

# Classification and Sentence Description of Human Motions Using Hybrid Generative-discriminative Method

その他のタイトル	生成と判別のハイブリッド手法を用いた人間動作の 識別および文章記述
学位授与年月日	2016-06-16
URL	<a href="http://doi.org/10.15083/00075149">http://doi.org/10.15083/00075149</a>

## 論文の内容の要旨

論文題目      Classification and Sentence Description of Human Motions Using Hybrid Generative-discriminative Method  
(生成と判別のハイブリッド手法を用いた人間動作の識別および文章記述)

氏 名      郷津 優介

産業用ロボットからサービス用ロボットへと需要が移行してくるに伴い、ロボットが人間を近くで観察し、その行動を理解し、意図を把握した上で適切な生活支援を行う必要性が増してきている。本論文では、人間と空間を共有する知能ロボットやシステムにおいて、人間の日常行動を理解するための動作識別と動作から連想される文章生成を扱った。知能ロボットやシステムによる動作識別に誤りが多いと、例えば動作群の履歴などを用いた予測を行う場合に適切な支援ができなくなるなど、ユーザーに不便を感じさせる可能性があるため、動作を正確に識別することは重要である。また、人間の知能が他の動物と異なる点は、言語を通じて実世界を理解し、他者と複雑なコミュニケーションを行えることである。同様に、ロボットが実世界の動作を理解するためには、運動と言語を結び付けることは重要である。従来の動作認識システムでは、身体運動の時系列情報を離散化した記号で表現する運動モデル（隠れマルコフモデル）、運動と単語の間の連想関係を表現する運動言語モデル、単語の並びを表現する自然言語モデルを用いて、観測した動作を識別すると共に複数の文章で記述した。

本論文では、以下の項目に基づいて運動モデルを拡張し、それを動作認識システムに適用することで、動作識別と文章生成の精度向上を図った。

1. 類似した運動パターンを持つ動作群があると身体運動の情報だけでは識別できない場合があるため、周囲環境や対象物体など他の情報まで考慮して識別性能を向上させることは重要である。これに対して、運動モデルの他に音声モデルを合わせたマルチモーダル統合システムを構築し、ジェスチャー識別における有効性の検証を行った。
2. 従来の運動モデルは識別性能が比較的到低く、動作の特徴を捉える運動モデル単体の識別性能を向上させることは重要である。これに対して、生成的アプローチと識別的アプローチのハイブリッド手法を用いて識別に特化したハイブリッドモデルに拡張し、ジェスチャー識別における有効性の検証を行った。

3. Kinectなどを用いたスケルトンベースの動作識別において、一般的によく用いられるスケルトン特徴としてスケルトンモデルにおけるマーカージョイント間の相対的な位置情報がある。しかし、類似した姿勢を持つ動作を識別するために相対的な速度と加速度（身体運動の微分情報）まで含めて識別性能を向上させることは重要である。これに対して、Kinectから直接取得できるマーカージョイントの位置情報から逆運動学計算を用いて取得できる身体運動の微分情報をスケルトン特徴として上記のハイブリッドモデルに適用し、ジェスチャー識別における有効性の検証を行った。
4. 動作の種類により着目すべき身体部位の箇所は異なり、全身による大局的な運動情報よりも動作と密接に関連した局所的な身体部位の運動情報を対象に応じて選択的に利用して動作を識別することは重要である。これに対して、上記のハイブリッドモデルから得られる運動特徴を様々な身体部位と対応させて、動作対象に応じた身体部位の重み付け統合を行うフレームワークを提案し、上半身運動のジェスチャー識別と全身運動の動作識別における有効性の検証を行った。
5. 生活支援における応用を考えた場合、日常生活における人間の様々な動作を識別できることは重要である。これに対して、125種類の動作を含んだ運動データセットを用いて、上記の運動モデルを拡張した提案手法を多クラスの動作識別に適用し、その有効性の検証を行った。また、従来の運動認識システムに適用することで文章生成の評価を行った。

これに対して本論文で得られた結果を以下に記す。

1. 運動モデルと音声モデルを統合するマルチモーダルジェスチャー識別の設計法を提案した。運動と音声の特徴抽出はそれぞれ逆運動学計算、ケプストラム解析により行い、運動と音声から得られる特徴の時系列情報を隠れマルコフモデルで学習した。各モデルから得られる識別スコアは提案手法により統合され、その出力を最終的な識別結果とした。2種類のモデルが相補的に寄与することで、マルチモーダル統合モデルの平均識別率が単一モーダルな運動モデルに比べて88%向上した。
2. 生成的アプローチと識別的アプローチのハイブリッド手法を適用して運動モデルの拡張を行った。複雑な関節運動を伴う動作に対して、高次の運動表現が可能な隠れマルコフモデルに関するフィッシャーベクトルを用いて、時系列情報の表現能力が高い隠れマルコフモデルと多次元の特徴ベクトルを扱えてクラス識別能力が高いサポートベクターマシンの双方の利点を活かすハイブリッドモデルを構築することで、平均識別率が隠れマルコフモデルによるアプローチ（従来の運動モデル）に比べて55%向上した。
3. 上記のハイブリッドモデルに対して、逆運動学計算を用いて取得したマーカージョイント間の相対的な位置・速度・加速度をスケルトン特徴として利用した。速度と

加速度を含めることで、姿勢は似ているが方向の異なる動作と姿勢は似ているが速さと方向の異なる動作の識別にそれぞれ有効に働き、身体運動の微分情報を含めることで平均識別率が相対位置のみの場合に比べて9.4%向上した。また、相対的な位置・速度・加速度を逆運動学計算により取得しているが、Kinectから直接取得できる位置情報のフレーム間差分で求めた場合に比べて平均識別率が高くなることを確認し、取得方法の有効性を示した。

4. 局所的な身体部位ごとの時系列情報から運動モデル群を構築し、隠れマルコフモデルに関するフィッシャーベクトルで運動表現された運動特徴群をマルチカーネル学習で重み付け統合する動作識別の手法を提案した。ジェスチャー識別と動作識別の両方において対象に応じた運動特徴の重み付け統合が識別に有効であることを確認し、それぞれの平均識別率はハイブリッドモデルに比べて25%と35%向上した。また、重みの大きかった運動特徴に対応した身体部位の可視化を行い、識別に有効な身体部位と人間が直感的に重要視する部分が近いことを確認した。
5. 運動モデルを拡張した提案手法を用いることで、non cross-subject test setting（訓練データと評価データは同じ被験者から取得した動作データ）の条件下で、125種類の動作の平均識別率が81%を達成した。また、運動モデルの識別性能が向上したことにより動作から連想される文章の精度も高くなることを確認した。

本論文では、人間と空間を共有する知能ロボットやシステムが人間の行動を理解して適切な生活支援を行うためには動作識別が重要であるとして、“マルチモーダル統合の利用”、“識別に特化したハイブリッドモデルの構築”、“身体運動の微分情報の利用”、“運動に特徴的な身体部位への着目”などの様々な観点からアプローチしてきた。また、これらに基づく提案手法を“日常生活における様々な動作に対応した多クラス識別”に適用することで運動モデルの識別性能が高くなることを確認し、これを用いて従来の動作認識システムを拡張することで連想される文章が正しく記述されることを確認した。運動と言語を結び付けたことで、言語の持つ性質を利用した連想・推論などによる知的な情報処理を行えるようになり、動作以外に物体やシーンなどの情報を同じレイヤーで扱うための統一的な表現形式として、人間が理解でき且つ容易に扱えるという点でも“言語”に変換することは有効である。特に、本論文では文章情報を用いて運動モデルにフィードバックして運動特徴の重みを修正してやることで、識別率および文章生成の精度を高められるという再学習可能な枠組みとなっている。