

転倒検出のための行動検出器の学習データセット構築方法

津田圭一¹ 堀淳二² 村上拓也³ 高原祥平⁴

三菱電機(株)¹⁻⁴

1.背景と目的

高齢化の進行に伴い、高齢者施設や独居高齢者の自宅における転倒や急病等の検出の自動化が望まれている。そこで我々は、カメラ画像から、骨格検出器と行動検出器を用いて転倒を検出するシステムを開発した。

これに際し、行動検出器の学習に、モーションキャプチャで取得した三次元骨格データを用いて学習データの増嵩やアノテーションの省力化を実現したので、ここに報告する。

2.ソフトウェア構成

本節では、転倒検出に用いるソフトウェア全体の構成を説明する。図 1が概略図である。

入力は 320×240 画素、10fps の時系列画像(動画)である。時系列画像からは骨格検出器(今回は Movenet [1]を使用)にて骨格が検出される。

次段が、行動検出器(今回は ST-GCN[2]を使用)であり、30 フレームの時系列骨格を入力とし、行動クラス(立・座位/転倒中/臥位/その他)ごとの尤度を算出する。

さらに最終段の転倒判定部が、行動尤度の時系列から、ルールベースで転倒発報の是非を判断し、転倒アラームを出力する。以降に各ブロックの詳細を示す。

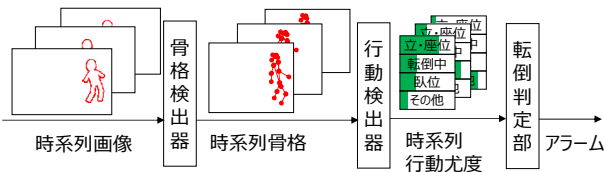


図 1 ソフトウェア概略構成

2.1.骨格検出器

本節では、図 1の初段の骨格検出器について述べる。Movenet には検出人数が一人のものと複数人のものがあるが、今回は複数人が映っている状況をスコープ外とし、前者を用いた。

2.2.行動検出器

本節では、図 1の中段の行動検出器について述べる。行動検出器の学習に関しては、今回、学習データ収録やアノテーションなども行い、結合強度学習も初期値から行った。

我々は、そのために必要なデータ量やアノテ

Fall Detection using Pose Estimation and Action Recognition
1 Keiichi Tsuda, 2 Junji Hori, 3 Takuya Murakami,
4 Shohei Takahara
1-4 Mitsubishi Electric Corp.

ション工数が増大する課題への対処のため、三次元骨格を元にデータを増嵩した。詳細は 3 節にて述べる。

2.3.転倒判定部

本節では、図 1の終段の転倒判定部について述べる。本システムでは、前段出力の行動尤度の時系列から、発報すべきと考えられるパターンを検出し、アラームを発出する。

3.ST-GCN 学習

ST-GCN を含む学習ベースの検出器では、多くの学習データが要求されること、それに伴い多大なアノテーション工数が必要なことが開発の足枷となっている。そこで我々は、三次元骨格を用いた学習データの増嵩を行った。これは、学習に必要な二次元骨格を、三次元骨格を様々な回転させて二次元面に投影することで多数生成するという方法である。以下に詳述する。

3.1.データ収録

まず、三次元モーションキャプチャを用いて転倒や歩行動作の三次元骨格を取得した。また、この時点でアノテーションも行うことで、結果的にアノテーションの工数も削減できる。アノテーションのクラスは前述の通り、立・座位/転倒中/臥位/その他である。

収録したデータは 61 シーン、それぞれのシーンは 1 分程度である。また、高齢者の転倒等のデータ収集は困難なため、俳優に高齢者を模した演技を指示してデータ収録を行った。

3.2.二次元投影

三次元データの投影には、平行投影法と透視投影法がある。今回は、パラメータが比較的少ない平行投影法を用いる。

回転パラメータは、オイラー角、すなわち、ロール、ピッチ、ヨーである。表 1にその範囲を、また回転イメージを図 3に示す。以下に、3 パラメータの設定の意味を示す。

まずロール角は、0° 固定である。これは、撮影光軸の周囲の回転だが、現実的に撮影シチュエーションによらず、一定と考えて構わない。

表 1 回転角範囲

	範囲	ステップ	パターン数
ロール角	0°		1
ピッチ角	20° ~50°	10°	4
ヨー角	0° ~340°	20°	18

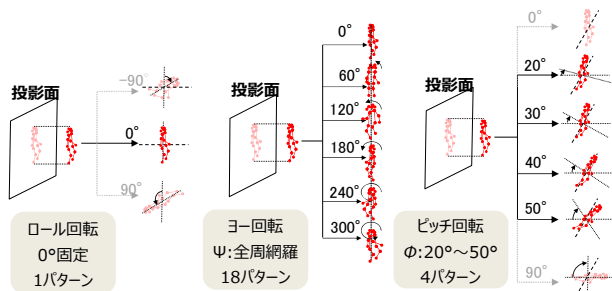


図 3 回転イメージ

理由は、撮影光軸周りの回転により、重力方向の見えが変わるが、人間の姿勢は重力方向に拘束されており、ロール方向の角度バリエーションは殆ど生じないことである。また、図 3左のように、0° では立位となる姿勢が、90° 回転をすると臥位となるように、姿勢の意味合いが変わるため、ロール角は固定することが望ましい。

次にヨー角は全周を網羅する。これはどれも、人の向きとして自然な姿であり、ロールのように姿勢としての意味合いが変わることもなく、また、ヨー角に制限を設けると、例えば右向きの転倒は検知でき、左向きの転倒は検知できないなどという事態が生じるためである。

最後にピッチ回転は、20° ~50° の範囲とした。これはカメラの見下ろし角にあたるが、ロール角のように一定値に定めることはできず、しかしヨー角のように任意の値を許容することもできない。前者の理由は、監視カメラや見守りカメラが実際に設置される際の見下ろし角は、天井の高さや部屋の形状によって異なり、後者の理由は、0° での立位が 90° では臥位になるなど、ロール角同様に姿勢としての意味合いが変わるからである。今回 20° ~50° としたが、これは実験で好成績となる範囲を選定した結果である。

以上のように、1 シーンの三次元骨格から、72通りの二次元骨格を得た。これは同時に、アノテーション工数を 1/72 に削減したことになる。

3.3. 学習結果

以上のデータを set0~3 に 4 分割して 50epoch の学習を行い、ST-GCN 単体評価を行った。方法は交差検証(該当 set を評価、残りを学習に使用)である。各 set の結果を表 2 に示す。

図のように、概ね 75%程度の正解率となっている。不正解となる一要因として、奥向きの転倒中や臥位の、立位との混同が挙げられる。

表 2 行動検出器単体評価結果

	set0	set1	set2	set3
正解率	74.38%	74.54%	77.44%	71.54%

4. 実動画への適用結果

本転倒検出器を動画に適用した例を図 2 に示す。

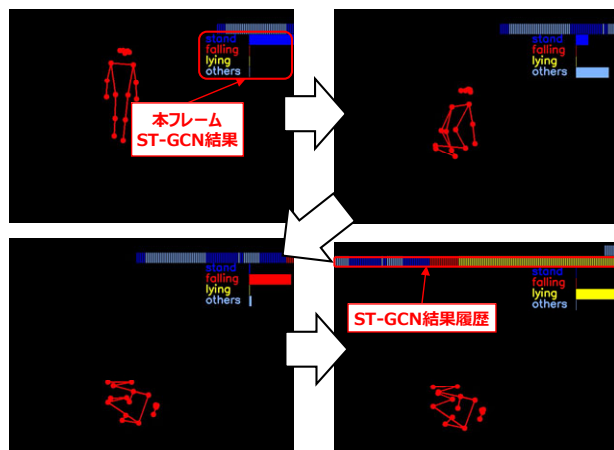


図 2 動画適用結果

図は、インターネット上に公開されているデータセット、Le2i Dataset[3]の一動画への適用結果である(図には動画そのものは表示せず、骨格と ST-GCN 結果のみを示す)。右上のバーグラフが該当フレームの ST-GCN 結果、その上のバーが ST-GCN の履歴(青:立・座位/赤:転倒中/黄:臥位/水色:その他)である。想定通り、3 コマ目に転倒中(赤)や 4 コマ目に臥位(黄)が現れている。

モーションキャプチャと Movenet の骨格の傾向には差がある可能性があったが、結果から判断すると、ST-GCN の汎化特性が差を吸収し、前者で構築した学習モデルは後者にも有効と言える。

5. まとめ

骨格検出器 Movenet と、行動検出器 ST-GCN を用いた転倒検出器を構築した。ST-GCN の学習データ取得の際、三次元骨格座標を取得してそれを回転後二次元投影することでデータを増嵩した。その際の回転角度として、姿勢の意味合いと角度のカバー範囲の観点から、実用上最適な範囲を設定できた。

この方法により、1 通りの三次元骨格から 72 通りの学習データを得ることができ、実動画へ適用しても概ね良好な結果を得た。

今後は、転倒検出システム全体としての評価や、その精度向上のための改良を行っていく。

引用文献

[1] MoveNet : 超高速で正確なポーズ検出モデル google LLC, <https://www.tensorflow.org/hub/tutorials/movenet?hl=ja>
 [2] Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition, AAAI, 2018.
 [3] Le2i Dataset : https://github.com/YifeiYang210/Fall_Detection_dataset