

役に立つ画像処理

立命館大学 白井良明

1 はじめに

コンピュータによる画像処理は、工場などにおける自動化や、われわれの生活を便利にすることなどを目的としている。画像処理の研究は、1960年代に始められた。最初は切り出された文字の認識が主な応用であり、そのための特徴抽出と、特徴を仮定した認識理論が中心であった。人工知能の立場での積み木の認識や、画像処理の立場での泡箱写真解析や顔抽出などの研究も始められていた。

1970年代に、文字認識以外にも工場での生産工程における部品の位置決め、計測、検査などに応用されるようになった。1980年代に入ると、工場以外の応用が広がり始めた。しかし、いずれも小規模であり、広範囲に普及する実用例は見当たらなかった。研究では、情報処理能力の向上とともに、三次元画像処理、動画像解析などが盛んになり、テレビカメラによる道路の自律走行も実験されるようになった。ヒューマンインターフェイスとしての画像処理も現実的となり、ジェスチャー認識、顔認識などの人を見る研究が盛んになった。

1990年以後は、コンピュータの処理能力の向上により、人を対象とする画像処理が実用化された。特に、顔の向きや視線の抽出、顔識別などの技術が進展した。顔の認識は応用範囲は広く、顔は文字に続くコンピュータビジョンの重要な対象となった。

このような状況を考慮し、ここでは、人、特に顔を見る技術を取り上げる。

2 顔の検出と認識

顔の検出と認識は、画像処理の対象として比較的早くから取り上げられてきた。顔の部品である目や口などを抽出することにより、正面顔を検出する試みは 1970 年代初期に始められた。顔の部品の位置関係から、顔を識別し、あ

らかじめ登録した写真と照合することにより、建物などの出入り口の管理を行おうという研究が長年行われたが、いずれも処理速度と信頼性が十分でないため、顔認識はしばらく休止状態となった。

1993年に、米国国防省が顔認識プロジェクト FERET(Face Recognition Technology)を開始した。主な目的は、顔認識システムの開発と評価であり、そのためにいち早く顔データベース (FERET Data Base とよぶ) の作成に着手した。1996年には、1199人合計 14,126 画像 (同じ人でも、異なる時、表情、向き、髪など) を完成し、一般に公開した。世界の顔認識の研究は、FERET データベースを使うことによって加速された。無断で顔写真を使うことは肖像権の侵害となるので、公開可能な顔データは貴重である。このプロジェクトは、2000年からは顔認識システムの評価 (Facial Recognition Vendor Test (FRVT)) に移行し、より大規模なデータベースを用いるようになった。

従来の顔検証のための顔認識は、画面の中央近くに、顔がある程度大きく写っている場合を想定していた。しかし、画面の一部に顔があるような一般的な画像から顔を検出するためには、顔部品候補を抽出し、それらの組み合わせとして顔領域を検出することが必要となる。そこでは、スケールと回転の自由度があるので、計算量が大きくなる欠点があった。

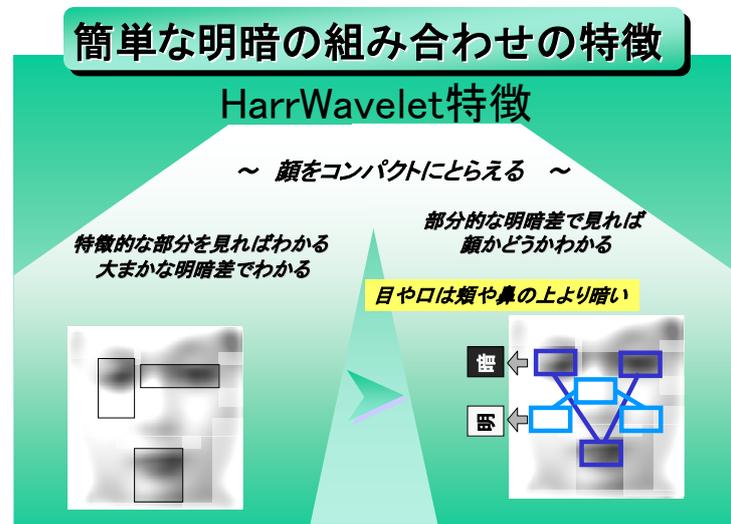


図1 顔候補の検出

2001年に、Viola等は、高速に顔を捜す方法を提案した。顔特徴としては、顔の種々の部分領域に対応する簡単な局所パターンに限定した。それは、顔の部分に対応する明暗の組み合わせを検出する方法である（図1）。

処理を効率よく行うため、まず眼に相当する特徴を検出し、それが検出されたとき（T）のみ口に相当する部分を検出する。このように段階的に処理をおこなう（図2）。顔でない領域の大部分では、早い段階で棄却される（F）ので、計算量が大幅に削減される。

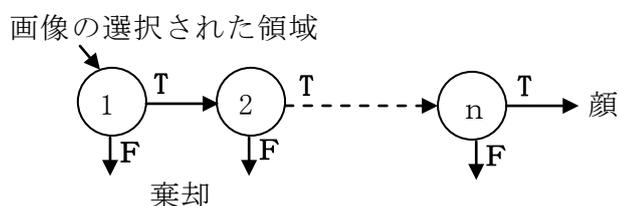


図2 段階的検出法

種々の顔の大きさへ対応するためには、図3のように縮小画像を作り、それを対象とすることによってより大きい顔を認識する。顔の向きに対してもあらかじめ複数の向きの顔に対する検出法を用意しておく。画面内の回転に対してもすべて回転画像を用意し、それに対して同じ処理を施す。計算量は多くなるが、高速化を工夫して実用的な速度で行っている。

このように、すべての可能性を試す方法は、コンピュータ処理の基本と考えられている。

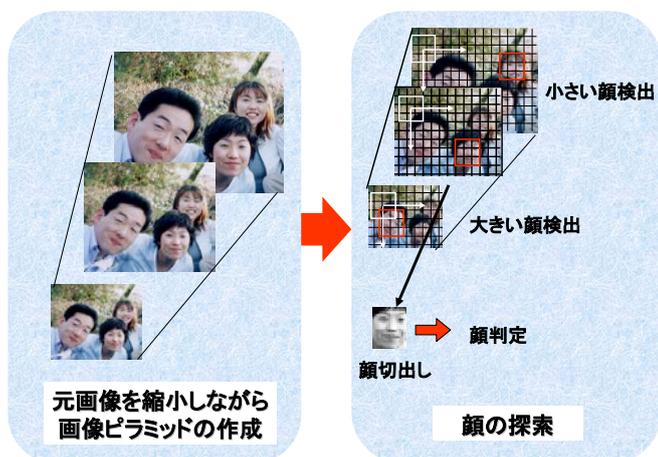


図3 顔の大きさの変化への対応

顔と顔の部品の位置が推定できると、それを標準的な顔の3次元モデルと照合することにより、3次元空間での顔の向きが推定できる。顔の向きと顔部品の位置がわかると、目や口の形をより詳しく抽出することができる。図4に、顔部品を3次元顔モデルに当てはめ、画像上で顔部品の主な特徴点を求める例を示す。このようにして得られた特徴点を利用して顔を認識する。顔検出は、デジカメの焦点や絞り調節、監視カメラなどに応用され、顔認識は機器の利用者の検定やデータベースからの検索に利用されている。

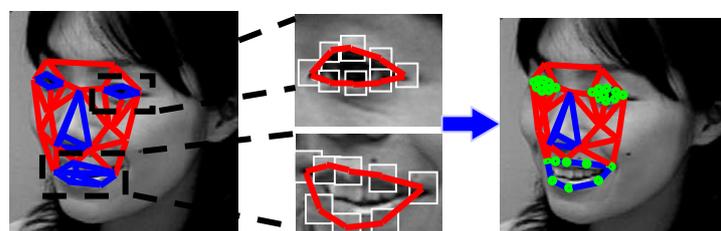


図4 顔部品の形状抽出と3次元モデル当てはめ

顔の表情認識の研究も古くからなされたが、喜び以外は自然な表情を出すことが少ないので、実用化されなかった。現在は喜びを検出してシャッターを切るデジカメが実用化した。

3 人体の認識

人体のシルエット画像から、2次元的な姿勢を推定する研究は1970年代から始められた。1990年代には、コンピュータグラフィクスでの人体の動きの作成やゲームのために、動いている人体の3次元形状を測定する需要が強くなり、人体に装着するモーションキャプチャ装置の代わりに、非接触で画像から推定する研究が盛んになった。

カメラによる監視も盛んになり、米国のDARPA (Defence Advanced Project Agency) は、1997から3年間 Video Surveillance and Monitoring (VSAM) とよばれるプロジェクトを実施し、主に屋外での人の動き、姿勢および荷物の放置などを監視する研究を支援した。図5に、駐車場の監視カメラによって人を検出している例を示す。

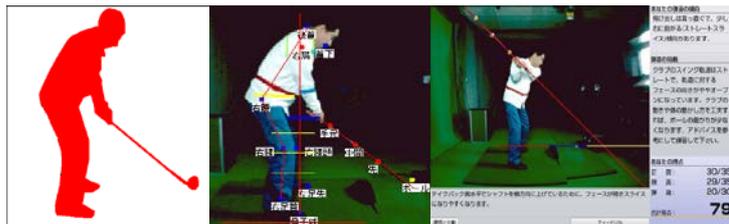


図5 Visual Surveillance の例

人の追跡では、追跡対象が他の人によって隠蔽される場合は、対象を取り違えることがある。その対策としては、動きの連続性、ステレオ視による距離情報の利用、対象領域の色や色のヒストグラムの利用などがある。

色が使えない場合は動きベクトル（オプティカルフロー）を求めて、動きの連続性を利用した追跡も行われている。

動画画像から人体の各部位とその形を求めることもできる。われわれが企業と共同で開発したゴルフ診断システムは、スイング動画画像からシルエットを求め、診断すべきフレームを決定し、手足、肩、ゴルフクラブなどの位置や方向から診断している。



人体シルエット 特徴点 診断結果
図7 動画画像からゴルフスイング診断

これまでは、テレビカメラが主な入力装置であったが、最近、距離を入力できる装置が実用化している。もっとも普及しているのがマイクロソフトが開発したキネクト (Kinect) である。これは、赤外線光源から点パターンを投影し、それをカメラで観測することにより、三角測量の原理で距離を求めている。これと通常のカラ

で用いられ、キネクトのデータを処理するプログラムも多数公開されている。図8には、キネクトから得られた人体の距離画像から骨格を求める例を示す。まだ、対象までの距離が数メートル以内であるため、その応用は限定される。しかし、従来のステレオ視より高性能であり、今後はテレビカメラに代わる入力装置となる可能性がある。

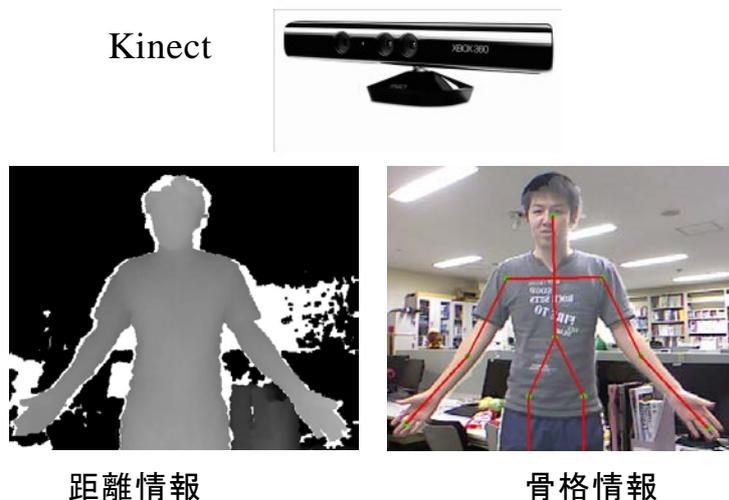


図8 キネクトからの距離情報と骨格情報

4 より高性能な人・顔の検出システムへ

現在、監視カメラが世界中で多数設置され、犯罪捜査などに用いられている。ボストンマラソンでの爆弾事件でも観客の中から犯罪者を特定している。監視カメラの分析には膨大な労力がかかるので、米国では入国者の顔写真を保存し、監視カメラの画像の顔と自動的に照合するシステムがある。しかし、ボストンマラソンの画像は、顔が正面からかなり外れていて、顔の大きさも十分でないため、該当者が見つからなかった。

わが国でも犯罪捜査のための画像処理の重要性が認識され、顔認識の性能向上や監視カメラの機能向上のためのプロジェクトを推進している。ここではその1つを紹介する。

それは、文科省の戦略推進費によるもので、科学技術振興機構が統括しているプロジェクトで、「環境適応型で実用的な人物照合システム」

で、安全・安心社会のための犯罪・テロ対策技術等の実用化を目指している。以下のスライドは、主契約のオムロンの責任者である労氏から提供され、私の講演で示したものの一部である。

科学技術振興機構が統括



文科省科学技術戦略推進費

「環境適応型で実用的な人物照合システム」

安全・安心な社会のための犯罪・テロ対策技術等の実用化

主契約：オムロン株式会社 顔認識、データベース照合

研究代表者 労世紅

協力：京都大学 人物追跡、セキュリティ

東京大学 照明むら対応

名古屋大学 低解像度対応

顔照合技術の課題

一般的な防犯カメラの映像

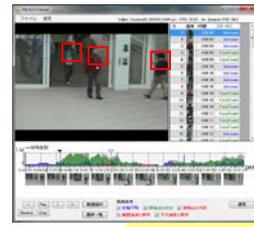


人物照合で移動経路を抽出



人物検出、トラッキング、照明対応、人体向き対応

実用的な捜査支援ツール



高速な顔・人体検出



人物のみ
要約表示

人物検出を用いた
ビデオ要約・検索技術

なお、最後の「実用的な捜査支援ツール」は、中間段階の副産物で、すでに完成している。それは、ビデオから人が移っている画面だけを取り出して人に見せるシステムであり、捜査の効率化を目指している。現状は人に見せているが、人の代わりに自動でデータベースと照合することが最終目標である。

5 おわりに

これ以外に、人の追跡、サービスロボットのための物体認識や移動のための画像処理を示した。これ以外にもロボットや医用などへの画像認識の実用化は進んでいる。現在は種々の処理プログラムも無料で公開されていて、開発環境は向上しているので、より多方面での応用が期待されている。