

## 審査の結果の要旨

氏名 包 蕊寒

本論文は、“A Hierarchical Action Recognition System Based on Spatio-Temporal Local Motion Feature Descriptors (和訳：ローカルな動きの時空間特徴表現を用いた階層的行動認識システム)”と題し、ビデオシーケンス中の人物の行動認識に関し、動きフィールドの時間変化をベースとして、ローカルな動きの時空間特徴表現を導入するとともに、これを用いて階層的な行動認識システムを構築する研究の成果を纏めたもので、全文 6 章よりなり英文で書かれている。

第 1 章は、序論であり、本研究の背景について議論するとともに、本論文の構成について述べている。

第 2 章は、“A System for Gesture Perception Using Compact Spatio-Temporal Motion Filed Patches”と題し、本研究で構築した行動認識システム全体の構成について述べている。人間の脳における階層的な動き認識のプロセスにヒントを得て、動画像の各フレームより方向性エッジを抽出、これに基づいて動き場を求めるとともに、この動き場の時間変化から行動を認識する構成となっている。静止している背景の情報を消去して動きのある部分のみに着目するため、時間的に変化するエッジフラグだけを残し、その総量が所定の値となるまで積分した Differential Edge Displacement (DED) マップを作成するが、これによって動作のスピードに依存しない認識アルゴリズムを実現している。そして、時間的に連続する DED マップのブロックマッチングから動きフィールドを検出し、行動を動きフィールドの時間シーケンスとしてとらえる。さらに、こうして得られた、空間軸 ( $x, y$ ) 及び時間軸 ( $t$ ) 上に広がる動きフィールドの 3 次元データを、 $dx, dy, dt$  の微小 3 次元領域の patch の集合体としてとらえるが、本研究では、patch 内の動きフィールドの各方向成分を、空間軸上、時間軸上に射影することで、コンパクトなベクトルに変換する方法を提案している。さまざまな種類の動作から patch をランダムに多数サンプリングし、K-means 法で典型的なパターンを 500 個選び出し、これを行動パターン表現の基底として用いる。そして一つの行動シーケンスは、その中に含まれるすべてのパッチと 500 の典型パターンとの類似度を求め、各パターンの最大類似度を要素とするベクトルとして表現される。このベクトル表現では、動作シーケンス内の時空間情報がすべて消去されているため、大変ロバストな認識ができる。腕を回して単純なパターンを描くという比較的簡単な動作に対し、nearest neighbor 分類器で 90%以上の正答率を得ており、隠れマルコフモデルを用いたもっと複雑なシステムよりも良い結果となっている。微小 3 次元領域 patch の時空間軸上への射影によるコンパクトなベクトル表現は、本研究の重要な成果である。

第 3 章は、“Complex Action Recognition Employing Essential Directional Edge Displacement

(E-DED) Maps and Updating Feature Calculations” と題し、前章で確立したシステムの性能向上と、もっと複雑な問題への適用の結果について述べている。認識の基となる動きフィールド検出の精度向上のため、その元となる DED マップに対し、よりノイズの少ない Essential DED (E-DED) マップを導入するとともに、得られた動きフィールドに blur フィルターによるボカシを施すことにより、特徴的な動きのみハイライトされるようにした。その結果、前章より複雑な動作を含む Weisemann human database や、さらに困難な KTH human database に対し高い正答率を得た。これは実用的に重要な成果である。

第 4 章は、“Spatio-temporal Motion Field Descriptors for the Hierarchical Action Recognition System” と題し、本アルゴリズムを VLSI ハードウェア上へ効率よく実装するための新たな微小 3 次元領域 patch の表現方法を提案している。パッチを時空間軸上で等間隔に分割したセルに分け、セル内の各動き方向成分の最大値を用いる MPM 表現法、及び合計値を用いる APM 表現法の二つを提案、第 1 章での表現よりコンパクトで認識性能も高いことが示されている。Weisemann database では、提供されている背景情報を用いることなく先行研究よりも良い結果が得られている。

第 5 章は、“Fisher Discrimination Dictionary Learning (FDDL) via Sparse Representation for Action Recognition” と題し、モデルの学習過程に関し、異なる方法についての検討結果を述べている。顔認識で良い結果の得られている Sparse Representation 法に FDDL 法を組み合わせたシステムでは、Weisemann database に対し 100% の認識率が得られた。しかし KTH database では、第 3 章の結果には及ばなかった。

第 6 章は結論である。

以上要するに本論文は、ビデオシーケンスから抽出した動きフィールドの時間変化をベースに、この中のローカルな時空間微小領域の動きをコンパクトな特徴ベクトルで表現し、これを用いて階層的な行動認識システムを構築するとともにその有用性を実験的に示した研究の成果を纏めたもので、電子工学の発展に寄与するところが少なくない。

よって本論文は博士（工学）の学位請求論文として合格と認められる。