

## 色競合における適応的事例収集法

山本 俊一<sup>†</sup>

和田 俊和<sup>‡</sup>

奥乃 博<sup>†</sup>

画像から指定した色を検出する方法として、色競合法が提案されている。これは色空間内で最近傍識別を行いルックアップテーブルを構成する手法で、ターゲットと非ターゲットのパターンを教示することにより精度よく、しかも高速に学習と識別が可能であることが知られている。この手法では学習パターンが多いほど検出精度が良くなるが、人手で大量の学習パターンを収集することは困難である。本研究では、動画像において実時間で肌色に基づいて人物追跡を行いながら肌色と非肌色の学習パターンを収集し、肌色の動的学習を行う手法を提案する。追跡中の顔領域と色の変化の連続性に基づいて学習パターン仮説を生成し、色空間における色の類似性と肌色検出への貢献度に基づいて仮説検証を行い、尤もらしい学習パターンを決定し学習する。

### Adaptive Collecting Examples in Color Competition

Shunichi Yamamoto<sup>†</sup>

Toshikazu Wada<sup>‡</sup>

Hiroshi G. Okuno<sup>†</sup>

The color competition method is proposed as a method of detecting the specified color from an image. This is the method of constituting a look up table using nearest neighbor in a color space. This method is known to be able to learn and classify fast and accurately if the method be trained by specifying target and nontarget color examples. This method can accurately detect if there are many training patterns. However, it is difficult to collect many training patterns manually. In this report, we propose a method that collect training patterns of skin color and non-skin color, tracking humans based on skin color, and learn skin color dynamically. This method generates hypotheses of training patterns based on regions of tracking faces and continuity in color variation, and verifies these hypotheses based on similarity of colors and contribution to detecting skin color, and learn likely training patterns.

### 1 はじめに

肌色検出などの色ターゲット検出は、ターゲットと非ターゲットの色の識別問題と見なすことができ、パターン認識的な手法が研究されてきた [1, 2]. 識別に関する研究では、ほとんどの場合色の分布がパラメトリックモデルで表現されるために、ターゲット色

の識別精度があまり高くないという問題がある。この問題の根本的な原因は、3要素からなる色ベクトルを確率や尤度、類似度などの単一のスカラー値に写像することにある。この問題を解決するには、識別を行う前にスカラー値への写像を行わずに、色空間内で単に識別面だけを設定することが考えられる。そのような一手法として、色競合法 [3, 4] があげられる。これは色空間内で最近傍識別を行いルックアップテーブルを構成する手法で、ターゲットと非ターゲットのパターンを教示することにより精度よく、し

<sup>†</sup>京都大学大学院情報科学専攻  
Kyoto University Information Science

<sup>‡</sup>和歌山大学  
Wakayama University

かも高速に学習と識別が可能であることが知られている。この手法では学習パターンが多いほど検出精度が良くなるが、人手で大量の学習パターンを収集することは困難である。そこで、我々は動画像において色の検出を行いながらターゲットと非ターゲットの色の学習パターンを収集し、自動的に学習させる手法を提案する。本研究では、ターゲットの色を肌色としてこの手法により肌色の動的学習を行う。

画像から自動的に肌色と非肌色の事例を収集する場合、いかに正しい肌色と非肌色の事例を収集するかが問題となる。最近傍識別器一般の研究では誤った学習パターンを棄却する Editing [5, 6] の研究が行われてきたが、ここでは色競合に基づく検証フィルタを構築する。4つの肌色の満たすべき性質を定義し、それらの性質に基づいて学習パターンとなる色を決定する。これらの性質は、それぞれ制約の強さが異なっておりそれらの特徴に合わせて仮説生成と検証に使い分ける。

## 2 肌色領域の満たすべき性質

画像から肌色と非肌色の学習パターンを収集するための手がかりとして、色の変化の連続性、顔領域に肌色が含まれること、色空間における色の類似性、肌色検出への貢献度の4つを利用する。これらは肌色の満たすべき性質を表しているため、画像中に含まれる色が肌色か非肌色かを決定する手がかりとなる。

### 2.1 肌色画素の空間的分布

人間の知覚に注目した、肌色領域の満たすべき性質について述べる。「全体が持つ性質から要素の振り方が決定される」という視点から心理的現象を扱う心理学にゲシュタルト心理学 [7] がある。人間は知覚したものの全体を簡潔な秩序あるよい形に体制化しようとする傾向があるとされている。ものは近くにまとまっているとか、連続していたりすると、1つにまとまって知覚される。この考え方を、画像の色が肌色かどうか決定するのに利用する。

具体的には肌色の検出結果のある部分が、回りはすべて肌色と検出されたが、中心部は非肌色と検出された場合を考える。図1の左側のような場合である。この場合、中心部だけが肌色と検出されない

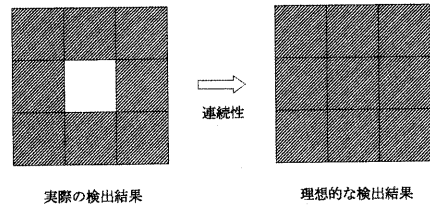


図 1: 3×3 画素の肌色と非肌色の検出パターン

いうことは普通は考えにくく、同図の右側のように中心部も肌色と検出されるべきである。したがって、中心部の色を肌色の学習パターンの候補とする。また、逆に周りが非肌色と検出され、中心部が肌色と検出された場合も同様に、中心部は非肌色と検出されるべきである。

### 2.2 顔領域と動き

入力画像の中で肌色と非肌色の存在する場所を決定するために、次のような仮定を設ける。

- 顔領域には肌色が含まれる。
- 顔領域ではないと判定された肌色領域には非肌色が含まれる。
- 動いているものには肌色が含まれる。

ここで、肌色領域とは色競合法によって検出した肌色を密度解析によってセグメンテーションした矩形領域を指し、そのうち顔の大きさとして妥当な大きさのものを顔領域とする。動きとは、フレーム間差分により画素ごとの色の変化を求め、肌色領域内で色の変化が検出された画素の割合を計算し、閾値以上の場合に動きがあるとする。これらのうち、一つが満たされるだけでは信頼性が低いので、次のような場合に肌色と非肌色の学習パターンの候補とする。

- 顔領域でかつ動きが検出されれば肌色を含む。
- 顔領域でない肌色領域でかつ、動きが検出されなければ非肌色を含む。

これらの条件をもとに、次のような手順で肌色と非肌色の学習パターンの候補を決定する。

- 肌色の学習パターン  
入力画像から得られた各顔領域に対して次の処

理を行う。顔領域内の動きの割合が閾値以上ならば、顔領域内に含まれる色すべてを肌色の学習パターンの候補とする。

- 非肌色の学習パターン

入力画像から得られた顔領域でない各肌色領域に対して次の処理を行う。肌色領域内の動きの割合が閾値以下ならば、肌色領域内に含まれる色すべてを非肌色の学習パターンの候補とする。

このようにして得られた学習パターンには誤りが多く含まれる。特に、顔領域から得た肌色は目や顔の外側など肌色でない部分を多く含む。

### 2.3 色空間における類似性

色空間のなかでの類似性に基づいて、ある色を学習パターンの候補とすべきかどうかを決定する。肌色と非肌色はそれぞれまとまって分布している。ある色が肌色の分布と非肌色の分布からそれぞれ同じくらい離れている場合、その色はどちらの学習パターンとしても良い。ある色が肌色の分布の近いところに存在し、非肌色の分布からは離れている場合、その色を非肌色の学習パターンとすると危険であるので学習パターンの候補としない。

あるパターン  $x$  をクラス A の学習パターンとして教示する場合を考える。パターン  $x$  がクラス A と識別されている場合は、そのパターン  $x$  をクラス A の学習パターンとしても問題ない。次に、パターン  $x$  がクラス B と識別されている場合、それぞれのクラスの最近傍のパターンからの距離は図 2 のような 2 つの場合が考えられる。パターン  $x$  からクラス A 内の最近傍のパターンまでの距離を  $d_A$ 、クラス B 内の最近傍のパターンまでの距離を  $d_B$  とする。距離の比  $\frac{d_A}{d_B}$  が大きい場合は、そのパターンが相対的に識別面に近く、学習パターンとして教示してもよい。しかし、距離の比  $\frac{d_A}{d_B}$  が小さい場合は、識別面から相対的に遠く、そのパターンを学習パターンとすることにより、識別面が大きく変化するため、危険である。

このようにして、ある色が学習パターンとして妥当かどうかを判定するが、LUT による現在の識別器の実装では、アルゴリズムの性質上最近傍のパターンからの距離しか得られない。LUT の更新の際に、それぞれのクラスの最近傍のパターンからの距離を

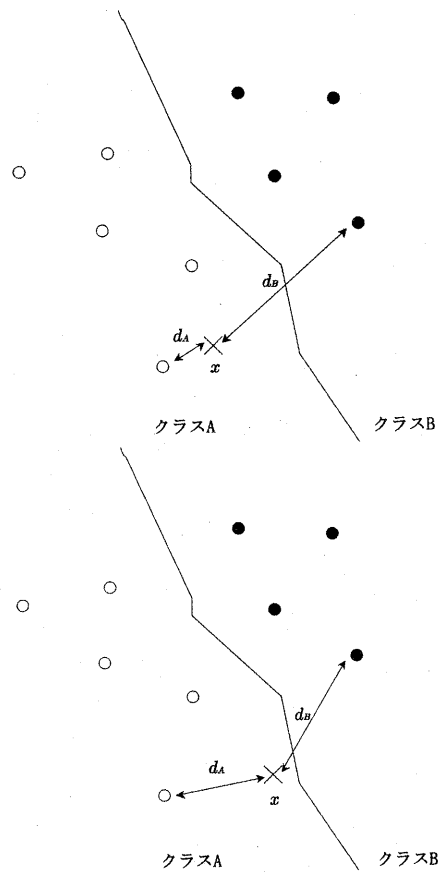


図 2: それぞれのクラスの最近傍のパターンからの距離

計算して置くようにすればよいが、識別面が決定した後も LUT の全要素に対して 2 つの距離を計算しなければならず、LUT の更新にかかる計算時間のコストが増加する。そこで、次のようなヒューリスティックを用いる。

- パターン  $x$  からクラス A の最近傍のパターンまでの距離  $d_A$  が小さければ、クラス B の最近傍のパターンまでの距離  $d_B$  は大きい。
- パターン  $x$  からクラス A の最近傍のパターンまでの距離  $d_A$  が大きければ、クラス B の最近傍のパターンまでの距離  $d_B$  は小さい。

このヒューリスティックを利用して学習パターンを取捨選択する手順は次のようになる。

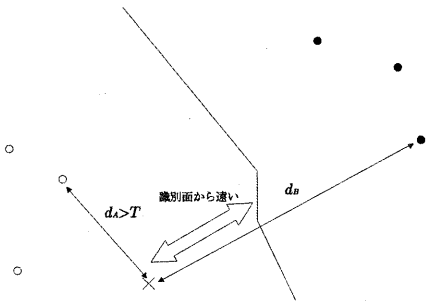


図 3: ヒューリスティックが妥当でない場合の例

1.  $d_A \geq T$  ならばパターン  $x$  はクラス A の学習パターンである
2.  $d_A < T$  ならばパターン  $x$  はクラス A の学習パターンではない

ただし,  $T$  は閾値である.

このヒューリスティックは, 図 2 のような場合には妥当である. しかし, 図 3 のように, クラス A とクラス B のパターン間の距離が大きい場合,  $d_A \geq T$  となる場合が存在し, 識別面から遠いためそのパターンを学習する場合の危険度が大きくなる.

## 2.4 肌色検出への貢献度

すべての色に対して貢献度を付与し, その貢献度に基づいてその色を学習パターンとすべきかどうかを検証する. 貢献度とは, その色が学習パターンとして学習された回数と検出に利用された回数とし, 時間がたつにつれて, 減少していくものとする. つまり, 何度も学習された色や何度も検出されている色は検出に貢献していると見なし, その色が学習されなければ, 時間の経過とともに忘れられていくようにする. この貢献度を一次遅れ系のようなものでモデル化する.

$$y(t, c) = \alpha y(t-1, c) + (1-\alpha)x(t, c)$$

ただし,  $y(t)$  を時刻  $t$  の色  $c$  の貢献度,  $x(t)$  は時刻  $t$  に色  $c$  が学習されたかどうかを表し,  $\alpha$  は減衰に関する定数である.  $x(t, c)$  は色  $c$  が学習された場合または検出に利用された場合は 1 を, それ以外の場合は 0 であるとする. 一時遅れ系ならば, この式によって貢献度が更新されるのは時刻  $t$  ごとに 1 回ずつで

あるが, 同じ時刻でも学習された回数と検出に利用された回数だけこの式によって更新するところが異なっている.

この貢献度に基づいて, 仮説を立てた肌色と非肌色の学習パターンを学習すべきかどうかを検証する手順を述べる. 色  $c$  の現在のクラスと同じクラスの学習パターンであるという仮説は当然成り立つとし, 現在のクラスと異なるクラスの学習パターンであるという仮説に対してのみ以下を適用する. ただし,  $T$  は閾値とする.

1. 貢献度  $y(0, c) = 0$  とする.
2. 時刻  $t$  のとき, 貢献度  $y(t-1, c) \geq T$  ならば, 色  $c$  が学習パターンであるという仮説は棄却する. 採用された場合は, 貢献度  $y(t-1, c) = 0$  とする.
3. すべての色  $c$  に対して,

$$y(t, c) = \alpha y(t-1, c) + (1-\alpha)x(t, c)$$

として貢献度  $y(t, c)$  を更新する. ただし,  $x(t, c) = 0$  とする. 学習された色と検出に利用された色すべてに対して  $x(t, c) = 1$  として上式によって貢献度を更新する.

## 3 適応的事例収集法

2 節では肌色の満たすべき性質によって学習パターンを決定する方法について述べた. これらの性質による制約の強さはすべて同等ではない.

肌色画素の空間的分布による制約は比較的強いものであり, 誤った学習パターンが抽出されることは少ない. 学習パターンの収集にこの制約しか用いない場合, 肌色に囲まれた部分から肌色を補完するというような場合には適しているが, 新たに尤もらしい肌色を獲得することは困難である. 一方, 顔領域と動きによる制約は弱いものであり, 誤った学習パターンが抽出されやすい. しかし, 新たに尤もらしい肌色を獲得するのに適している. これら 2 つの制約は, 逆のクラスの学習パターンとすべき色を収集するのに適している.

色空間における類似性による制約は, 識別面付近の色は肌色と非肌色のどちらの学習パターンとしてもよいというものである. 肌色検出への貢献度によ

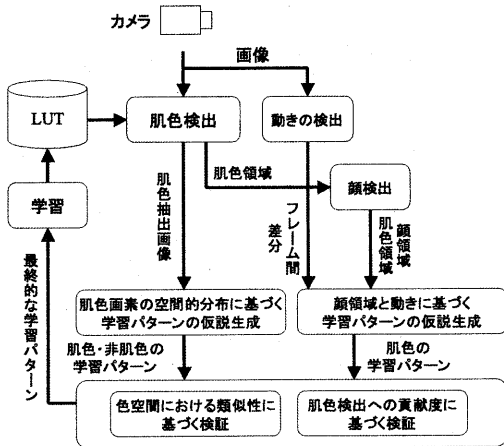


図 4: 肌色の動的学習のシステム構成

る制約は、ある色の現在のクラスへの貢献度が低いものは逆のクラスとして学習してもよいというものである。これら2つの制約は、どちらのクラスの学習パターンとしてもよい場合が存在し、その場合にどちらのクラスの学習パターンとするかを決定できない。

この4つの制約の性質を考慮し、「肌色画素の空間的分布」と「顔領域と動き」に基づいて学習パターンの仮説を生成し、「色空間における類似性」と「肌色検出への貢献度」に基づいて仮説の検証を行うことにする。そこで、色競合における適応的事例集手法による肌色の動的学習のシステム構成を図4のようにする。

## 4 実験

学習パターンの仮説生成と仮説検証フィルタが有効に働くかどうか、また実時間で行えるかどうかを、実験により検証する。カメラにより撮像した60秒間の動画を用いて実験を行い、入力画像のサイズは320×240ピクセルである。カメラのキャリブレーションを行っていないため、入力画像の中心部から外側に行くほど画像が暗くなる。始めに顔をどのように検出するかは与えられるものとし、まず顔が検出されたとしてその後照明の変化などに追従できるかどうかを実験する。初期値として、始めの顔の位置を与え、少し肌色が検出されるように肌色と非肌色の



図 5: 入力画像

学習パターンを手動で与える。

### 4.1 仮説検証の有無の比較実験

次の2つの条件について代表的な検出結果を示し、仮説の検証を行う場合と行わない場合を比較し、仮説検証フィルタの有効性を確かめる。

1. 仮説生成を行うが仮説の検証を行わずに学習する場合。
2. 仮説生成および仮説の検証ありで学習を行う場合。

### 4.2 実験結果

テスト動画の入力に対して、開始から500フレーム付近(約17秒後)の入力画像を図5に、仮説検証フィルタなしの場合の肌色の検出結果を図6に、仮説検証フィルタありの場合の肌色の検出結果を図7に示す。また、仮説検証フィルタありの場合の肌色の検出結果は、動画の60秒間の間誤検出が増大することもなく安定して検出されていた。仮説検証フィルタありの場合の検出結果は次の実験結果で示す。

### 4.3 動的学習の有無の比較実験

次の2つの条件の場合について、学習しない場合と動的に学習する場合の肌色の検出結果を比較し、動的に学習を行うことで照明の変化に追従できることを示す。



図 6: 仮説検証フィルタなしの場合の肌色検出結果



図 7: 仮説検証フィルタありの場合の肌色検出結果

1. 学習しない場合.
2. 仮説生成および仮説の検証ありで学習を行う場合.

#### 4.4 実験結果

学習なしの場合の肌色の検出結果を図 8 に、仮説生成および仮説の検証ありで学習を行なった場合の肌色の検出結果を図 9 に示す。検出を行なった各画像毎に対して水平プロファイルを出力結果とした。これは、 $x$  座標ごとに検出された肌色の画素数のを計算し、値が大きいほど黒く、値が小さいほど白くなるような画像を時系列に沿ってならべたものである。横軸は時刻で、縦軸は入力画像の  $x$  座標である。また、学習ありの場合に検出された顔領域の中心の位

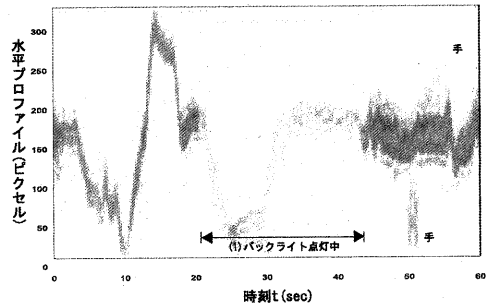


図 8: 肌色の動的学習なし

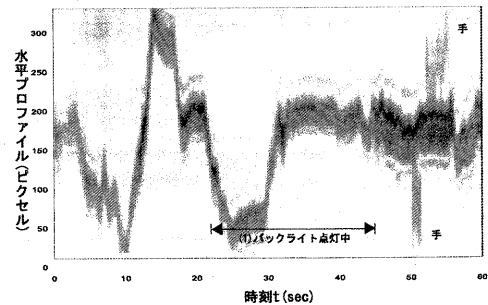


図 9: 肌色の動的学習あり

置を図 10 に示す。肌色領域のうち顔とみなされたものには ID を付けている。そして、距離が近いもの同士が同じ ID となるようにしている。ID0 のグラフは顔であるが、50 秒付近で 2 つの ID1 と ID2 の肌色領域が検出されているが、これは手が視界内に入ったためである。もともと顔であった ID0 が手になり ID1 が顔になっている。このように、肌色だけに基づく顔の検出では重なった時に、区別できない。仮説生成のよって得られた学習パターンの数と仮説検証フィルタに採択された学習パターンを数の推移を図 11 に示す。

1 回の学習に要する時間の平均は 1935ms、1 枚の入力画像の肌色識別や学習パターンの抽出などを行うのに要する時間の平均は 80ms であった。単独で学習を行なった場合、学習に要する時間は 670ms であった。1 回の肌色の検出に要する時間と比較すると、計算時間のコストは大きいので、肌色の検出と学習を交互に行なうと肌色の検出時間が遅くなってしまふ。そこで、学習と識別を並行して行い、実時間での検出を可能としている。LUT を 2 つ持ち検出

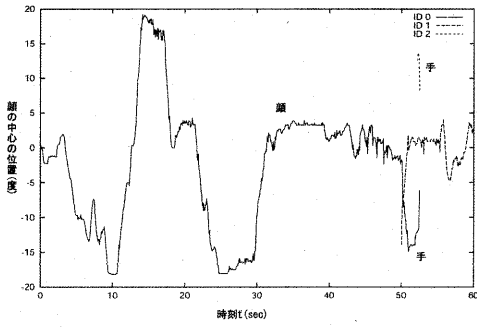


図 10: 顔領域の中心の検出結果

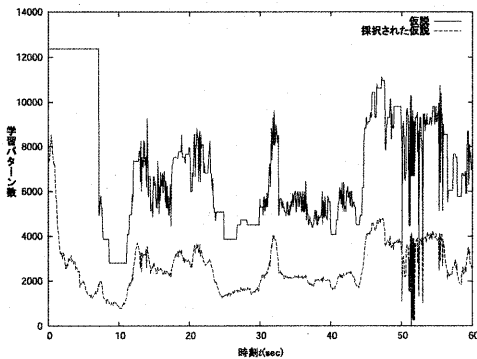


図 11: 生成された仮説と採択された肌色の学習パターン数

用と学習用に分け、検出と学習のスレッドを別々にすることで実現している。検出用スレッドでは検出用 LUT を用いて検出を行ない、学習パターンを収集する。学習の準備が整えば、学習用 LUT の内容を検出用 LUT に反映し、また学習を開始する。余分な処理が増えるため、全体としてのパフォーマンスは低下する。

#### 4.5 考察

学習ありとなしの場合を比較すると学習ありの場合の方が、肌色が精度良く検出されている。学習なしの場合は、あまり多く肌色が検出されておらず、図 8 の (1) の時刻にカメラのバックライトが点灯し、その瞬間に肌色の検出精度が大きく低下している。一方、学習ありの場合は、肌色と非肌色の動的学習が行



なわれているため、肌色と非肌色の学習パターンが適切に収集され、カメラのバックライトが点灯した後も、顔を検出するのに十分なだけ肌色が検出されている。図 12 はバックライト点灯中のある時刻の入力画像と肌色の検出結果である。バックライト点灯中はこのような動的学習ありの場合でも、顔の右側が肌色として検出されることはなかった。情報統合に基づいて仮説生成が行なわれているが、仮説検証フィルタに棄却されたため学習できなかった。この学習できなかった部分の色を取り出してみると、RGB 値で (255,255,255) 付近の色、つまり白がほとんどであった。もしこれを学習してしまった場合は、背景中に多く存在する白い部分を検出してしまい、正常に肌色を検出できなくなる恐れがある。よって、この色は学習しない方が妥当であると言えるだろう。

図 11 より、仮説検証フィルタによって学習パターンの仮説が棄却されている様子が分かる。肌色の学習パターンの仮説と棄却された色の例を図 13 に示す。これを見ると、肌色でない部分が正しく棄却されているのが分かる。

## 5 まとめ

本論文では、色競合法により肌色を検出する場合に、動画から肌色の検出を行いながら自動的に肌色と非肌色の事例を収集し、肌色の動的学習を行う方法について述べた。肌色の満たすべき 4 つの性質を定義し、これらの特徴に合わせて仮説生成と検証という枠組み学習パターンを決定するシステムを構築した。顔領域に基づく方法では誤った仮説が多く生成されるが、仮説検証を行うことでその大部分を

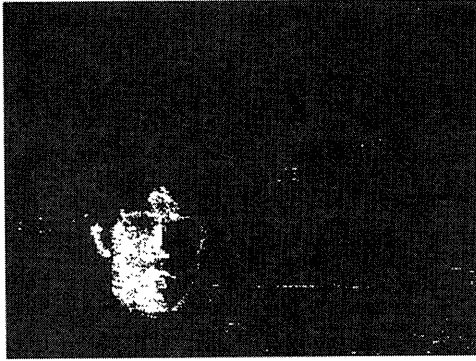


図 12: バックライト点灯中の入力画像と肌色検出結果



図 13: 学習パターンの仮説と棄却された色

除去することができた。これらの仮説生成と仮説検証フィルタの有効性および、実時間での動作が可能かどうかを実験によって確認した。仮説検証フィルタにより顔領域と動きに基づく仮説生成の信頼性の低さを補うことができ、誤学習を防ぐことができた。初期値として、始めの顔の位置と少しの肌色と非肌色を教示することで肌色の動的学習を可能とし、照明の変化に対して追従することができた。また、学習と識別の処理を並行して行うことで実時間での学習と識別が可能となった。

## 参考文献

[1] Crisman, J. and Thorpe, C.: SCARF: A Color Vision System that Tracks Roads and Intersec-

tions, Vol. 9, No. 1, pp. 49-58 (1993).

[2] Nakamura, T. and Ogasawara, T.: On-Line Visual Learning Method for Color Images Segmentation and Object Tracking, pp. 222-228 (1999).

[3] 和田俊和: 最近傍識別器を用いた色ターゲット検出-「らしさ」に基づかない識別とコンピュータビジョンへの応用-, Vol. 24, pp. 17-24 (2002).

[4] 山本俊一, 戸田充彦, 和田俊和, 奥乃博: 実時間人物発見のための視覚と聴覚の情報統合に基づく肌色の動的学習, pp. 269-270 (2003).

[5] Hart, P. E.: The condensed nearest-neighbor rule, Vol. IT-4, No. 5, pp. 515-516 (1968).

[6] 大杉, 上原: プロトタイプ理論に基づく極小事例ベースの構成, Vol. 39, No. 11, pp. 2949-2959 (1998).

[7] Gibson: 『生態学的視覚論』, サイエンス社 ('85).