

論 文

対話を用いた物体認識のための照明変化への適応

榎原 靖[†] 滝澤 正夫^{††} 白井 良明[†] 島田 伸敬[†]

Adaptation to Change of Lighting Conditions for Interactive Object Recognition

Yasushi MAKIHIRA[†], Masao TAKIZAWA^{††}, Yoshiaki SHIRAI[†],
and Nobutaka SHIMADA[†]

あらまし 本論文ではユーザとの対話を用いた物体認識システムに必要な機能の一部として、様々な照明条件下で物体認識を行うための色変換の学習方法について述べる。システムは最初に対象となる物体のモデルをある基準の照明条件下で作成しておく。物体認識時の照明条件は物体モデル作成時のものとは異なるので、システムは認識時の照明条件に合わせて物体モデルの色を変換する必要がある。この色変換の推定には認識時に観測される一つの参照物体の色を用いる。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。ただし、新しい環境下ではそのような色の組が蓄えられていないので、適当な関係を仮定して色変換する。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色の組を用いて関係を更新する。そして物体を認識することに色変換の推定精度を向上させることで、ユーザの対話の負担を軽減する。最後に、実際の冷蔵庫内のシーンをを用いた実験により、本手法の有効性を示す。

キーワード 色変換, 照明条件, 学習, 物体認識, 対話

1. ま え が き

近年、高齢化社会の到来により人を助けるサービスロボットの必要性が高まっている。高橋ら [1] は、画像を用いて認識した物体を持ってくるサービスロボットの研究を行った。しかし、サービスロボットの対象シーンは複雑で自動認識が困難な場合がある。このような場合に、視覚以外の情報を利用することが考えられる。

その中には、画像情報と言語情報を組み合わせた研究 [2]~[4] がある。我々も言語情報を利用することを考え、自動物体認識に失敗した場合には、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、それをもとに物体認識を試みる [5], [6]。

また、一般に物体認識を行う視覚システムにとって、照明条件の変化への対応も重要である。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 [7] に関する研究が多くなされている。その応用例としては影領域ともと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 [8] がある。

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 [9] や最大の反射率をも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 [10] 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。

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 [11]。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 [12]。Drew ら [13] は、物体のモデルと認識対象の画像の間で特徴の対応をとり、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。

一方では、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 [14] は異なる照明条件下で撮

[†] 大阪大学大学院工学研究科, 吹田市
Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita-shi, 565-0871 Japan

^{††} ミナト医科学株式会社中央研究所, 笹山市
Central Laboratory, Minato Medical Science Inc., 137 Kawanishi, Sasayama-shi, 669-2735 Japan

影した同じシーンの画像中の色の組から color flow を学習し、色変換を行う手法を提案している.[13]や[14]の手法では変換の推定に複数の色の組が必要となるが、ロボットビジョン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ータについては、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。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、新しい環境下で観測された色の組が存在しないため、新たにパラメー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ー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色の組を用いて色変換を推定するためのパラメータを更新する。物体を認識するごと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。

本論文では、まず 2. で冷蔵庫から必要なものを取ってくるサービスロボットののための視覚システムの概要について述べる。次に、3. で色変換のモデルを導入して、認識時の照明条件に合わせた色変換を推定する方法について述べ、4. で推定のためのパラメータの具体的な計算方法について述べる。5. で照明条件の変化への適応に関する実験を行い、6. で結論を述べる。

2. 物体認識システムの概要

サービスロボットの概観を図 1(a) に示す。ユーザは手もとのディスプレイを見ながらロボットに音声で指示を出す(図 1(b))。ロボットは最初にあらかじめ登録された物体モデルを用いて指定された物体(缶・瓶・ペットボトル)を自動で認識し、結果をディスプレイと音声を用いてユーザに伝える。認識が失敗した場合やユーザが間違いを指摘した場合には、ユーザとの音声対話を通して得られる情報を用いて再度物体認識を試みる。以下では、まず 2.1~2.3 で、物体モ

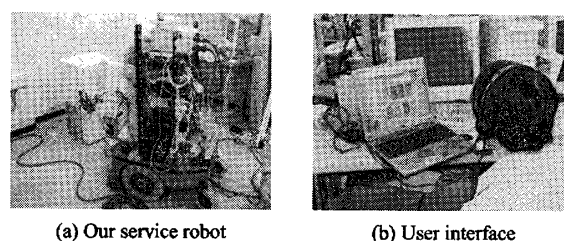


図 1 システムの概要
Fig. 1 Our system.

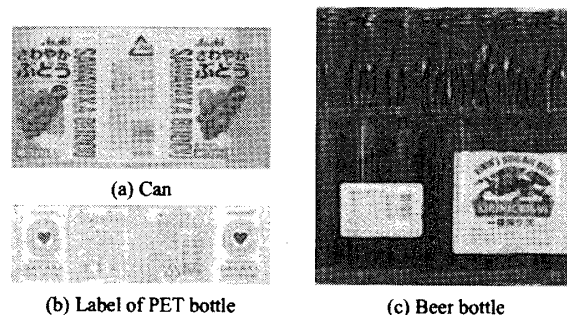


図 2 投影画像
Fig. 2 Projected images.

デルの作成、自動物体認識、対話を用いた物体認識の 3 項目についての概要を述べるが、詳細はそれぞれ文献[15]~[17]を参照されたい。また、本論文のサービスロボットは、音声対話や移動マニピュレータを用いた物体の操作も行うが、その詳細についてはそれぞれ文献[18],[19]を参照されたい。最後に 2.4 で、照明変化と認識精度の関係について述べる。

2.1 物体モデルの作成

物体の特徴としては、物体の実際の大きさ、代表色(最大の面積をもつ一様色領域の色)、及び 2 次特徴(代表色以外の一様色領域の色、位置、面積)を用いる。また、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を取得する必要があるが、効率化のために代わりに 1 枚の投影画像(図 2)を作成する。投影画像は、いくつかの方向から撮影した物体の画像から物体領域を切り出して、それらを重ね合わせて作成する。その際には照明条件を考慮した補正を行う(付録を参照)。最後に、その投影画像から特徴を抽出してモデルとして保存する。以上の方法により、150 種類の物体モデルを作成した。

2.2 自動物体認識

システムは以下の手順に従って物体認識を行う。

- (1) 照明条件に合わせて物体モデルを変換(次節を参照)
- (2) 物体の候補領域を抽出



(a) Original image (b) Extracted regions (c) Recognition result
In (c), black region is extracted secondary feature.

図 3 物体モデルとの照合
Fig. 3 Matching to object model.

(3) 物体の種類を照合して物体の領域を決定

(4) 抽出された領域を物体モデルと照合

図 3 は自動物体認識の例である。この例は、ユーザが左上の黄色の缶（グレープフルーツジュース）を指定した場合である。代表色を用いて缶の領域を抽出すると、図 3(b) のように指定物体に加えて下の棚の黄色の缶（レモンティー）も抽出される。しかし、次に指定物体の 2 次特徴を照合すると左上の缶だけから特徴が抽出される（図 3(c) の黒の領域）ので、左上の缶を指定物体として認識することができる。

2.3 対話を用いた物体認識

扉の影などによって色の变化した物体を自動で抽出することは困難である。そのような場合に、システムはユーザにその物体が見つからなかったことを知らせ、物体のおおよその場所を尋ねる。ユーザから場所が指定されると、システムは影や照明変化などによって起こり得る色の变化を考慮して物体の代表色の範囲を広げて、再び認識を試みる。

2.4 照明変化と認識精度の関係

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 4(a) に示す。ここでは、水色の缶と青色の缶の R , B 成分を例にとっているが、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 4(b) に示すように、作成された全物体モデルの代表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。以下ではこの色推定について述べる。

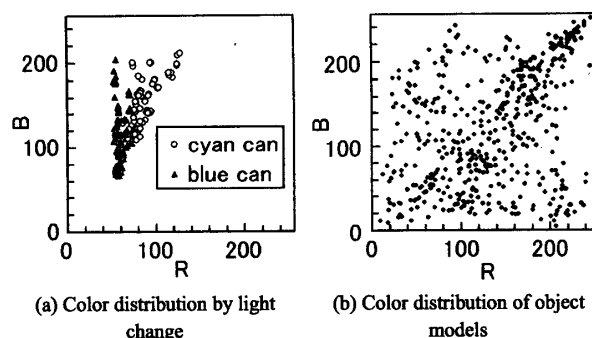


図 4 照明変化による色分布と物体モデルの色分布
Fig. 4 Color distribution by light change and that of object models.

3. 未知照明条件下での色の推定

まず光の反射の物理モデルに基づく線形色変換を定式化する。次に、一つの参照色を用いて未知照明条件下での色変換を推定する方法について述べる。

3.1 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回転体であり、鏡面反射（ハイライト）は物体領域のうちの一部にしか生じないため、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（付録を参照）、認識時には拡散反射のみが生じる部分、すなわち Lambertian 表面とみ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 $c^{OL} = [c_1^{OL}, c_2^{OL}, c_3^{OL}]^T = [R^{OL}, G^{OL}, B^{OL}]^T$ は、以下で表される [20]。

$$c_p^{OL} = \int F_p(\lambda) S^O(\lambda) E^L(\lambda) d\lambda \quad (1)$$

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 の表面反射率の波長分布、 E^L は照明条件 L のもとで物体に到達する光強度の波長分布を表す。また、これらの分布がいくつかの基底関数の線形和で表されるという有限次元線形モデルの仮定 [21] がよく知られており、その仮定を用いると式 (1) は以下ようになる。

$$c_p^{OL} = \sum_{k=1}^{n_S} \sum_{l=1}^{n_E} f_{pkl} s_k^O e_l^L \quad (2)$$

ここで、 s_k^O , e_l^L はそれぞれ $S^O(\lambda)$, $E^L(\lambda)$ を基底関

数の線形和として表したときの重みであり, n_S , n_E は基底関数の数である. 基底関数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 $n_S = n_L = 3$ とする. また, f_{pkl} はカメラレスポンスやその他の基底関数の積分項をまとめたものである. 更に上式を s_k^O について整理すると, 以下のように行列で表現することができる.

$$\mathbf{c}^{OL} = A^L \mathbf{s}^O \quad (3)$$

ここで A^L の pk 成分は $\sum_{l=1}^{n_E} f_{pkl} e_l^L$ である.

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) より以下の式が得られる.

$$\mathbf{c}^{OL_1} = A^{L_1} \mathbf{s}^O, \mathbf{c}^{OL_2} = A^{L_2} \mathbf{s}^O \quad (4)$$

これより線形な色変換が以下のように得られる.

$$\mathbf{c}^{OL_2} = A^{L_2 L_1} \mathbf{c}^{OL_1} \quad (5)$$

ここで, $A^{L_2 L_1} = A^{L_2} (A^{L_1})^{-1}$ である.

3.2 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 (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 5 に示す. 最初に, モデル作成時の基準の照明条件 L_r のもとで観測される参照物体 O_r の色 $\mathbf{c}^{O_r L_r}$ と物体 O の色 $\mathbf{c}^{O L_r}$ が与えられると, 各物体の表面反射率 $\mathbf{s}$ が式 (3) を用いて以下のように表される.

$$\mathbf{s}^{O_r} = (A^{L_r})^{-1} \mathbf{c}^{O_r L_r} \quad (6)$$

$$\mathbf{s}^O = (A^{L_r})^{-1} \mathbf{c}^{O L_r} \quad (7)$$

次に, 未知照明条件 L のもとで観測される参照物体 O_r の色 $\mathbf{c}^{O_r L}$ が与えられると, 式 (2) より,

$$c_p^{O_r L} = \sum_k \sum_l f_{pkl} s_k^{O_r} e_l^L \quad (8)$$

となる. 式 (6) から得られる $\mathbf{s}^{O_r}$ を上式に代入して整理すると, 未知照明条件 $\mathbf{e}^L$ は以下ようになる.

$$\mathbf{e}^L = (B^{O_r})^{-1} \mathbf{c}^{O_r L} \quad (9)$$

ここで, 行列 B^{O_r} の pl 成分は $(B^{O_r})_{pl} = \sum_k \sum_j f_{pkl} ((A^{L_r})^{-1})_{kj} c_j^{O_r L_r}$ である. 結局, 式 (2), (7), (9) から求めるべき $\mathbf{c}^{OL}$ が以下のように得られる.

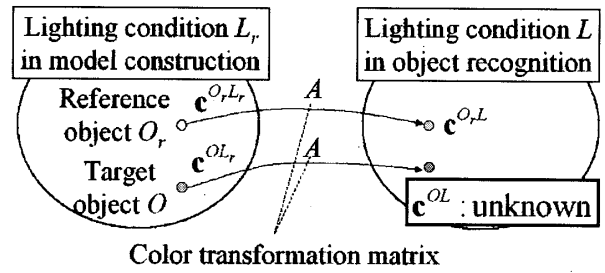


図 5 色推定の問題設定

Fig. 5 Problem setting of color estimation.

$$c_p^{OL} = \sum_j \sum_m \alpha_{pmj} c_m^{O_r L} c_j^{O L_r} \quad (10)$$

ここで, 各 α_{pmj} は色変換の推定パラメータであり, $\alpha_{pmj} = \sum_k \sum_l f_{pkl} ((A^{L_r})^{-1})_{kj} ((B^{O_r})^{-1})_{lm}$ 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 $\mathbf{c}^{O L_r}$, 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 $\mathbf{c}^{O_r L}$ を各々 $\mathbf{c}^m$, $\mathbf{c}^o$, $\mathbf{c}^r$ として定義して, 式 (10) を以下の二つの式で表現する.

$$\mathbf{c}^o = A(\mathbf{c}^r; \boldsymbol{\alpha}) \mathbf{c}^m \quad (11)$$

$$A_{ij}(\mathbf{c}^r; \boldsymbol{\alpha}) = \sum_k \alpha_{ijk} c_k^r \quad (12)$$

前者は物体の色が変換行列 $A(\mathbf{c}^r; \boldsymbol{\alpha})$ 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 $A(\mathbf{c}^r; \boldsymbol{\alpha})$ の各要素 $A_{ij}(\mathbf{c}^r; \boldsymbol{\alpha})$ が観測される参照物体の色 $\mathbf{c}^r$ の線形結合で表されることを意味している.

実際には観測される色は相互反射, 冷蔵庫内部の照明, ガンマ補正やコントラスト補正等によって影響される. ここでは, 相互反射による影響は小さいとみて無視する. また, 冷蔵庫内部の照明光は一定とする. カメラについては, ガンマ補正やコントラスト補正に関するキャリブレーションをしていないものを用いる. これらの影響を考慮して, 近似的に色変換に定数項を追加する. そのために色 $\mathbf{c}$ を $\mathbf{c} = [c_1, c_2, c_3, 1]^T = [R, G, B, 1]^T$ として同次座標で表現して, 基底関数の数を 4 にする. これにより色変換の行列 A は 4×4 となるが 4 行目の要素については, $[0, 0, 0, 1]$ で固定であるので, 実質 12 個のパラメータで変換することになる. よって, 推定パラメータ $\boldsymbol{\alpha}$ の個数は, $12 \times 4 = 48$ となる.

4. 色変換の推定パラメータの学習

ここでは, 前章で述べた色変換を推定するためのパラメータの学習方法について述べる. 最初にパラメータを学習するのに十分な色の組がある場合の学習方法

について述べる．次に，新しい環境下で十分な色の組が得られていない初期段階における推定パラメータの学習方法について述べる．

4.1 学習のための色の組が十分にある場合

色変換の推定パラメータ 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^(注1)．しかし，実際には線形化の仮定などにより，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 2 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 2 乗法の評価式 S_o は以下ようになる．

$$S_o = \sum_{i=1}^N \sum_{j=1}^{n_i} \delta_{ij}^T \delta_{ij} \quad (13)$$

$$\delta_{ij} = A(\mathbf{c}_{ij}^r; \alpha) \mathbf{c}_i^m - \mathbf{c}_{ij}^o \quad (14)$$

ここで N は物体の数， n_i は物体 i に対する観測数， δ_{ij} は物体 i の 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^(注2)．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ー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

$$\Delta_{ij} = A(\mathbf{c}_{ij}^r; \alpha)^{-1} \delta_{ij} = \mathbf{c}_i^m - A(\mathbf{c}_{ij}^r; \alpha)^{-1} \mathbf{c}_{ij}^o \quad (15)$$

を用いる．更に，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，物体ごと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，推定パラメータ 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．

$$S = \sum_{i=1}^N \sum_{j=1}^{n_i} \Delta_{ij}^T \Sigma_i^{-1} \Delta_{ij} \quad (16)$$

ここで Σ_i は物体 i に対する色変換誤差の共分散行列である ($\Sigma_i = (1/n_i) \sum_{j=1}^{n_i} \Delta_{ij} \Delta_{ij}^T$)．式 (16) 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 (13) の S_o を最小化することで求める．

学習した推定パラメータを用いて，ある物体モデルの色を変換した例を図 6 に示す．もとの物体モデル

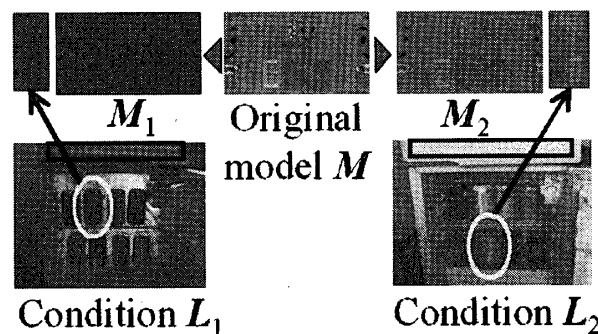


図 6 色変換の結果

Fig. 6 Color transformation result.

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．

また，認識の際には変換誤差をマハラノビス距離によって評価するが，変換誤差が正規分布に従うものと仮定して，マハラノビス距離が 90% の信頼区間に対応する値より小さいときに対象の物体の色とマッチしたとみなす．

4.2 新しい環境下での推定パラメータの学習

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ないので，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，2.3 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．ただし，すべての推定パラメータを求めるためには色の組が足りないため，推定パラメータの中で重要なパラメータのみを求めて，残りのパラメータを 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ー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ータから順番に拘束を解いて，パラメータを更新する．

ここで推定パラメー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) パラメータの拘束をどの順番で解いていくか
- (2) パラメータの拘束をいつ解くか (いくつの色

(注1)：1 組の色から 3 本の方程式が求まる．

(注2)： $\mathbf{c}_i^m$ は物体 i のモデルの色で，照明変化とは関係なく一定値であるために添字 j が落ちる．

の組が得られた時点で解くか)

(1) については、推定パラメータを式 (11) の変換行列 A の対角項 A_{ii} や定数項 A_{i4} に用いられるパラメータのように性質の似たパラメータごとにグループ化しておき、実験的に認識率の向上に効率的に寄与するパラメー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。具体的には、以下のグループ数 $n_p = 9$ のパラメータグループ G_i に分けた。

$$\begin{aligned} G_1\{\alpha_{iii}\}, G_2\{\alpha_{i44}\}, G_3\{\alpha_{i44}\}, G_4\{\alpha_{i4i}\}, \\ G_5\{\alpha_{iji}\}, G_6\{\alpha_{iij}\}, G_7\{\alpha_{ij4}\}, \\ G_8\{\alpha_{i4j}\}, G_9\{\alpha_{ijk}\}. \end{aligned}$$

ここで、 $1 \leq i, j, k \leq 3, j \neq i, k \neq i$ である。また、明度補正のみのパラメータを α_0 として、それに各グループ 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 はすべての拘束を解いた推定パラメータである。

(2) については、各推定パラメー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ータが算出されて認識率が低下する場合があるので、パラメー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、少ないパラメー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ータの拘束を解く。

ただし、パラメー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、パラメータグループの拘束を解くと、その時点で推定パラメータが急激に変化して認識率の変化が起こることがある。そこで、各段階での推定パラメータ α_i ($i = 0, \dots, n_p$) 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ータ α を求める。

$$\alpha = \sum_{i=0}^{n_p} w_i \alpha_i, \quad w_i = \frac{d_i}{\sum_{i=0}^{n_p} d_i} \quad (17)$$

$$d_i = \begin{cases} \delta_{i0} & (n < 2n_i) \\ 1 - (1 - \delta_{i0}) \frac{4n_i - n}{2n_i} & (2n_i \leq n < 4n_i) \\ 1 - (1 - \delta_{in_p}) \frac{n - 4n_i}{2n_i} & (4n_i \leq n < 6n_i) \\ \delta_{in_p} & (n \geq 6n_i) \end{cases} \quad (18)$$

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ータ数、 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、パラメー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、推定パラメー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。

推定パラメー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 (16) 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換えるものとする。

5. 実験

まず、十分に推定パラメー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、参照物体（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 [12] についても同様の実験を行った。

実験結果を表 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ほとん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 (16) 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 2 乗誤差の平均 $S_m = S/(\sum_i n_i)$ 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。

次に初期段階における推定パラメータ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

表 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

Table 1 Success rate of extracting candidate regions and color transformation errors for each method.

(a) Learning sets

method	success1	success2	failure	S_m
diag.	58.1	37.7	4.2	8.16
ours	91.4	8.6	0.0	2.89

(b) Test sets

method	success1	success2	failure	S_m
diag.	58.3	37.3	4.4	8.21
ours	91.2	8.8	0.0	3.01

Success1 and success2 indicate success rates (%) of extraction of the candidate regions in automatic and interactive object recognition respectively. S_m indicates average normalized squared errors of color transformation.

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、色の組を追加するごとに推定パラメータを学習していき、その推定パラメー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7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、学習セットのうちの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ータ数の少ない推定パラメータ($\alpha_0 \sim \alpha_2$)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、少ないパラメー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ータ数の多い推定パラメータ($\alpha_5, \alpha_8, \alpha_9$)では、最初のうち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ータの認識率を上回っている。実際に用いる推定パラメータ α は学習セットの数の少ない間はパラメータ数の少ないパラメータを用いておき、そこから徐々にパラメー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、4.1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。

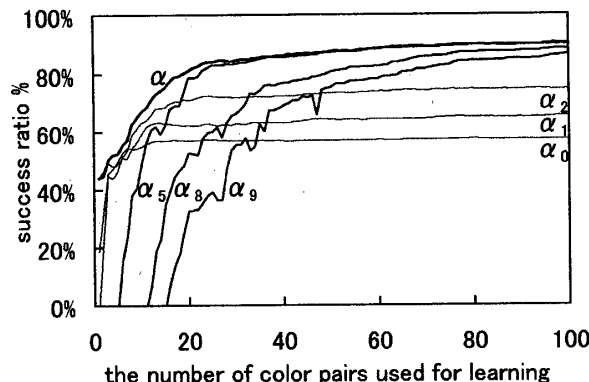


図7 初期段階における推定パラメータ学習の予備実験
Fig. 7 Pretest of performance of parameters for color transformation estimation in initial phase.

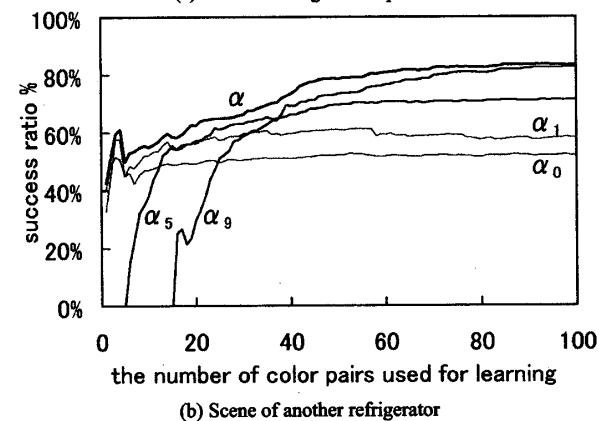
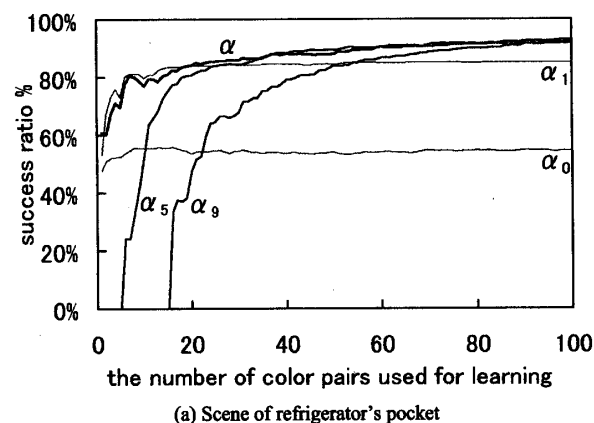


図8 初期段階における推定パラメータの性能実験
Fig. 8 Performance of parameters for color transformation estimation in initial phase.

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 8 は各シーンに対する推定パラメータの学習結果について示している。予備実験のとき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。

6. む す び

本論文では、最初にユーザとの対話を利用した物体認識システムの概要について述べた。また、一つの参照物体の色を観測することで物体認識時の色の変換を推定する手法を提案した。色の変換の推定パラメー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下での誤差を物体ごと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でない初期段階において、推定パラメー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。最後に実験によって、十分に色の組があるときと十分な色の組がな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ータが有効であることを確認した。

今後の課題は、以下のとおりである。

- 光沢のある物体や透明な物体の認識
- 果物やタッパウェア等の登録と認識
- 移動マニピュレータへの実装

文 献

- [1] 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. ICMA2000, pp.549-554, 2000.
- [2] S. Wachsmuth and G. Sagerer, "Integrated analysis of speech and images as a probabilistic decoding process," Proc. 16th ICPR, vol.2, pp.588-592, 2002.
- [3] U. Ahlrichs, J. Fischer, J. Denzler, C. Drexler, H. Niemann, E. Noth, and D. Paulus, "Knowledge based image and speech analysis for service robots," Proc. Workshop on Integration of Speech and Image Understanding, pp.1-27, Kerkyra, Greece, 1999.
- [4] 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.
- [5] 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. Robotics Mechatronics, vol.14, no.2, pp.124-132, 2002.
- [6] 矢野恵生, 横原 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, pp.145-148, 2002.
- [7] B.A. Wandell, "Foundation of vision," Sinauer Associates, Sunderland, MA, 1995.
- [8] S. Nadimi and B. Bhanu, "Moving shadow detection using a physics-based approach," Proc. 16th ICPR, vol.2, pp.701-706, 2002.
- [9] G. Buchsbaum, "A spatial processor model for object color perception," J. Franklin Inst., vol.310, pp.1-26, 1980.
- [10] J.J. McCann, J.A. Hall, and E.H. Land, "Color monodrian experiments: The study of average spectral distributions," J. Opt. Soc. Am., vol.A-67, p.1380, 1977.
- [11] 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.
- [12] M.S. Drew, J. Wei, and Z.-N. Li, "Illumination-invariant image retrieval and video segmentation," Pattern Recognit., vol.32, pp.1369-1388, 1999.
- [13] M.S. Drew, Z. Li, and Z. Tauber, "Illumination color covariant locale-based visual object retrieval," Pattern Recognit., vol.35, no.8, pp.1687-1704, 2002.
- [14] E.G. Miller and K. Tieu, "Color eigenflows: Statistical modeling of joint color changes," Proc. 8th ICCV, vol.1, pp.607-614, 2001.
- [15] 横原 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.
- [16] Y. Makihara, M. Takizawa, Y. Shirai, J. Miura, and N. Shimada, "Object recognition supported by user interaction for service robots," Proc. 16th ICPR, vol.3, pp.561-564, 2002.
- [17] Y. Makihara, M. Takizawa, Y. Shirai, J. Miura, and N. Shimada, "Object recognition supported by user interaction for service robots," Proc. 5th ACCV, vol.2, pp.719-724, 2002.
- [18] 滝澤正夫, 横原 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, 2003.
- [19] 矢野恵生, 三浦 純, 白井良明, 島田伸敬, "移動マニピュレータのための直観的作業教示法," 第 20 回日本ロボット学会学術講演会講演論文集, CD-ROM, 2002.
- [20] 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.
- [21] 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.

付 録

物体モデル作成時の照明条件の補正

物体撮影時には、照明条件やカメラの絞りの変化による画像全体の色の変化、ハイライト、物体表面の法

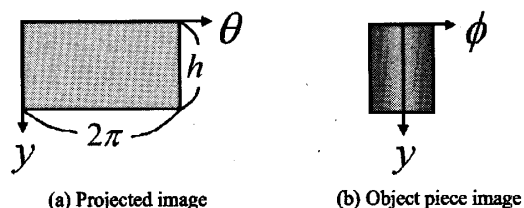


図 A.1 投影画像と物体の断片画像の座標系

Fig. A.1 Coordinates of projected image and object piece image.

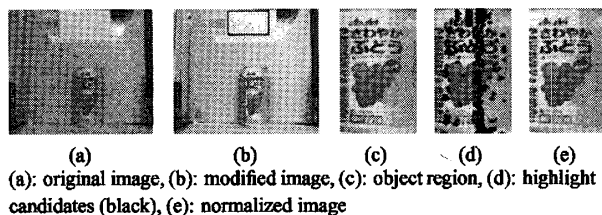


図 A.2 投影画像作成時の照明の補正

Fig. A.2 Modification of lighting condition in constructing projected image.

線方向に依存する照度分布による物体上の明るさの変化などが生じることがあるので、投影画像の作成の際にはそれらを補正する必要がある。

・画像全体の色補正

モデルの撮影場所に参照白色（図 A.2 (b) の黒枠）を用意しておき、観測された参照白色 $\mathbf{c}^r = [c_1^r, c_2^r, c_3^r]^T$ が基準の白色 $\mathbf{c}^s = [c_1^s, c_2^s, c_3^s]^T = [255, 255, 255]^T$ になるように 3.1 で述べた線形変換の対角項を用いて色補正をする。

・ハイライトの補正

最初に物体の領域（図 A.2 (c)）から周囲と比較して輝度の高い領域をハイライトの候補として抽出する（図 A.2 (d)）。位置合わせの際には、候補領域を除外して最もよく合うようにして、モザイクングの際には、ハイライトの候補領域を同じ位置に重なったもう一方の断片画像の色で代用する。

・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 A.1 (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 A.1 (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、

$$I_o(\phi, y; \theta) = l(\phi) I_p(\theta + \phi, y) \quad (\text{A.1})$$

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、すべての方向から見

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。

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

投影画像の周期性より式 (A.2) 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、すなわち定数となるので、 $S(\phi)$ 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 (A.2) の積分は離散的な計算で近似することになり、誤差を含むことになる。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方向から到来する光による余弦項と、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。

$$\hat{l}(\phi) = a \cos(\phi + \phi_0) + b \quad (\text{A.3})$$

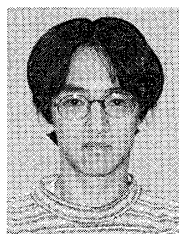
ここで、 a , b , ϕ_0 は分布のパラメータである。ここで、ハイライトの候補領域が抽出された方向を除いて、誤差 $\delta(\phi) = \hat{l}(\phi) - l(\phi)$ の 2 乗和を最小とするパラメータを求める。そして、物体表面上の最大照度で照らしたときの色に合わせるように補正係数 $(a+b)/(a \cos(\phi + \phi_0) + b)$ を各物体の画像の画素値に掛ける。図 A.2 (e) は図 A.2 (c) の照度分布を正規化したものである。

(平成 15 年 5 月 19 日受付, 8 月 26 日再受付)



横原 靖

2001 阪大・工・応用理工卒。2002 同大学院修士課程了。現在、同大学院後期課程在学中。人との対話を用いたサービスロボットのための物体認識に関する研究に従事。2001 年度日本機械学会ロボティクス・メカトロニクス部門ベストプレゼンテーション表彰受賞。日本ロボット学会、日本機械学会各会員。

**滝澤 正夫**

2001 阪大・工・応用理工卒。2003 同大学大学院修士課程了。現在、ミナト医科学中央研究所に勤務。人との対話を用いたサービスロボットのための音声対話に関する研究に従事。日本ロボット学会会員。

**白井 良明 (正員)**

1964 名大・工・機械卒。1969 東大大学院博士課程修了。同年 4 月電子技術総合研究所研究官、研究室長、部長となる。その間、1971 年 8 月より 1 年間米国 MIT の客員研究員。1988 大阪大学工学部教授となり、現在に至る。その間、1996 年 4 月より 3 年間、東京大学大学院工学研究科教授併任。知能ロボットに関する研究に従事。工博。1975 Pattern Recognition Society Award, 1983, 1994 本会論文賞。情報処理学会、人工知能学会各会員。

**島田 伸敬 (正員)**

1992 阪大・工・電子制御機械卒。1997 同大学院博士後期課程了。工博。同年同専攻助手。2001 同研究科研究連携推進室情報ネットワーク部門講師を経て、現在、同研究科助教授。コンピュータビジョン、ジェスチャ認識、ヒューマンインターフェース、インターネットソリューションの研究に従事。情報処理学会、IEEE 各会員。