介護施設における 2 次元姿勢推定を用いた行動識別と言語化

Behavior Classification and Description Generation on 2D Human Postures in Care Facilities

○正　高野渉 (大阪大学)　　李海妍（トヨタ自動車）

Wataru TAKANO, Osaka University, takano@sigmath.es.osaka-u.ac.jp
Haeyeon LEE, Toyota Motor Coorporation

This paper describes a novel approach to classifying whole body motions from estimated human postures in 2D camera images, and subsequent generating their relevant descriptions. The motions are encoded into the stochastic motion models that are referred to as motion primitives. Words are connected to the motion primitives, and word n-grams are represented in the stochastic manners. More specifically, the motion observation is classified into the motion primitive, from which its relevant words are generated. These words are arranged in a grammatically correct order to make the descriptions for the observation. This approach was tested on actions performed by the elderlies in the care facility and its validity was demonstrated.

Key Words: Human Motion, Recognition, Language, Learning

1　緒言

現実世界と仮想世界を繋ぐ新たな社会システムを通じて，社会課題を解決する人間を中心とする社会の到来が期待されている．インターネットが代表するような仮想世界は，言語を介してコンテンツを有機的に結びつけることによって，利便性の高い集合知を形成してきた．一方，現実世界における人間の知能も，言語を用いて世界を記述，記録し，それを伝達，継承することによって，脈々と知識体系を築いてきた歴史でもある．すなわち，仮想世界と現実世界を融合する道具は言語である．

Table 1 21 different kinds of motion patterns

Motion index	Motion label
1	sit at a table
2	turn in a wheel chair
3	talk on a wheel chair
4	get up from a bed
5	slide from a bed
6	move from a bed to a wheel chair
7	move from a wheel chair to a bed
8	lie on a bed
9	move in a wheel chair
10	move from a chair to a wheel chair
11	stand up from a wheel chair
12	move from a sofa to a wheel chair
13	eat at a table
14	drink
15	be fed
16	open a door
17	wash hands
18	walk along the corridor
19	walk along the corridor holding on to the railing
20	fall over
21	lose a balance

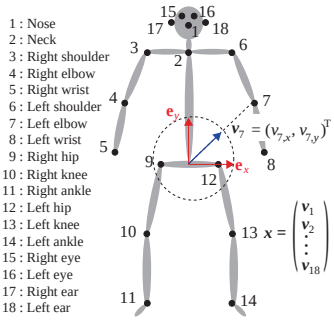

Fig.1 Keypoints and features for whole body motions

社会インフラとして日常生活にカメラが浸透し，人間の行動を見守る社会が進みつつある．様々な時間・場所で記録された人間行動のデータは膨大である．そのデータの利活用や支援分野の広がりは，行動データを如何に構造化し・再利用しやすい形で管理するかに懸かっている．高齢化が進む社会において，若年者が高齢者に常に付き添い，介護することに限界が来ている．この喫緊の社会課題に対して，行動データの蓄積と最先端情報技術はいかなる貢献ができるだろうか？

本研究では，介護施設においてカメラで計測した人間行動から姿勢推定を行い，行動を識別する方法を紹介する．さらに，行動識別と言語を結びつけることによって，行動データを直感的に理解しやすい文章へ変換する言語化技術を紹介する．介護施設での行動を全て機械の手によって言語化することは，自動で介護日誌を作成することになる．カメラ映像を再生して見ることなく，介護日誌を読むだけで日々の施設の状況を把握できる．さらに，普段と異なる行動や非常事態を連想させる文章を検知することによって，必要に応じて迅速に通報連絡する支援システムへ繋がる．

2　全身運動から言語の計算

人間の全身運動をカメラにて撮影する．図 1 に示す 18 個所の特徴点の画像中での位置を検出する [1]．左右の腰（“Left hip” と “Right hip”）の中点を原点，右腰から左腰へ向かう方向を X 軸，原点から首 (“Neck”) へ向かう方向を Y 軸とする 2 次元座標系を考える．ただし，Y 軸は X 軸と直交するように修正するものとする．すなわち，X 方向単位ベクトルを $\boldsymbol{e}_x$，Y 方向単位ベクトルを $\boldsymbol{e}_y$ とする．この 2 次元座標系における原点から第 k 番目の特徴点への単位方向ベクトルを $\boldsymbol{v}_k$ とする．この単位方向ベクトル $(\boldsymbol{v}_k : k = 1, 2, \cdots, 18)$ を並べたベクトル $\boldsymbol{x}$ を全身の姿

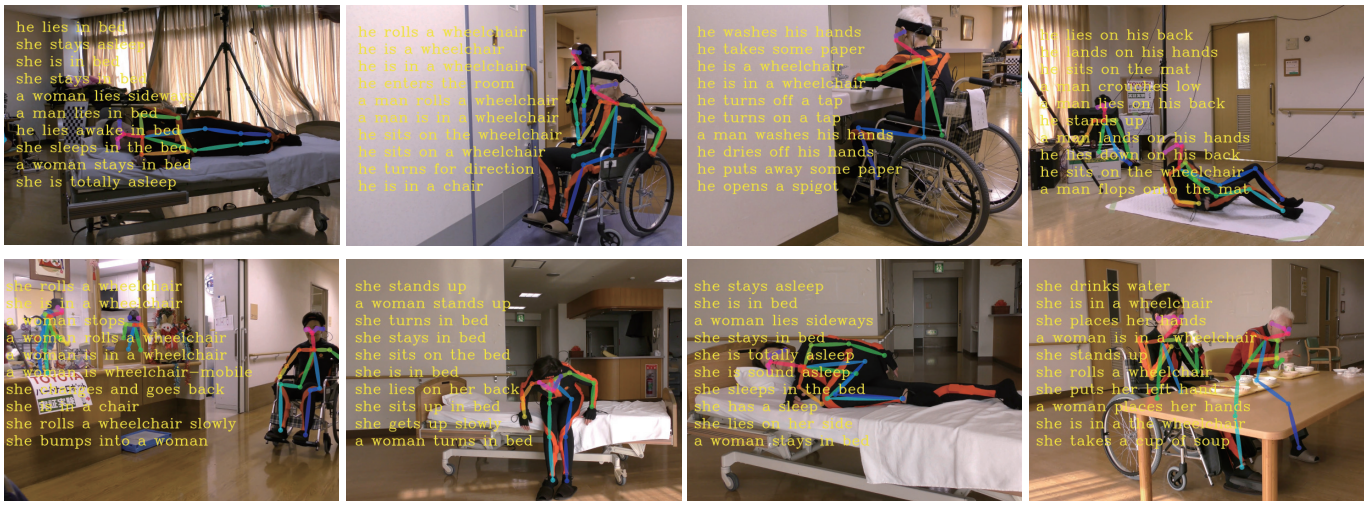

Fig.2 Behaviors are classified into the motion patterns, and subsequently converted to the descriptions.

勢を表す特徴ベクトルとする．この特徴ベクトルは，被験者の体型や並進位置に対して不変である．

全身運動は，特徴ベクトル $\boldsymbol{x}$ の時系列信号として表される，この時系列信号を運動パターン毎の隠れマルコフモデル (Hidden Markov Model:HMM) を用いて学習する．学習は，各運動パターンの時系列特徴ベクトルが HMM から生成される確率を最大にするようにパラメータを最適化する計算である．運動の識別は，観察した運動の特徴ベクトル時系列を生成する確率が最大となる HMM を見つける計算となる．

識別された運動を言語と結びつける統計数理の枠組みは，2 つのモジュール（意味と統語機能）から構成される [2]．意味機能モジュールは，運動と文章中の単語の連想関係を運動識別結果の条件における各単語の生成確率として表現する．統語機能モジュールは，文章中の単語並びに関する規則を単語 n-gram の確率として表現する．運動認識結果に応じて，意味機能モジュールに従って各単語が生成される確率と，統語機能モジュールに従ってそれら単語を並び替えてできる文章の確率を統合した指標に基づいて，動きを表現するに最適な文章を作成することができる．

3　実験

介護施設にて高齢者（男 1 名，女 1 名）の行動をカメラにて撮影した．運動パターンは，表 1 に示す 21 種類であり，各運動パターンに対して 5 試行の行動を撮影した．合計 210 種類の運動データを 10 分割したデータセット（各データセットには，各運動パターンの運動データを 1 個含む）を用意する．9 個のデータセットを学習データ，残り 1 個のデータセットをテストデータとする検定を，すべてのデータがテストデータとなるように繰り返す交差検定を行った，

図 3 に各被験者の運動の識別結果を示す．各被験者のテストデータ数はともに 105 個である．男性被験者の運動データ中，正しく識別できたデータ数は 80 個であり，識別率は 76%である．また，女性被験者の運動データ中，正しく識別できたデータ数は 70 個であり，識別率は 67%である．誤識別の例として，男性被験者の運動 “move in a wheel chair”が “turn in a wheel chair” として識別されている．両運動は，車いすにて移動する動作であり，被験者の姿勢が似ていることが誤認識につながっている．同様に，女性被験者の運動 “turn in a wheel chair”が “open a door”として識別されている．“open a door”の動きは，車いすに乗りながらドアを開けるものである．これも，共に車いすに乗る姿勢が類似している．

運動識別の結果を用いて，運動を説明する言語表現を生成する実験を行った．図 2 では，各運動に対して確率値が高い 10 個の文章が表示されている．男性被験者がベッドで横たわる動きに対して，第 1 候補として “he lies in bed”という文章が生成されている．第 2 候補として，“she stays asleep”という文章が作成されている．本研究では，被験者の推定計算が含まれていないため，“she”という誤った主語を用いた文章が生成されているが，主語以下の “stays asleep”部分では，動きが適切に説明されている．女性が車いすにて移動する動きに対して，“she rolls a wheelchair” という文章が生成されている．このように，高齢者被験者の動きに対して説明文として適切な文章が生成されていることが確認できる．

4　結言

本論文では，人間の全身運動をカメラにて撮影し，被験者の体型や並進位置に不変な特徴量を提案し，その特徴量の時系列データから運動を識別および言語化する方法を提案した．介護施設での高齢者の 21 種類の動きを約 70%で正しく識別できること，およびそれら動きに対して適切な説明文が生成できることを実験によって確認した．

本研究は，平成 30 年度文部科学省科学研究費補助金挑戦的研究 (萌芽)(17K20000) の支援を受けて行った．

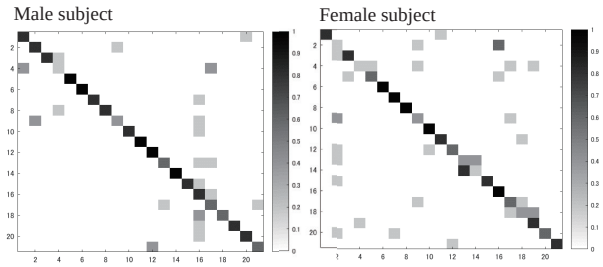

Fig.3 Confusion matrices shows the accuracies of motion classification on the individual subjects.

参考文献

[1] Z.Cao, T. Simon, S.E.Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *In Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, pages 7291–7299, 2017.

[2] W. Takano and Y. Nakamura. Statistical mutual conversion between whole body motion primitives and linguistic sentences for human motions. *International Journal of Robotics Research*, 34(10):1314–1328, 2015.