs, J. Fischer, J. Denzler, C. Drexler, H. Niemann, E. Noth and D. Paulus, ”Knowledge Based Image and Speech Analysis for Service Robots”, Workshop on Integration of Speech and Image Understanding, 1999.

[48] 高橋拓弥，中西周作，久野義徳，白井良明, ”音声とジェスチャによる対話に基づくヒューマンロボットインターフェース”, インタラクション’98 論文集, pp. 161-168, 1998.

[49] S. Zeki, ”A vision of the brain”, Blackwell Scientific Publications, London, 1993.

[50] B.A. Wandell, ”Foundation of Vision”, Sinauer Associates, Sunderland, MA, 1995.

[51] S. Nadimi and B. Bhanu, ”Moving Shadow Detection Using a Physics-based Approach”, Proc. of 16th Int. Conf. on Pattern Recognition, Vol. 2, pp. 701-706, 2002.

[52] G. Buchsbaum, ”A spatial processor model for object color perception”, J. Franklin Inst. Vol. 310, pp. 1-26, 1980.

[53] J.J. McCann, J.A. Hall, and E.H. Land, ”Color mondrian experiments: The study of average spectral distributions”, J. Opt. Soc. Am., Vol. A-67, p.1380, 1977.

[54] D.H. Marimont and B.A. Wandell, ”Linear models of surface and illuminant spectra”, J. Opt. Soc. Am., Vol. A-9, pp. 1905-1913, 1992.

[55] M.S. Drew, J. Wei, and Z.N. Li, ”Illumination-invariant image retrieval and video segmentation”, Pattern Recognition, Vol. 32, No. 8, pp. 1369-1388, 1999.

[56] M.S. Drew, Z.N. Li, and Z. Tauber, ”Illumination color covariant locale-based visual object retrieval”, Pattern Recognition, Vol. 35, No. 8, pp. 1687-1704, 2002.

[57] E.G. Miller and K. Tieu, ”Color Eigenflows: Statistical Modeling of Joint Color Changes”, Proc. of 8th Int. Conf. on Computer Vision, Vol. 1, pp. 607-614, Vancouver, Canada, Jul. 2001.

[58] D. Forsyth, ”A novel algorithm for color constancy”, Int. J. Computer Vision, Vol. 5, pp. 5-36, 1990.

[59] G.D. Finlayson, ”Color constancy in diagonal chromaticity space”, Proc. of 5th Int. Conf. on Computer Vision, pp. 218-223, 1995.

[60] E.H. Land, ”Retinex Theory of Color Vision”, Science Am., pp. 108-129, 1997.

[61] G.D. Finlayson, B.V. Funt, and K. Barnard, ”Color constancy under varying illumination”, Proc. of 5th Int. Conf. on Computer Vision, pp. 720-725, 1995.

[62] K. Barnard, G. finlayson, and B. Funt, ”Color Constancy for Scene with Varying Illumination”, Computer Vision and Image Understanding, Vol. 65, No. 2, pp. 311-312, Feb. 1997.

[63] J.J. McCann, J.A. Hall, and E.H. Land, ”Color mondrian experiments: The study of average spectral distributions”, J. Opt. Soc. Am., Vol. A-67, p. 1380, 1977.

[64] S. A. Shafer, ”Using color to separate reflection components”, Color Research and Application, Vol. 10, pp. 210-218, 1985.

[65] 松山隆司，浅田尚紀，”多重画像の統合による多機能高精度画像計測”，コンピュータビジョン技術評論と将来展望, pp. 23-36, 新技術コミュニケーションズ, 1998.

[66] D.B. Judd, D.L. Macadam, and G.W. Wyszecki, ”Spectral distribution of typical daylight as a function of correlated color temperature”, J. Opt. Soc. Am., Vol. 54, pp. 1031-1040, 1964.

[67] J. Cohen, ”Dependency of the spectral reflectance curves of the Munsell color chips”, Psychonomical Science, Vol. 1, pp. 369-370, 1964.

[68] L.T. Malony and B.A. Wandell, ”Color constancy: A method for recovering surface spectral reflectances”, J. Opt. Soc. Am., Vol. A-3, pp. 29-33, 1986.

[69] S. Nakauchi, P. Silfsten, J. Parkkinen, and S. Usui, ”Computational theory for color transparency: recover of spectral information for overlapping surfaces”, J. Opt. Soc. Am., Vol. A-16, No. 11, pp. 2612-2624, 1999.

[70] 滝澤正夫，槇原靖，白井良明，島田伸敬，三浦純, ”サービスロボットのための対話システム”, システム制御情報学会論文誌, Vol. 16, No. 4, pp. 174-182, 2002.

[71] 濱田純, 槇原靖, 矢野恵生, 滝澤正夫, 白井良明, 三浦純, ”必要に応じてユーザと対話しながら行動するサービスロボット”, 第 21 回日本ロボット学会学術講演会, CD-ROM, 2003.

[72] 矢野恵生，槇原靖，滝澤正夫，白井良明，三浦純，島田伸敬，”福祉用サービスロボットの開発”，第 2 回福祉工学シンポジウム講演会講演論文集, 2002.

付 録A　冷蔵庫領域の抽出

冷蔵庫内のシーンの物体認識をする際に，あらかじめ冷蔵庫の領域を抽出しておくことは，以下の 4 つの点で有効である．

1. 探索範囲が画像の全体から冷蔵庫内のみになるために，処理時間が短縮されると共に冷蔵庫外の物体の誤検出を防止できる．

2. ユーザが冷蔵庫に対する物体の位置を指定できる．例えば，ユーザに認識した物体を報告するときに，上下の棚のいずれに存在するかを伝えることが可能であるし，また，指定した物体が見つからなかったときに，ユーザが冷蔵庫内の場所を指定することによって，その場所に絞って再探索することも可能である．

3. 物体モデルには実際の物体の大きさが登録されているため，そのままでは画像上の物体の大きさとして使うことができないが，画像上で映る冷蔵庫の大きさから，画像上での物体の大きさを推定できる．

4. 冷蔵庫領域の抽出によって，色変換に必要な参照色のための扉等の領域を自動的に得ることができる．

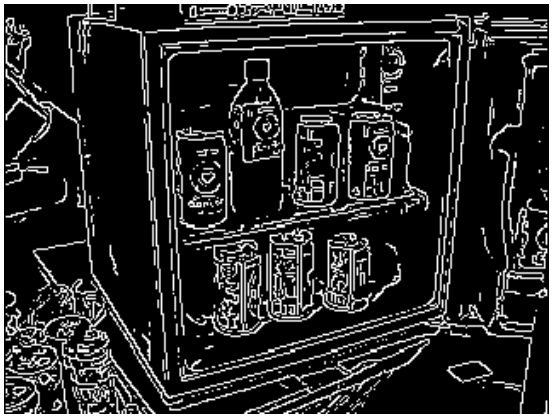
逆に上記の 4 番目の点で利用するためには，様々な照明条件に対して色変換を行う前に冷蔵庫領域を抽出する必要がある．よって，冷蔵庫領域の抽出に輝度情報や色情報を用いることはできない．よって，冷蔵庫領域の抽出には，照明条件の変化にあまり影響されないエッジ画像によるハフ変換 [10][11] を用いる．

まずエッジ画像を求めなければならないが，エッジ強度は RGB 空間のユークリッドノルムのカラーのソーベルオペレータによるエッジを用いて，その強度が 10 以上のものに対して非極大値抑制を行うことによってエッジ画像を取得する．原画像 A.1(a) に対するエッジ画像は図 A.1(b) のようになる．このエッジ画像をハフ変換して水平・垂直に近いものを取り出すと図 A.1(c) の直線が出てくる．ここで，投票数の閾値は 75 とし，水平・垂直については，直線の傾きが [deg] 単位でそれぞれ [-30,30],[60,120] に入るものを選択した．次に，これらの直線から，画像中心から左右に向かっていって，初めて当たる垂直な線を冷蔵庫の左右の枠として採用し，画像の上下それぞれ 1/3 の部分から最も中心に近い水平な線を冷蔵庫の上下の棚として採用する．また，真中の棚の線は上下の棚の線の中間付近に存在する線

を採用する．これにより，図 A.1(c) の太線が冷蔵庫の枠や棚の直線として得られ，最終的には図 A.1(d) のような冷蔵庫領域を抽出することができる．

また，冷蔵庫の扉の方についても同様の手法を用いて領域を抽出することができるが，ここでは詳細の説明を省く．

(a) 原画像



(b) エッジ画像

(c) ハフ変換による直線の抽出

(d) 冷蔵庫領域の抽出結果

図 A.1: ハフ変換による冷蔵庫領域の抽出

付 録B　色に基づく領域分割

領域分割法には色ヒストグラムによる判別基準法を用いるもの [35] や色ヒストグラムと距離情報統合した領域分割 [36] 等があるが，ここでは単眼カメラによる領域分割を行うため，色ヒストグラムによる領域分裂法を用いる．ここでは，色ヒストグラムとして YIQ 成分のヒストグラムを用いるが，これは輝度成分と色あり成分を分けて考えるためである．以下にここで用いる領域分裂法の手順を示す.

1. 冷蔵庫内の 1 つの領域に対して YIQ 成分のヒストグラムを作成する.

2. 各成分のヒストグラムに対して，分割に最も適した谷とその評価値を計算する.
   ここで評価値の計算方法について説明する．今，ある領域のある成分のヒストグラムが図 B.1 のように得られたとする．まず，谷の候補となる点として，ヒストグラムの極小点の位置 $v(i)$ とそこでの高さ $H_v(i)$ を取得し，山の候補となる点として，極大値の位置 $r(i)$ とそこでの高さ $H_r(i)$ を取得する．次に，ある谷 $v(i)$ に対する隣り合う山との高さの差 $H_d(i)$ を，

   $$H_d(i) = \min\{H_r(i), H_r(i+1)\} - H_v(i) \tag{B.1}$$

   として定義しておき，全ての谷に対して $H_d(i)$ を計算して，それが 10 以下になる谷に対しては，その谷と隣り合う両側の山の低い方を候補から外す．例えば図 B.1 の破線の丸で囲んだ山と谷がそれに当たる．残った谷に対して，谷の高さ $H_v(i)$ がどれだけ低いか，隣り合う山との高さの差 $H_d(i)$ がどれだけ深いか，隣り合う山の間隔 $W(i) = r(i+1) - r(i)$ がどれだけ離れているかを考慮した評価式 (B.2) の値が最大のものを最適な谷として採用する.

   $$R(i) = w_0 \frac{H_v(i)}{H_{max}} + w_1 \frac{H_d(i)}{H_{max}} + w_2 \frac{W(i)}{W_{max}} \tag{B.2}$$

   ここで，w_0, w_1, w_2 は実験的に決めた適当な重み係数でそれぞれ $w_0 = 0.25, w_1 = 0.5, w_2 = 0.25$，$H_{max}$ は最頻値での度数，W_{max} はヒストグラムの幅 (YIQ の値をそれぞれ [0,255] に入るように変換しているので，ここでは 255 とする) である.

3. 最も良い評価値を持つ成分を選択して，評価値が閾値 (=0.4) を超えていれば，その成分の谷で領域を分割してラベリングする.

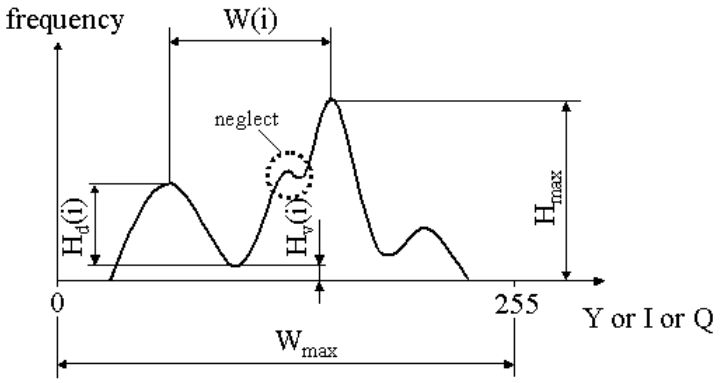


図 B.1: ヒストグラムの例

4. 分割が行われれば分割された領域の一つを選んで 1 に戻り，そうでなければ次の領域に移り 1 に戻る．

5. 以上の操作を分割する領域が無くなるまで繰り返す．

今回の方法による領域分割画像の例を図 B.2 に示す．この図は，各領域を平均色で表した図である．

(a) 原画像　　(b) 領域分割画像

図 B.2: 領域分裂法による領域分割画像

付 録C　物体の3次元位置の取得

画像上で認識した物体をマニピュレータで掴むためには，その 3 次元位置を取得する必要がある．ここでは，マニピュレータの座標系 (鉛直方向上向きを z 座標とする) を世界座標系として，マニピュレータから見た把持すべき物体の 3 次元位置を求めることを考える．序論でも述べたように，本研究ではあらかじめキャリブレーションされたカメラと冷蔵庫の棚の高さを用いることで，物体の 3 次元位置を求めるものとする．

まず，カメラキャリブレーションについて簡単に述べておく．画像座標系 (x_f, y_f) から世界座標系 (x_w, y_w, z_w) への変換を行う際には，以下のように途中にカメラ座標系 (x_c, y_c, z_c) を間に入れて考えることが多い．

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \frac{z_c}{f} \begin{bmatrix} x_f - c_x \\ y_f - c_y \end{bmatrix} \tag{C.1}$$

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T \tag{C.2}$$

ここで，f は画素値に換算した焦点距離，c_x, c_y は画像中心の x, y 座標，R, T は世界座標からカメラ座標系へ変換するための回転行列と並進ベクトルであり，これらをカメラキャリブレーションによってあらかじめ求めておく必要がある．そのためには，世界座標系と画像座標系での複数の対応点を取得して，Tsai の方法 [39] によりこれらのパラメタを算出する．具体的な対応点取得は，ある姿勢にあるマニピュレータの手先の 3 次元位置を各関節角より算出して，画像上でのその手先の 2 次元位置を手動で抽出することで行う．また，この操作をマニピュレータが作業を行う範囲内の様々な点に対して行うことで，パラメタ算出に必要な対応点を取得する．

次に，上で挙げた関係式を基にして，図 C.1 のように画像上で認識した物体の画像上での接地点 (x_{fb}, y_{fb}) と世界座標上での接地点 (x_{wb}, y_{wb}, z_{wb}) を，その物体が接地している冷蔵庫の棚の高さを

$$z_w = z_{shelf} \tag{C.3}$$

として与えることで，算出する[1]．

[1]ここで，6 個の未知数 $x_c, y_c, z_c, x_{wb}, y_{wb}, z_{wb}$ に対して，式 (C.1),(C.2),(C.3) の 6 本の方程式が得られるため．

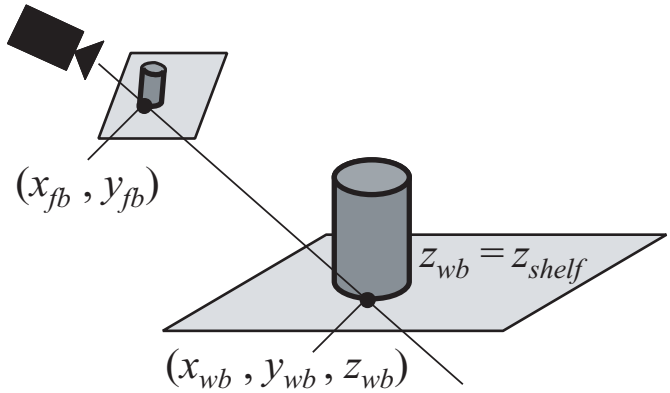


図 C.1: 3 次元位置取得の問題設定

ここまでで求まったものは，接地面の周上のある一点の 3 次元位置であるので，最後に物体の接地面の中心点を求めることを考える．ここでは，問題の簡単化のため，底面が円形になる物体 (四角型のペットボトルと果物以外) について扱う．また，表記の簡略化のため，図 C.2 のように $z_w = z_{shelf}$ に投影した平面 (x_w, y_w) 上で問題を扱う．まず，画像上における物体の接地線分 (物体領域の下端) の中点に対応する接地点を (x_{wb}, y_{wb}) として求める．次いでこの平面上に投影したレンズ中心の座標を式 (C.2) に $(x_c, y_c, z_c) = (0, 0, 0)$ を代入することで，(x_{wl}, y_{wl}) として求める．ここで，近似的にレンズ中心と接地点を結ぶ直線が接地面中心 (x_{wo}, y_{wo}) を通るものとすると，接地面中心は物体モデルの幅 w を用いて，以下のように表される．

$$x_{wo} = x_{wb} + \frac{w}{2}\cos\theta \tag{C.4}$$

$$y_{wo} = y_{wb} + \frac{w}{2}\sin\theta \tag{C.5}$$

$$\theta = \arctan\left(\frac{y_{wb} - y_{wl}}{x_{wb} - x_{wl}}\right) \tag{C.6}$$

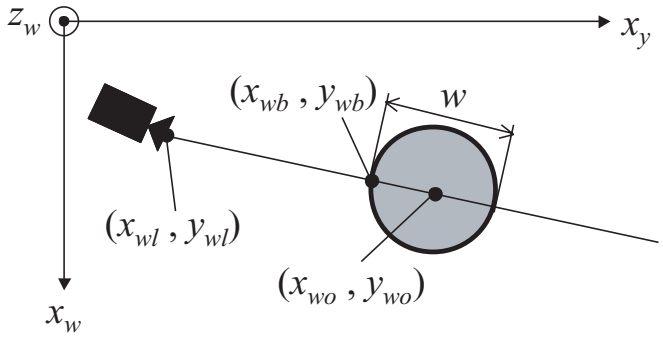


図 C.2: 接地面の中心点の算出

この接地面中心点と物体モデルをもとにして，物体の世界座標における範囲が求まる．例えば，円筒形近似できる缶のような物体の場合には，以下のような範囲 R_o になる．

$$R_o = \{(x_w, y_w, z_w) | (x_w - x_{wb})^2 + (y_w - y_{wb})^2 \leq \left(\frac{w}{2}\right)^2,\ z_{shelf} \leq z_w \leq z_{shelf} + h\} \tag{C.7}$$

本論文に関する論文および口頭発表 (投稿中も含む)

論文 (Full paper 査読)

- Y. Makihara, M. Takizawa, K. Ninokata, Y. Shirai, J. Miura, and N. Shimada, ”A Service Robot Acting by Occasional Dialog -Object Recognition Using Dialog with User and Sensor-Based Manipulation-”, J. of Robotics and Mechatronics, Vol. 14, No. 2, pp. 124-132, 2002.
- 槇原靖，滝澤正夫，白井良明，島田伸敬, ”対話を用いた物体認識のための照明変化への適応”, 電子情報通信学会論文誌, Vol. J87-D-II, No. 2, pp. 629-638, Feb. 2004.
- 槇原靖, 白井良明, 島田伸敬, ”対話を用いた物体認識のためのローカルな照度変化を考慮した色変換の学習”, 電子情報通信学会論文誌 (掲載予定).

国際会議 (Full paper 査読)

- Y. Makihara, M. Takizawa, Y. Shirai, J. Miura, and N. Shimada，”Object Recognition Supported by User Interaction for Service Robots”, Proc. of 5th Asian Conf. on Computer Vision, Vol. 2, pp. 719-724, Melbourne, Australia, Jan. 2002.
- Y. Makihara, M. Takizawa, Y. Shirai, J. Miura, and N. Shimada，”Object Recognition Supported by User Interaction for Service Robots”, Proc. of 16th Int. Conf. on Pattern Recognition, Vol. 3, pp. 561-564, Quebec, Canada, Aug. 2002.
- Y. Makihara, M. Takizawa, Y. Shirai, and N. Shimada, ”Object Recognition in Various Lighting Conditions”, Proc. of 13th Scandinavia Conf. on Image Analysis, pp. 899-906, Gothenberg, Sweden, Jun. 2003
- Y. Makihara, Y. Shirai, and N. Shimada，”Online Learning of Color Transformation for Interactive Object Recognition under Various Lighting Conditions”, Proc.

of 17th Int. Conf. on Pattern Recognition, Vol. 3, pp. 161-164, Cambridge, UK, Aug. 2004.

国内発表

- 槇原靖，滝澤正夫，二ノ方一生，白井良明，三浦純，島田伸敬, ”必要に応じてユーザと対話しながら行動するロボット-対話を利用した物体の認識と操作-”, ロボティックス・メカトロニクス講演会 2001 講演論文集, CD-ROM, 2001(ベストプレゼンテーション表彰受賞).
- 槇原靖，滝澤正夫，白井良明，三浦純，島田伸敬, ”ユーザとの対話を用いたサービスロボットのための物体登録”, 第 19 回日本ロボット学会学術講演会講演論文集, CD-ROM, 2001.
- 槇原靖，滝澤正夫，白井良明，三浦純，島田伸敬, ”ユーザとの対話を用いたサービスロボットのための物体認識”, 第 20 回日本ロボット学会学術講演会講演論文集, CD-ROM, 2002(研究奨励賞受賞).
- 槇原靖, 白井良明, 島田伸敬, ”対話を用いた物体認識のための認識結果表示法”, 電子情報通信学会 パターン認識・メディア理解 (PRMU) 研究会, Feb. 2005(掲載予定).

謝辞

本研究を進める上で，御多忙の中，熱心に御指導をして頂いた白井良明 教授，三浦純 助教授，島田伸敬 助教授，先山卓郎 助手に心より感謝致します.

特に，白井教授と島田助教授には学部 4 年生として研究室に来て以来 6 年もの間，毎週の研究ミーティングで御指導を頂き，研究の指針を示して頂いたことは，博士論文を完成させるに当たって非常に大きな助けとなりました.

事務官の岡田さんには，発表の際に本来自分で作成すべき出張手続きなどの事務処理をお任せすることが多く，大変迷惑をおかけしましたが，御陰様で安心して会議等に参加することができましたことを大変感謝致します.

また，両親には，博士後期課程に進むに当たって快く了承して頂き，学費を含め生活全般を支えてくれたことを感謝します.

最後になりましたが，日頃から多くのご支援，御助言を頂いた方々，並びに素晴らしい研究環境を提供して頂いた白井研究室の皆様に心より感謝致します.