

人物姿勢推定アルゴリズムをあれこれ試す

ラズパイで エッジAI

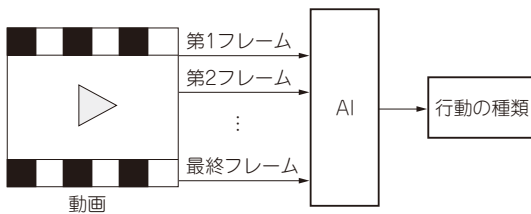


最終回

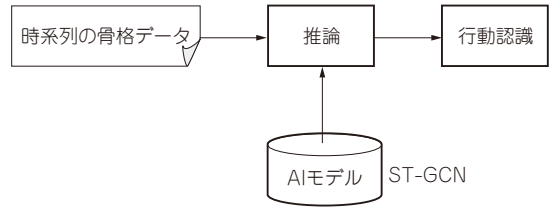
第5回

時系列の骨格情報から行動を認識する

笹尾 幸良



(a) 各フレーム全てをAIに入力して推定する



(b) 時系列の骨格位置をAIに入力して推定する

図1 行動認識を実現する方法には2つの方法がある

人間は、物を投げる、拍手をする、食べ物を食べる、スケートボードをするなど多様な行動をとります。これらは静止した人の状態からは判別できず、動画で判断する必要があります。動画の中の人物がどの行動種類をしているか識別するタスクは、行動認識 (Action Recognition) と呼ばれます。今回はこの行動認識について、AIモデルMMSkeletonを使って試してみます。

● 行動認識をAIで実現する方法

行動認識には次の2つの方法があります。

1. 動画の各フレーム画像全てをAIに入力する [図1 (a)]
2. 動画の各フレームで人の骨格位置を推定し、その時系列の骨格位置をAIに入力して推定する [図1 (b)]

2は骨だけの人間が動く様子を見て何の行動をしているか当てるようなものですが、1の動画の画像全てをAIに入力するよりは高速に処理を行うことが可能です。本稿では2の方法を紹介します。

使用するアルゴリズム

アルゴリズムはST-GCN⁽¹⁾を使います。ST-GCN (Spatial Temporal Graph Convolutional Networks) は2018年に提案されたアルゴリズムで次の手順を踏みます。

1. 時系列の骨格の座標を入力として空間方向である同じフレーム内の関節位置同士を結ぶ。
2. 時間方向である時系列の関節位置同士を結び時空間グラフを形成する。
3. このグラフを畳み込みニューラル・ネットワークによって処理し、行動種類を識別する。

使用するAIモデル

ST-GCNの実装とモデルはMMSkeleton⁽²⁾を利用します。ST-GCNが技術名(論文タイトル)で、MMSkeletonはそれをオープンソースのコードとして実装したものです。作者は同じです。MMSkeletonはPyTorchで学習されており、モデルも量子化されていないPyTorch形式です。今回はこのまま、この事前学習済みモデルをPyTorchでラズベリー・パイ上で実行します。

MMSkeletonのST-GCNモデルには幾つかの種類がありますが、今回はkinetics-skeletonを用います。このモデルは400種類の行動を識別するように学習されており、例えば

- arm wrestling : 腕ずもう
- bowling : ボウリング
- changing wheel : ホイール交換
- crawling baby : 赤ちゃんのハイハイ
- feeding birds : 餌を鳥にあげる

などがあります。詳しくは文献(3)を参照してください