

## 論文の内容の要旨

論文題目 人を中心としたシーン理解のための実時間認識システムに関する研究

氏名 依田育士

計算機とユーザのより自然な対話を考え、最も手前にあるものが何であるのかを理解し、その動きを捉えるといった特定の対象やモデルに特化しないリアルタイム認識システムに関して研究を行った。このとき、対象の特定領域に着目するキーとして主に距離情報と動き情報を用いて、かつ対象を認識する特徴としてみえ（テクスチャ）をそのまま識別し、認識すべき対象物を前もって学習させる方法を用いた。

実際のシステムは、計算機とユーザとのインタラクションのためにビジョン技術を用いて円滑にすることを目的とし、計算機に見せることによって対象物を計算機に学習させ、さらに学習した対象物を複数個同時に実時間で検出し認識する手法を提案した。具体的には、計算機ユーザの正面ステレオ画像から、奥行き情報とオプティカルフローを実時間で同時に取得し、最も手前から3番目までにあるものとそれらの動きに着目し、それらを常に実時間で認識する手法をシステムとして実装した。この手法をリアルタイム（＝ビデオレート）で実現するために、多重解像度戦略を利用し、入力画像から、ピラミッド画像を生成し、低解像度画像で奥行き情報とオプティカルフローを計算し、そこで注目する小領域を高解像度画像で認識を行った。

その際の認識手法は、モデルを一切持たないテクスチャによる認識を基本とした。これは顔とハンドサイン、さらに身の回りのものといった人を中心としながらも、形が異なるものを同時にビデオレートで認識するためである。具体的には、画像特徴に高次局所自己相関特徴を、識別手法には線形判別分析を用いた。高次局所自己相関特徴は局所情報であるので平行移動に対して不変である。そこで、ステレオカメラからの入力により位置の特定を行い、その切り取られた対象

を認識した。

また、研究の出発点は1組のステレオカメラと1ユーザというものが基本設定であった。しかしながら、さらなる認識性能の向上を考えた場合、処理過程の中で得られた多重解像度画像の利用方法、さらにカメラを多数設置した場合のその多眼画像の利用法、またその両方である多眼多重解像度画像の利用方法とその効果を図るための基礎的な認識実験も行った。

そして最後に今後の研究の方向性を示すものとして、これらステレオカメラを多地点化することで、より広い空間に適應させ、多ユーザにも対応するユビキタスステレオビジョンに関する研究について述べた。

以下、各章単位にその要旨を述べた。

第2章においては、研究の背景について述べた。特に関連研究として顔を中心とする認識、手指の認識とジェスチャ認識について述べた後、本研究が目指す実時間性を伴う人を中心とした総合的な認識について述べた。

第3章においては、高速化のための多重解像度戦略について述べ、特にその奥行き情報とオプティカルフローの取得方法、その両情報を使う意義について言及した。また、距離情報を基にした複数対象の切り出し方法とその各領域の動きの判定について説明した。このとき、距離情報によって対象を分節し、同時に画面全体から得られたオプティカルフローの情報も分節することで、初めて分割された対象単位での動き情報として利用できることを示した。

第4章においては、認識手法について中心的に説明を行った。本論文を通して画像特徴には高次局所自己相関特徴を、認識手法には線形判別分析を用いた。このとき線形判別分析の階層化に関する方法と認識結果の評価方法についても定義を行った。特にここでは、距離情報により注目対象を切り抜く場合、この切り出しが不安定であっても、平行移動に対して不変である高次局所自己相関特徴を用いることで、取得された特徴ベクトルが実際に近似する事例を示した。

第5章においては、実装したインタラクティブシステムについて述べた。このシステムでは、常に1~3個の手前にある複数領域を認識し、そのおおまかな動きを知ることができた。また、カメラの手前に対象を置くだけその領域を取り込めるので、ユーザが提示したものを学習させるインタフェースになり得るものであった。このとき4章で示した認識手法を用いて、個々の対象とカテゴリ単位での認識率向上のために、大分類と小分類という形で判別式を2度計算する手法を実装した。その上で、暗幕背景と複雑な背景の2つの環境下での実験を行い、複雑背景下でも実用に耐え得る認識率が得られることを示した。また典型的なアプリケーション例として、簡単なジェスチャを識別するユーザとのインタラクティブシステムを実装し、ハンドサインによるコントロールを実現した。

第6章においては、特に顔と手に対象を絞って、それらを実時間で同時認識する手法について述べた。はじめに多眼多重解像度画像の利用方法の基礎に関して述べ、高次局所自己相関特徴を認識特徴として用いた場合の多視点での画像の効果、多重解像度の効果、およびその両方の効果を測るために多眼多重解像度画像の生成から特徴の取得方法について定義した。次に実際に顔とハンドサインの認識実験について説明した。まず顔においては個人識別に対する多眼多重解像度

画像の効果を調べ、次に個人識別と同時にビデオレートで計測可能な顔の向きでのその効果を調べた。顔認識については、学習時に定量的な複数方向の画像を与えることにより、1台のカメラで一定範囲内の顔画像認識の精度を高めることが可能であることを示した。また顔の向きの計測については、特に水平方向の向きは、同じ特徴量を用いてビデオレートで同時測定可能であることを示した。一方、ハンドサインに関しては、ジャンケン、指の本数、手話の数詞という3種類について認識時の効果を調べ、手話の数詞の際に判別分析階層化の効果についても調べた。多眼と多重解像度の効果については、単眼よりは多眼、単解像度よりは多重解像度にするすることで認識率の向上に寄与することを確認した。同時に、ハンドサインのように明らかに似通る形同士でも、判別分析階層化の効果があることを明らかにした。

第7章については、1ステレオカメラによる1ユーザという基本的な枠組みを多地点ステレオカメラに拡張することで、対象空間を拡張し、多ユーザへの適応を可能とするユビキタスビジョンについて示した。システム構成と処理戦略を示すとともに、現実シーンにおける実アプリケーション例をあげ、将来の研究の方向性を示した。

なお付録としてGAによる構造的画像処理手順の獲得に関する研究について掲載した。本論文は、人を中心とした認識システムに関する研究であるが、実際の画像処理、パターン認識、コンピュータビジョンの問題を実行する際には、画像処理手順の順序や閾値、各種パラメータなどを実験的、または経験的手法に頼って決定することが不可欠となってしまう。このようなヒューリスティックな問題を自動化するという課題に対して、何らかの解を与えることを目的に基礎的な画像処理手順自動獲得手法について述べた。これは極めて汎用的なものを目指し、本論文内の研究に関して有効なだけでなく、画像処理一般に広く適用可能なものである。