

# 学位論文の要約

## 「奥行き情報を利用した顔・髪を検出と追跡に関する研究」

鈴木 一正

### 1 研究概要

本研究では、奥行き情報を利用した顔や髪領域の検出・追跡システムを提案する。画像中に含まれる人の顔を検出・追跡する技術は、コンピュータビジョンの中でも重要な研究分野であり、監視カメラや、マシンインタフェース、ロボットとの対話など様々な応用がある。髪領域に関する研究報告は少ないが、髪領域は個人を特徴付ける重要な要素であり、個人識別の特徴として用いたり、髪型を利用したアプリケーションなどの応用が考えられる。また、実際に様々な応用システムを開発するために、高速かつ安定なシステムが要求されている。

本研究の目標は、高速かつ安定な顔や髪の検出・追跡システムを構築することであり、顔検出の高速化と追跡処理の安定化について議論する。また、奥行き情報を利用し、従来手法を改良することでこの目標を達成する。提案システムでは、顔検出で得られる情報に基づいて環境に応じた追跡モデルを自動的に構築し、追跡を開始することができる。

一般的に顔検出を行う場合、画像上から位置や大きさが未知の顔を検出するため、位置や大きさを変えて抜き出した数十万種類のサブウィンドウに対し、識別器を用いた判定を行う必要がある（ピラミッドスキャン）。また、ステレオカメラで顔検出を行った場合、視点や視野の増加により検出精度の向上が期待できるが、単純にピラミッドスキャンを行うと処理コストが倍増してしまう。本研究では、ステレオ処理によって得られる奥行き情報を活用し、識別回数を削減することで顔検出の高速化を行う手法を提案する。また、識別回数の減少にともなう誤検出も減少させることができるため、検出精度の向上も期待することができる。提案手法では、ステレオカメラを用いることで、検出精度を向上させながら単眼カメラよりも高速なビデオレートでの顔検出を実現する。

提案する顔追跡システムは、改良した色弁別度追跡法と顔検出の組み合わせで構築される。先行研究における

色弁別度を用いた顔の検出・追跡システムは、色ベースで顔の検出・追跡を行っているため、手など肌色と似た色の物体を誤って検出してしまうという問題が残っている。また、検出や追跡を行う前に予め肌色と背景の色を手動で教示する必要があるため、照明環境の変化などによってターゲット色と背景色が変わる度に色の再教示を行わなければ、検出や追跡結果が不安定になるという問題も残っている。そこで、本研究では顔識別器を用いてより正確に顔だけを検出できるようにし、検出結果の情報を基に色の学習を自動で行う手法を提案する。提案手法では、顔検出で得られた顔領域内外の色情報を用いてターゲット（肌色）、非ターゲット（背景の色）のヒストグラムを構築し、背景には少なく追跡しやすいターゲット色を選択しながら色学習を行う。また、ターゲット・非ターゲット色を随時更新することで周辺環境の変化に対する頑健性を向上させ、ターゲットまでの奥行き情報を用いた制約により顔追跡の安定化を実現する。

提案する髪追跡システムは、改良した K-means Tracker と顔検出の組み合わせで構築される。髪領域は、テクスチャや色情報が少なく形状も変化するため、モデル化や安定な特徴を抽出することが難しく、これまでビデオレートでの追跡を行っている例は見当たらない。K-means Tracker は、位置情報（画像の  $x, y$  座標）と色情報（ $R, G, B$ ）から構成される 5 次元特徴空間内にて K-means クラスタリングを行い、画像中の画素を対象か背景に分類することで追跡を行う手法である。本論文では、奥行き情報を追加した 6D K-means Tracker を提案し、3 次元化された位置情報によって特徴同士の分離性能を高めることで追跡の安定化を行い、色彩情報の乏しい髪領域の追跡を実現する。また、髪領域を追跡することで、頭部の姿勢や顔の見え方にかかわらず頑健な頭部追跡が可能となる。

## 2 奥行き情報を用いた顔検出の高速化

一般的な顔検出では、ピラミッドスキャンにより膨大な数の識別を行う必要があるが、距離情報を利用した限定的な探索を行うことで、検出を高速、高精度化することができる。人によって顔の大きさにはそれほど大きな違いがないため、ワールド座標空間内の顔の平均的な大きさをあらかじめ決めておくことで、カメラから顔までの距離が分かれば、画像上での顔の大きさ  $I_{size}$  を次式により推定することができる。

$$I_{size} = \frac{f}{Z} W_{size} \quad (1)$$

ここで、 $W_{size}$  はワールド座標における顔の平均サイズ、 $f$  は画素単位の焦点距離で、それぞれ既知の定数であり、 $Z$  はカメラからの距離である。

画像内の各領域を距離に応じた探索サイズ (図 1) のみで顔を探索することによって識別回数を大幅に削減することができるため、顔を高速に検出することができると同時に、誤検出を減少させることができる。

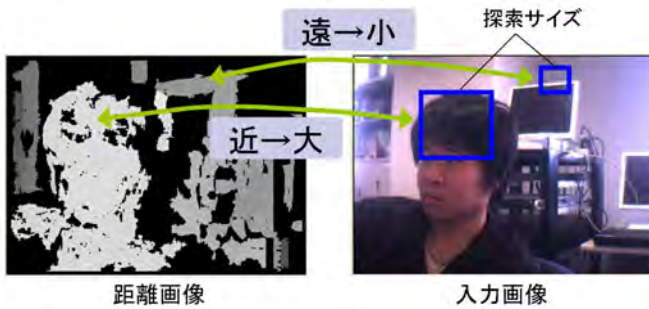


図 1 距離画像から顔サイズの推定

図 2 は、入力画像と距離画像に処理結果を示したものである。ステレオカメラにおける距離画像の構築は、ステレオ処理によって行われるが、処理の高速化のためスパスに配置されたサンプリング点でのみ行っている。距離画像には、限定された探索領域を示している。この領域内を推定されたスケールだけで探索し、顔検出を行った結果を入力画像上に示している。

## 3 顔検出と色弁別度追跡法を組み合わせた顔追跡システム

本研究では、高速顔検出法と色弁別度追跡法を統合することで、自動的に追跡が開始できる高速かつ安定な顔の検出・追跡システムを構築する。色弁別度追跡法は、色に基づく弁別性マップを構築し、楕円フィッティングにより追跡を行う手法である。画素毎に弁別度を算出す

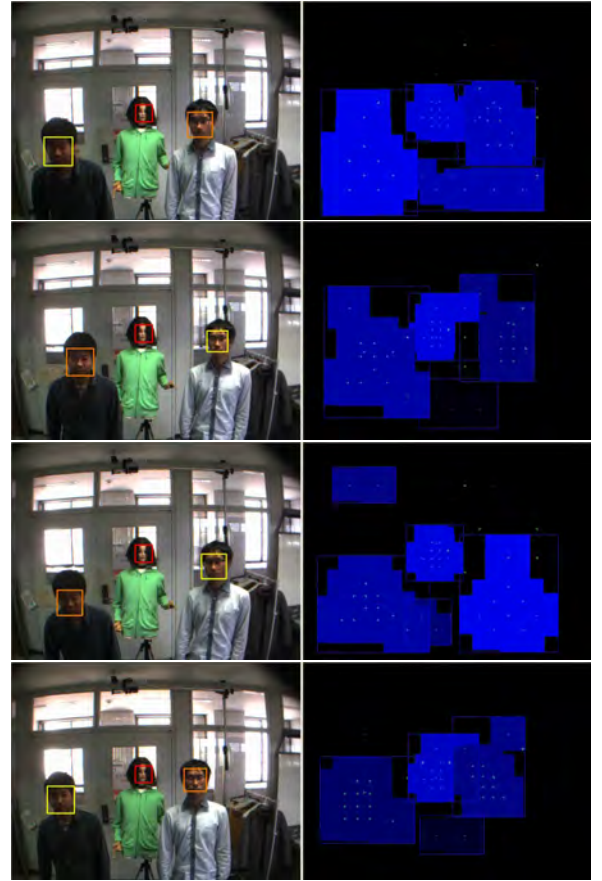


図 2 検出結果例 (左：入力画像と検出結果、右：距離画像と探索領域)

るため、ワイヤオブジェクトの追跡やオクルージョンに対する頑健性を持っている。また、テーブル化することによって高速化した弁別度算出によって弁別性マップを構築しているため、非常に高速な追跡処理が行える。また、弁別度は再近傍識別によるターゲット検出としても利用でき、検出処理と追跡処理を統合することでターゲットの検出・追跡システムを構築している。

しかし従来システムでは、あらかじめ肌色情報 (ターゲットプロトタイプ) を教示し、ターゲット色に基づく検出を行ってため、教示した色と似た色が画像中に入ってくると誤ってターゲットとして検出し、追跡されてしまう問題がある。提案システムでは、色情報ではなく顔識別器を用い、入力画像中から顔だけを正確に検出することでこの問題を解決する。また従来手法では、照明環境の変化や人種、個人差による肌色の違いなどによって、あらかじめ教示したターゲット色と異なる追跡対象者を検出できなかったり、検出に成功したとしても追跡が安定に行えない問題がある。提案手法では、顔検出によって画像中の顔領域が分かるため、検出された顔領域に基づいて、ターゲットと背景の色のプロトタイプを選択し、環境に応じた弁別度 LUT を自動的に構築する。弁別度

表 1 安定化処理時の追跡楕円パラメータの標準偏差  
(単位: 画素)

	x 座標	y 座標	長軸	短軸
安定化処理なし	93.61	20.10	49.38	28.98
楕円制約	2.13	1.13	3.70	0.49
弁別度 LUT 更新	1.38	1.24	1.34	0.60

LUT の構築を自動化することで、提案システムでは起動時に設置環境に合わせたカラーモデルを学習する必要がなくなる。さらに、追跡中にも顔検出を行い、弁別度 LUT のアップデートを行うことで、周辺環境の変化に対する頑健性を向上させる。

従来手法では、背景に対象と非常に似ている色が現れ対象と隣接してしまうと、追跡対象周辺の画素も高い弁別度を持ち、追跡範囲である楕円が広がりすぎてしまうという問題がある。提案手法では、従来研究によりステレオ視によって追跡対象までの距離を測ることができる利点を活かし、カメラから対象までの距離と画像内の楕円サイズを対応付けることで、より安定な探索範囲の決定ができる。提案手法では、人の顔検出・追跡問題に特化しているため、顔の大きさを予め定義しておく、ステレオ視によって得られた距離情報から画像上での顔の大きさを推定することができる。この推定された顔サイズにより、追跡楕円がこれよりも大きくならないように制限することで、楕円が広がりすぎることなく安定な追跡が行える。

図 3 は、追跡の安定化に関する実験の結果を示している。表 1 は、追跡中の楕円パラメータの標準偏差を示したもので、安定化処理を行った場合は数ピクセルの誤差となっており、追跡が安定していることがわかる。

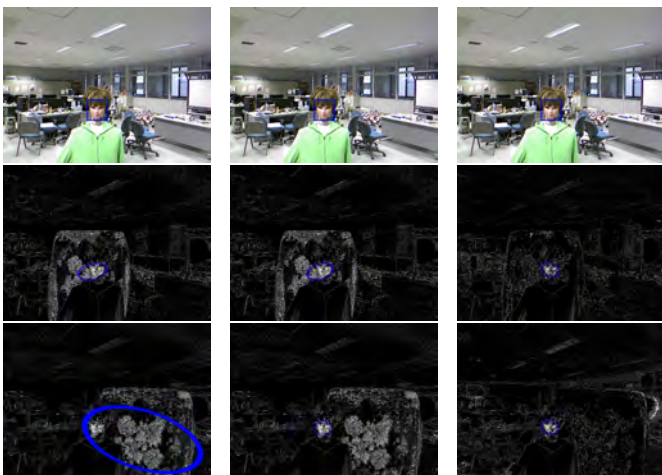


図 3 安定化処理による追跡結果の比較 (左: 安定化処理なし, 中: 楕円制約, 右: 弁別度 LUT 更新)

## 4 顔検出と 6D K-means Tracker を組み合わせた髪追跡システム

5 次元の K-means Tracker に奥行き情報を用いた拡張を行うことで、より安定に顔や髪領域を追跡するシステムの構築を行う。K-means Tracker は、画像上の追跡対象と背景の両方に対して複数のクラスタ中心を割り当て、5 次元特徴空間 (色, 画素位置) における K-means クラスタリングによって、サーチエリア内の各画素をターゲットと非ターゲットにラベリングすることにより追跡を行う手法である。特徴空間内でクラスタ中心と入力画素の距離を計算し、その距離に基づいてターゲットか非ターゲットのラベルが入力画素に付けられる。追跡対象が動けば、クラスタリングの結果、画像上でのターゲットクラスタ中心の重心位置が更新され、ターゲットの追跡が行われる。また、K-means Tracker は画素単位でのクラスタリングを行い、可変楕円をターゲット画素にフィッティングさせることで、追跡対象の大きさや形状の変化にも追従することができるため、様々な形状の髪領域を追跡するのに有効である。

しかし、5 次元の K-means Tracker には次のような問題点がある。追跡対象と類似する色を持つ背景画素がサーチエリアに混入した場合、5 次元特徴空間ではその背景画素が誤ってターゲット画素としてクラスタリングされる可能性が高く、その影響によりターゲット領域やサーチエリアが不安定となり、追跡が失敗してしまう場合がある。また、空間的距離を画像上の画素単位で計算するため、ターゲットが小さく写っている時は 2 次元座標における距離は小さくなり、ターゲットが大きく写っている時は 2 次元座標における距離は大きくなる。このように、画像内のターゲットの大きさによって、5 次元特徴空間内における 2 次元位置特徴ベクトルの割合が変わることで、大きく写っているターゲットの画素が、すべてターゲット画素としてクラスタリングされない可能性が高くなる。さらに、クラスタ中心の初期位置を手動で指定しなければならないという問題がある。

本研究では、奥行き情報を活用し、これらの問題を解決しつつ提案システムの構築を行う。提案システムの特徴を以下にあげる。

1. 奥行き情報により 6 次元に拡張した 6D K-means Tracker を用い、安定な追跡を行う。
2. 奥行き情報を利用した高速な顔検出を行い、検出された顔領域に基づいて、前景 (髪) と背景のクラス



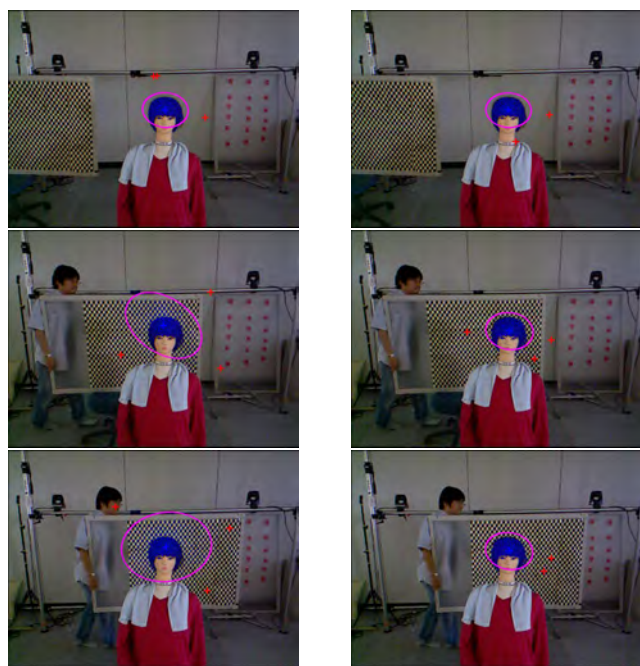
表 2 追跡楕円パラメータの標準偏差（単位：画素）

	x 座標	y 座標	長軸	短軸
従来手法	4.33	4.6	13.19	12.06
提案手法	0.63	0.67	1.07	1.19

タ中心の初期特徴ベクトルを自動的に決定する。

図 4 は、安定性に関する比較を行ったものである。従来手法では、背景にターゲットとの類似色が近づいた時に追跡楕円が広がってしまっているのに対し、提案手法では、全フレームにおいて安定した追跡を行えていることが確認できた。

また、表 2 は、追跡結果の楕円パラメータ（楕円中心の x 座標, y 座標, 楕円の長軸, 短軸）の標準偏差を比較したものであるが、提案手法は数ピクセルの誤差となっており、追跡が安定していることがわかる。



(a) 5D K-menas Tracker (b) 6D K-means Tracker

図 4 類似背景の髪追跡結果例

図 5 は、提案手法を用いて顔だけを追跡した場合と、頭部（顔と髪）を追跡した場合の比較結果である。顔だけの追跡では、顔領域が追跡中に小さくなると追跡が失敗してしまうのに対し、髪領域も同時に追跡した場合は、最終フレームまで失敗することなく追跡を継続することができている。このように、髪領域を追跡することによって、頭部の姿勢変化に関わらず頑健な追跡を行うことが可能となる。



(a) 顔追跡

(b) 頭部追跡

図 5 顔追跡と頭部追跡の比較

## 5 今後の課題

本研究による検出・追跡システムでは、複数の人物を追跡しようとした場合は、新たな人物が画像中に入ってくるのを検出するため、追跡処理時にも検出処理を行わなければならない。追跡処理よりも比較的処理コストが高い検出処理を常に行う必要がある。そのため、追跡対象が増えるにつれ、システム全体としての処理をビデオレートで行うことは難しくなっていく。この問題を解決するためには、検出処理を行う領域を削減するなど、検出処理のコストを抑えることが考えられる。また、K-means Tracker における各画素のクラスタリングを並列処理にするなど追跡処理を高速化することや、検出処理と追跡処理を並列化することでシステム全体を高速化することも考えられる。

以上。