

論 文 要 旨

論 文 題 目

単眼動画からの移動物体の位置と形状の推定に関する研究

身振り手ぶりなどのジェスチャを推定する手法として、データグローブ、磁気式モーションキャプチャ、加速度センサ、関節角度検出センサといったセンサを体に装着する方式がある。これらはセンサを直接体に装着することと専用の機材を用いることで処理の高速性、安定性、計測精度を実現できるが、特殊な機材を装着することになるためユーザに対して拘束感や不自然な使用感を与えてしまい、快適とは言いがたい。

これに対してカメラによるジェスチャの推定は、装着物を使用せずに非接触で行えるため、ユーザに対して拘束感や不自然な使用感を与えないという利点がある。しかし、間接的にジェスチャを推定するため、画像からの特徴抽出などの処理が複雑になるため計算量が高くなり、専用の機材を用いなければ処理の高速性を満たすのが難しくなる。そのため、計算量を削減するために追跡領域の低次のモーメント特徴を用いる手法などが提案されているが、低次のモーメントは対象の大きさや各方向の広がりなどの情報程度しか持たないため、対象が回転すると広がりの方向が変わり誤認識を起こす、細かな形状が扱えないなどの欠点を有する。

ここで人間の手の動作による計算機とのインタラクションを考えると、ハンドサインだけでなく手指の動作からのマウス様な機能などがあり、その実現には手指の形や向きなどの手指の形状認識と手指の運動方向、移動量などの手指の運動認識が必要である。運動認識においては、高解像度による部分的な詳細情報はあまり必要無く、移動物体の検知など外界の変化に対する情報を収集する機能が必要である。それに対し、形状認識においては、詳細な情報を収集する機能が必要である。

本研究では、人間と計算機のコミュニケーションシステムの構築を目的として、カメラから取り込まれた画像から人間の手振りジェスチャを単眼カメラから推定することでマウス様な機能を実現する手法を提案する。

具体的には、まず、時系列画像を直交座標系から対数極座標系に変換を行ない、時間差分、空間差分、色情報を用いて手領域を抽出する。開始直後は、対数極座標の原点を取り込み画像の中心点とするが、手領域の抽出後は、手の位置と速度から次フレームの手領域の推定重心位置を対数極座標の原点として変換を行なう。手形状の識別は、対数極座標空間の高次局所自己相関特徴を用いた識別により行なう。これより、手形状と手の運動軌跡からジェスチャ認識を行なうことで、マウス様な機能を実現する。

対数極座標系に変換することで、半径方向のサンプリングが対数的なものとなり、原点付近で高解像度、背景などの周辺部にいくにつれ低解像度の画像がえられる。これより、空間の状況をだまかに捉えデータ量を大幅に削減し、処理時間を減少させることができる。また、原点を中心として対象物体の回転や大きさの変化は、変換後の画像上では平行移動となることから、この性質を用いて変換後の画像から平行移動に不変な特徴を抽出することで、対象物体の回転や大きさの変化に影響を受け難い特徴をえる。

氏 名 小渡悟

琉球大学大学院
理工学研究科長 殿

論文審査委員

主査 氏名 高良 富夫
副査 氏名 玉城 史朗
副査 氏名 Mohammad Reza Asharif
副査 氏名 星野 聖



学位 (博士) 論文審査及び最終試験の終了報告書

学位 (博士) の申請に対し、学位論文の審査及び最終試験を終了したので、下記の通り報告します。

記

申請者	専攻名 総合知能情報工学専攻 氏名 小渡 悟 学籍番号 [REDACTED]		
指導教官	高良 富夫		
成績評価	学位論文 <input checked="" type="radio"/> 合格 <input type="radio"/> 不合格	最終試験 <input checked="" type="radio"/> 合格 <input type="radio"/> 不合格	
論文題目	単眼動画像からの移動物体の位置と形状の推定に関する研究		
審査要旨 (2000 次以内) 人間と計算機のコミュニケーションの際にも計算機が人間から受け取る情報の一つとしてジェスチャを理解できれば、自然な入力手段となり得るであろうと思われる。そのため人間の手振りなどのジェスチャ認識に基づいた入力デバイスの実現は、従来の計算機のキーボードやマウスに加え、更に快適に利用できる人間と計算機のインタフェースとして注目されている。 これらの研究では、ユーザに物理的なセンサを付けて認識する方法と、画像情報から認識する方法とに大きく分けられ、特に機械視覚システムを用いた画像情報から認識を行う方法は、ユーザに拘束感を与えないことからユーザがより良い操作性や自然な操作感覚をえられるインタフェースになりえるといえる。			

(次頁へ続く)

しかし、現在の機械視覚システムの能力は、人間の視覚と比べると柔軟性・頑健性・適応性などの点でまだ劣っている。人間の視覚は、現在の機械視覚システムとはかなり異なった仕組みに基づいた高い処理能力を発揮している。そのため、この人間の視覚系の機能を模倣すると、実環境での機械視覚システムの精度が向上すると期待される。

そこで本論文では、生体の視覚系の知見を機械視覚システムに応用することで、実環境でも安定して利用できる機械視覚システムについて検討を行い、人間と計算機とのインタフェースを目的とした単眼動画像からの移動物体の位置と形状の推定に適用し、その有効性を示すものである。本研究の主な手法は(1)視覚のモデルとして対数極座標変換(log-polar mapping; LPM)を用いること、および(2)形状の識別には最適化学習法LVQ1により学習したニューラルネットを用いることである。

霊長類の網膜と大脳視覚領との写像関係は、視野の中心部を重視した不均一構造をしており、注目点近傍で密で視野周辺では粗となるサンプリングを行っている。この写像関係のモデルについては、対数極座標変換(log-polar mapping; LPM)が広く使われている。このLPMの特徴としては、少ない画素数で高解像度、広域の作業範囲を得ることだけでなく、対象物体の回転、拡大、縮小が、それぞれLPM画像内の平行移動に置き換えられるというのがある。提案手法では、特徴量として取り込み画像を直交座標系からLPMにて対数極座標系に変換し、空間差分、時間差分、色情報をもとに切り出した対象領域から算出した高次局所自己相関特徴を用いる。

形状の識別には最適化学習率LVQ1により学習したニューラルネットにより行う。ジェスチャ動作の認識は、ユーザの手の動きが極小となり、手形状がシステムに登録されているジェスチャ入力キーと一致したときを動作の開始、または、終了の点として扱い、その間の区間をジェスチャ区間とみなし単純DPによる照合を行う。一般に画像からの動作区間の推定は、手指の動作が極小となる時点を正確に検出することは難しいが、提案手法では動作区間の推定情報として手指動作が極小であるという情報に加え、そのときの手指形状がシステムに登録されているかという情報を加えることで、手指動作の動作区間の決定を実現する。

本提案手法は動画像処理による非接触型の枠組みによる手法である。動画像処理による非接触型にすることでユーザに対して拘束感や不自然な使用感を与えないという利点はあるが、動画像処理による手法では計算量が多くなってしまい実時間での処理が困難になる。従来手法では、入力画像を低解像度に変換し特徴量としてAppearance based method(見え方に基づく手法)を用いることで計算量を減らし処理速度の高速化をはかっているが、低解像度にすることで誤認識が起りやすくなり、また細かな形状や動作が扱えないなどの欠点を有している。その問題に対して提案手法では、入力画像をLPM画像に変換することで、少ない画素数で高解像度、広域の作業範囲を得ることができ、画像のデータ量を減らすことで動画像処理の計算時間を減少させる。また、LPMにはスケーリング不変性、変換中心に対する回転不変性を有しているので、LPM画像から平行移動に対して不変な高次局所自己相関特徴を抽出することで、入力画像上の対象の大きさの変動や対象の回転に対して頑健な特徴量を得ることができる。このように、提案手法は従来手法と比較して、対象物体の細かな形状や動作が扱え、しかも、大きさの変動や対象の回転に対して頑健な特徴量をえるなどの優位性を有している。

以上のように、本論文には工学的に価値のある新しい研究成果が示されており、また、上記の者は専門分野及び関連分野の十分な知識を有することも確認できたので、学位論文及び最終試験とも合格とする。

小渡 悟