

OpenPose



my recommendations on research & development tools.

稲田健太郎^{†1}, 大澤遼平^{†2}, 山川敦也^{†3}

キーワード: 人物解析, 運動解析, 姿勢推定

1. まえがき

2020年の東京オリンピック・パラリンピック競技大会の開催を控え、試合中の審判の判定の補助、スポーツ選手の育成、視聴者のための新たな映像コンテンツの提供などを目的としたスポーツ映像処理技術の需要が高まっています。実際にスポーツ分野で映像技術が応用されている例として、テニスの国際大会で審判の判定の補助として採用されているThe Hawk-Eye Officiating System¹⁾などが挙げられます。しかし、このような試みの多くはプロ選手のために行われています。実用には、解析用の機器や複数の高性能カメラなど導入にコストが必要となります。このため、アマチュア選手への実用は難しいのが現状です。そこで、デジタルカメラやスマートフォンなどの一般的なカメラで撮影された映像に対して有効な動作解析手法が求められています。これにより、過去の映像に対しても解析を進めることができ、引退選手との比較などが可能となります。これからのスポーツ映像処理の分野の技術進歩が期待されます。

今回紹介するOpenPose²⁾は、従来使用されていたKinectより同時に多くの人数の検出が可能です。また、検出精度に関してもKinect以上なツールです。そして、カメラだけで検出が行えるため、現在入手困難なKinectに頼らず人物の動作解析が可能です。しかし、OpenPoseは検出が2次元であることとリアルタイムでの処理が難しいことが欠点として挙げられます。前者は、OpenPoseとは別に2次元データを3次元データにするツールを使用することで解決できるようになります。後者は、リアルタイムでの処理はGPUやカメラの性能に依存しますが、多くても毎秒10フレーム程度なので注意が必要です。

2. OpenPoseの手法

2.1 アーキテクチャ

従来のmulti-personの姿勢推定手法としてはtop-downのアプローチ、すなわち画像中の人物を検出し、検出した各人に対してsingle-personのpose-estimationを施す方法がとられてきました。これでは画像中の人物の検出がうまくいかなかった時点でその人の姿勢推定は失敗となり、また画像中の人数が増えれば増えるほど計算量は膨大になっていきます。これに対して提案手法はbottom-upのアプローチをとっています。具体的には、まず画像中の関節点を推定します。左手首なら画像中の全員の左手首を、右肩なら全員の右肩をそれぞれまとめて推定していき、各関節点の確率分布(2次元)を生成します。これがPart Confidence Maps(以下PCM)と呼ばれるもので(図1)、入力画像サイズを $w \times h$ 、関節点を j 個として、PCM S は

$$S = (S_1, S_2, \dots, S_j) \text{ where } S_j \in \mathbb{R}^{w \times h}, j \in \{1, 2, \dots, j\}$$

として得られます。そして得られた関節点をPart Affinity Fields(以下PAF)に基づいてつなぎ合わせていきます(図2)。辺の数(関節同士のつなぎ合わせ)を C 個として、PAF L は、

$$L = (L_1, L_2, \dots, L_C) \text{ where } L_c \in \mathbb{R}^{w \times h \times 2}, c \in \{1, 2, \dots, C\}$$

として与えられます。

こうしたアプローチによって人物検出の精度に依存せず、画像内の人数によって計算負荷が変化することのないmulti-personの姿勢推定が実現できます。二つのbranch

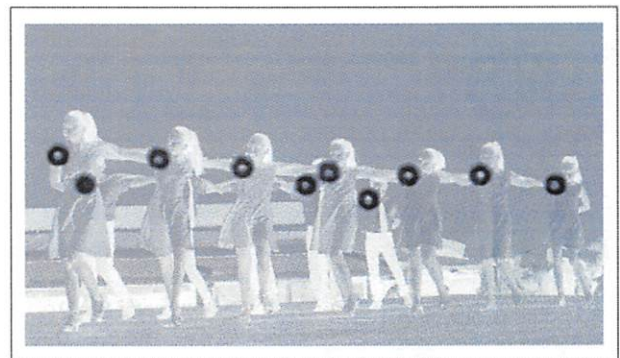


図1 Part Confidence Maps

†1 早稲田大学 大学院基幹理工学研究科 情報理工・情報通信専攻

†2 早稲田大学 基幹理工学部 情報理工学科

†3 早稲田大学 基幹理工学部 情報通信学科

"OpenPose" by Kentaro Inada (Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo), Ryohei Osawa (School of Computer Science and Engineering, Waseda University, Tokyo) and Atsuya Yamakawa (School of Communications and Computer Engineering, Waseda University, Tokyo)

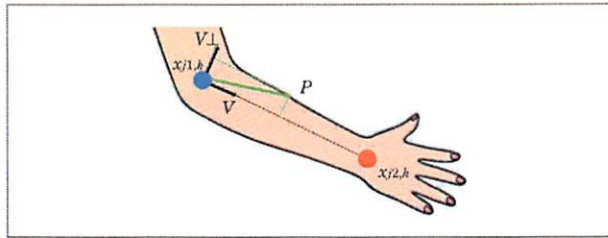


図2 Part Affinity Fields

(PCMとPAF)をまとめたネットワークを一つのstageとし、そのstageを反復的に複数個連ねていくことでより推定精度が向上します。

2.2 Part Confidence Maps

まず、画像内の人物 k の関節点 j に対してPCM $S_{j,k}^*$ を生成します。ここで、人物 k の関節点 j の位置のgroundtruthを $x_{j,k} \in \mathbb{R}^2$ としたとき、画像内の位置 $p \in \mathbb{R}^2$ における $S_{j,k}^*$ は、

$$S_{j,k}^*(p) = \exp\left(-\frac{\|p - x_{j,k}\|_2^2}{\sigma^2}\right)$$

として与えられ、 σ で確率分布の分散を調整します。こうして関節毎に得られた $S_{j,k}^*(p)$ は重ね合わされ、その際に確率分布のピークは保てるようにmax operatorを用いて、

$$S_j^*(p) = \max_k S_{j,k}^*(p)$$

と表されます。

2.3 Part Affinity Field

PCMで求めた関節同士をつなぎ合わせていくためのfieldがPAFです。画像内の位置 $p \in \mathbb{R}^2$ 、関節のつながり(limb) c におけるPAF $L_{c,k}^*$ を、

$$L_{c,k}^*(p) = \begin{cases} v & \text{if } p \text{ on limb } c, k \\ 0 & \text{otherwise.} \end{cases}$$

$$0 \leq v \cdot (p - x_{j1,k}) \leq l_{c,k} \text{ and } |v \perp \cdot (p - x_{j1,k})| \leq \sigma_l$$

とすることで、limb上、あるいはlimbからある範囲 σ_l に存在するピクセルには、2次元ベクトル v が与えられます。そしてlimb毎に得られたPAFはaverage operatorを用いて、

$$L_c^*(p) = \frac{1}{n_c(p)} \sum_k L_{c,k}^*(p)$$

と表されます。

2.4 マッチング

こうして得られたPCMとPAFを組み合わせて関節をつなぎ合わせていきます。しかし、単純に順番に関節同士をつないでみようとすると組み合わせ爆発により膨大な計算時間がかかってしまいます。そこで、まずつなぎ合わせる事のない関節間の接続関係は考えないことで計算量を大幅に削ります(図3(c))。また、関節同士のマッチングを分割していき、二部グラフのマッチングとして考えること

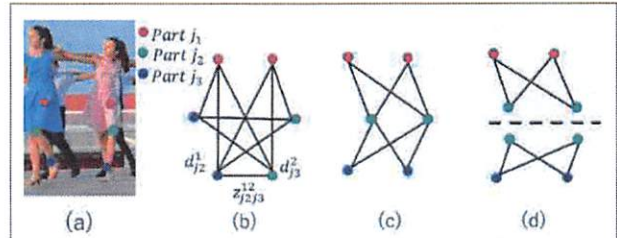


図3 マッチングの様子

で問題を細分化していきます(図3(d))。同じ関節同士が繋がることがないことを根拠とした工夫です。最終的にこれを欲張り法によってつなぎ合わせていくことで計算時間を大幅に短縮し、高速なマッチングを実現しています。

3. 導入方法

OpenPoseには、CUDA、cuDNN、Caffe、OpenCVのインストールが必要となります。CUDA 8.0とcuDNN 5.1、もしくはCUDA 10.0とcuDNN 7.5の組み合わせを推奨されています。各OSのインストール方法は、OpenPoseのGitHub³⁾に詳細が掲載されています。

4. 機能

4.1. モデル

上述したように、OpenPoseは人間の関節などの特徴点を検出し、リアルタイムに人物の姿勢を推定するシステムですが、モデルによって検出できる特徴点の数が異なります。特に、人物の姿勢推定においてはCOCOモデルとBODY_25モデルの二つがよく使われています。COCOモデルは目や鼻、肘、膝など18点の特徴点を検出します。一方でBODY_25モデルはCOCOモデルで検出可能な18点に加えて、つま先やかかとなどの位置を推定することが可能です。また、この二つのモデルに加えて15点の特徴点のみを検出するMP IIモデルもあります。OpenPoseの適用前後の様子を図4に示します。

さらに、顔と手に関しては、画像内にはっきり写っていれば、より多くの特徴点を検出可能なモデルも存在します。顔は輪郭や口などの特徴点を細かく検出し、70個の特徴点の位置を推定できます。また、手の検出では各指の関節や手のひらなどの特徴点を推定し、左右合わせて42点の特徴点の情報を得ることが可能になります。

上記のモデルは複数人が写っている映像に対しても適用できます。それに対して写っているのが1人のみの映像に限られてはしましますが、複数のカメラ映像から3次元の特徴点検出を行うこともできます。

4.2. 使用オプション

OpenPoseを実行する際にはオプションをつけることによってさまざまな機能を使うことができます。使用モデルに関してはデフォルトではBODY_25モデルが指定されますが、COCOモデル等を使用したい場合にはモデル指定オ

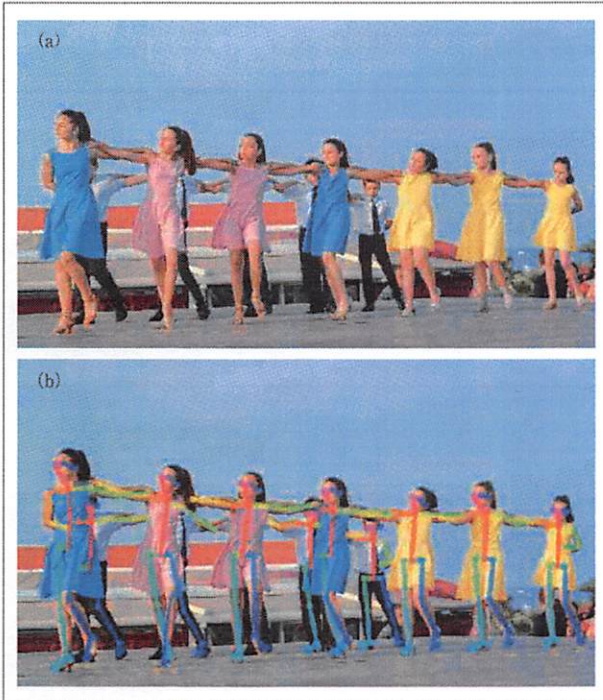


図4 (a) OpenPose適用前、(b) OpenPose適用後

クションをつけることによりモデルを変更することができます。また、それらのモデルとは別に“-face”や“-hand”といったオプションを実行時につけることにより、それぞれ顔や手の細かい検出が可能となります。

また、OpenPoseを実行するには入力画像であるのか動画であるのかを指定する必要があり、画像や動画があるフォルダも指定します。これによって、画像内もしくは動画の各フレーム内の人物に対して特徴点の検出を行います。また、WebカメラやIPカメラの映像を入力することも可能で、リアルタイムの映像に対しても特徴点検出ができます。

出力結果に関しては、デフォルト設定では何も保存されません。そのため、出力結果を保存したい場合にはオプションをつける必要があります。例えば、特徴点の座標情報をJSONファイルやXMLファイル等に出力して保存することができます。それに加えて、入力の画像や動画に、検出した特徴点とそれを結びつけた骨格を描画した画像もしくは動画を保存することも可能になります。

5. 用途

OpenPoseは研究等の非商用利用であれば無料で使うことができます。また、商用利用の場合はスポーツ分野では使用できませんが、非商用利用の場合はスポーツを含めさまざまな分野で活用することができます。

例えば、スポーツ分野では体にマーカ等を装着して関節等のデータを取得したりしていますが、OpenPoseを用いれば装置を装着する必要なくデータを得ることができます。そのため、得られた情報から選手の動きを分析するこ

とができ、その分析結果からトレーニングに応用したりすることもできるかもしれません。また、優秀な選手と動作を比較することにより、どこが違うのか、どこを修正すれば良いのかなどが座標情報から定量的にわかる可能性もあります。

また、何かの作業をしている人に対して姿勢推定を行うことにより、得られた情報から疲労度等を推定するアルゴリズムがあれば各人物の疲労度や体調などがわかるかもしれません。これが実現できれば、映像を撮っておくだけで、各人物の体調を推測し休憩の最適なタイミングを自動的に算出することができるようになる可能性もあります。

さらに、寝たきりの生活をしている人やベッドにいる赤ちゃんの映像からOpenPoseを用いて関節等の座標を抽出し、その座標情報からどのような動作を行っているかを特定します。そうすれば、何か危険な行動を行った際には介護をする人や子供の親などにその旨を伝え、より素早い対応の手助けとなるかもしれません。

6. むすび

スポーツの分野だけでなく、さまざまな分野で人物の動作解析はこれからの社会において必要になっていくと思います。しかし、映像から人物解析はまだ研究が進んでいない分野はあります。これから取り組むべき課題はたくさんあると思います。

(2019年6月26日受付)

【文献】

- 1) Hawk-Eye Innovations Ltd, Hawk-Eye, <https://www.hawkeyeinnovations.com/>
- 2) Zhe, Cao, T. Simon, S-E Wei, Y. Sheikh: "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 121, pp.1302-1310 (July 2017)
- 3) GitHub: "Openpose", <https://github.com/CMU-Perceptual-Computing-Lab/openpose> (2019年7月11日アクセス)



稲田健太郎 現在、早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻在学中。



大澤 遼平 現在、早稲田大学基幹理工学部情報理工学科在学中。



山川 敦也 現在、早稲田大学基幹理工学部情報通信学科在学中。