

姿勢推定可能な画像フィルタによるプライバシー保護

山本呼人¹ 酒澤茂之¹ 田坂和之²

概要: 画像認識技術を用いた遠隔コーチングによるスポーツ指導が注目されている。利用者の観点からは、インターネットへ自分の容姿の画像が送信されることへの懸念があり、映像をぼかすなどの抽象化処理を行いたいニーズがある。一方で、画像認識の観点からは、抽象化された画像から姿勢を推定することの困難さがある。本研究では、プライバシー保護のための画像フィルタリングと、姿勢推定精度を両立させるため、学習データセットの加工による再学習を行った結果について報告する。

キーワード: プライバシー保護, 姿勢推定, 深層学習

Image Filtering for Privacy Protection Maintaining Pose Estimation Accuracy

YOHITO YAMAMOTO^{†1} SHIGEYUKI SAKAZAWA^{†1}
KAZUYUKI TASAKA^{†2}

Abstract: A remote coaching for various sport activities is a promising application of image recognition technique. Considering a user acceptability, it is desirable to blur the image to be sent over the Internet. However, such image blurring may cause accuracy degradation of pose estimation which is necessary to instruct the user. In this paper, we report a result of retraining of a deep learning model for pose estimation with additional blurred image data set, which is a filtered version of the MS-COCO data set.

Keywords: Privacy protection, Pose estimation, Deep learning

1. はじめに

昨今 ICT 技術の発達によりインターネットを経由し様々なサービスを遠隔で受けることができる。その1つに遠隔でスポーツ指導を受けることができるスマートコーチが挙げられる。しかしそのサービスを受けるには高いコストがかかってしまう。そこでコーチングを人間ではなく AI が行うことでユーザがより低コストで気軽に利用できるサービスになると考える。現時点ではユーザは自分で映像を撮影し、ネットワークを通じてクラウド上のシステムに映像を送信する。概要図を図 1-1 として示す。

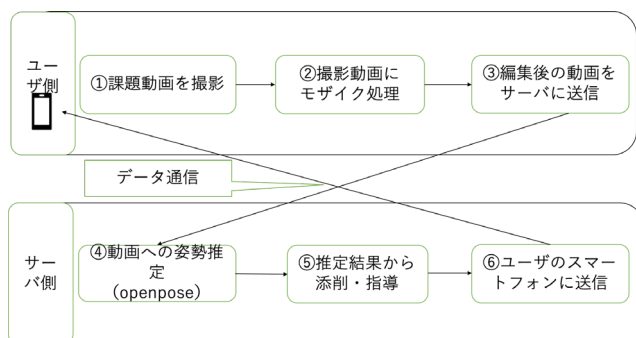


図 1-1 システム概要

しかし一般のユーザの視点からすれば自宅や自分の顔、体型の映った映像をネットワーク上に晒すことに対して恐怖感を抱く人も存在する。またユーザは手軽にサービスを受けることが目的であるため、普段使用しているスマートフォンやタブレットから映像を撮影するため、処理に対して計算量の制約がある。

そこで本論文では、利用者が撮影した映像に対して低演算量のフィルタをかけることでプライバシーの保護をし、そのフィルタ処理をした映像に対しても姿勢推定が可能なディープラーニングの学習モデルを作成し、その精度を測るものである。

2. 画像の加工と姿勢推定への影響

この章ではシステムに使用するフィルタの調査及び姿勢推定技術の紹介をする。

(1) 物理的フィルタの探索

前章でも記述した通り、映像に対してのフィルタ処理を行ったのちに姿勢推定を行う。そこでまず、カメラをフィルタで覆い撮影した画像を調査した。今回は手軽に用意できるセロハンテープ、養生テープを枚数を変化させながら

¹ 大阪工業大学
Osaka Institute of Technology
² KDDI 総合研究所
KDDI Research, Inc.

撮影した。図 2-1 はセロハンテープ 1 枚、養生テープ 1 枚でカメラを覆い撮影した画像である。

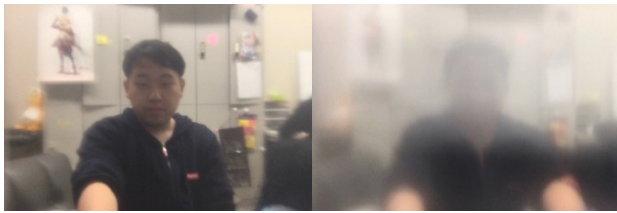


図 2-1 撮影画像（左：セロハンテープ、右：養生テープ）

(2) 画像認識への影響

本研究の実験では画像認識によって姿勢推定をする。そこで今回は“openpose”と“yolo”の 2 つの画像認識を使用した。

openpose は画像及び映像内の人物の各関節のデータを二次元の座標データとして出力する技術である。図 2-2 は openpose で検出可能な関節データを簡易的に表示したものである。

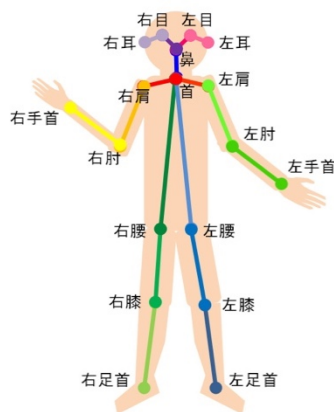


図 2-2 関節の簡易図[1]

次に yolo について説明する。yolo はリアルタイムで画像及び映像内の物体を判別、領域推定をする技術である[2]。この 2 つの技術を使用し、図 2-1 の画像に対して画像認識を行った。図 2-2、2-3 は識別画像を表示し、表 2-1 はセロハンテープと養生テープの枚数を変更して認識を行った比較結果をまとめたものである。

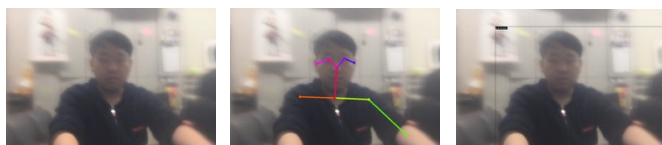


図 2-2 セロハンテープの結果（左から元画像、openpose、yolo）

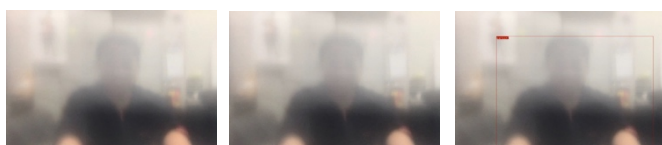


図 2-3 養生テープの結果（左から元画像、openpose、yolo）

表 2-1 画像の認識度合い

使用フィルタ/比較	openpose	yolo
セロハン 1 枚	○	○
セロハン 2 枚	○	○
セロハン 3 枚	○	○
養生 1 枚	×	○
養生 2 枚	×	○

表 2-1 より、yolo はぼかしの具合が強くなっても“人間”を判別できる。しかし openpose はセロハンテープでは推論可能だが養生テープでは推論ができない。一方で、ユーザの受容性の観点からは、養生テープによるぼかしの方が好ましい。そこで、本研究では養生テープを利用した際に、OpenPose で姿勢推定可能とすることを目標とする。

3. 画像加工に対応した姿勢推定モデルの構築

3.1 画像加工のソフトウェア化

2 章の結果より物理的フィルタは養生テープが最適となった。本研究では openpose の再学習を行うため再学習用のデータセットを作成する必要がある。そこでまずは“photoshop”を使用し擬似的に画像にフィルタ加工を施した。結果を図 3-1 として示す。

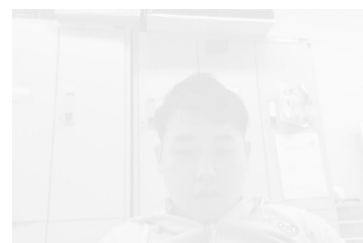


図 3-1 photoshop によるフィルタ加工

この加工では彩度を-91 に明度を+90 に変更した。

次に OpenCV でのフィルタ加工について述べる。OpenCV では画像透過でフィルタ処理を行った。元画像を 70% 真っ白の画像を 30% の割合で合成させた。その結果を図 3-2 として示す。OpenCV でのフィルタ処理は“MS-COCO”の画像を元画像とする。“MS-COCO”については次節で説明する。

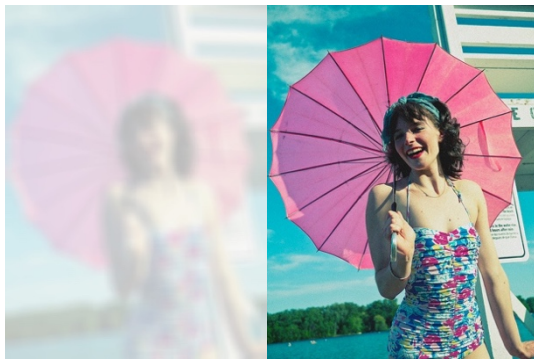


図 3-2 OpenCV によるフィルタ加工

3.2 姿勢推定モデルの学習

本章では openpose の姿勢推定に使用する学習モデルの再学習についてデータセット、再学習の実行の順に述べる。

(1) MS-COCO について

本実験で使用する学習用データセットの“MS-COCO”は MicroSoft が提供するデータセットで、データ数は学習用データが 118,287 個、バリデーション用データが 5,000 個それぞれに対してのマスクや関節情報などのアノテーションが用意されている。

(2) 画像加工後の学習用データの準備

前節で述べた学習用データ、バリデーション用データ両方に対して 2 章で述べた OpenCV でのフィルタ処理を施した画像を再学習用のデータセットとする。

(3) 学習の実行

本実験で使った openPose のソースコードは githab[3]を参照した。学習時のイテレーション数は各学習の結果とともに示す。

4. 実験結果

この章では 3 章で生成した再学習用のデータセットで学習をさせた結果を述べる。

4.1 データセットでの評価

まずフィルタ処理をする前のデータセットで学習したモデルによる推論結果を図 4-1 として示す。使用するモデルは 120,000 イテレーションのものとする。



図 4-1 推論結果 (左：ぼかしあり、右：ぼかし無し)

ぼかし無し、ありともに推論は可能になっているが、ぼかしありの推論では目の位置や腰の推論にズレが生じていることがわかる。

次にぼかしありの画像データとぼかし無しの画像データを 39,179 枚ずつ混合させたデータセットで学習させた結果を図 4-2 として示す。この学習ではイテレーション数を 2,000,000 とする。

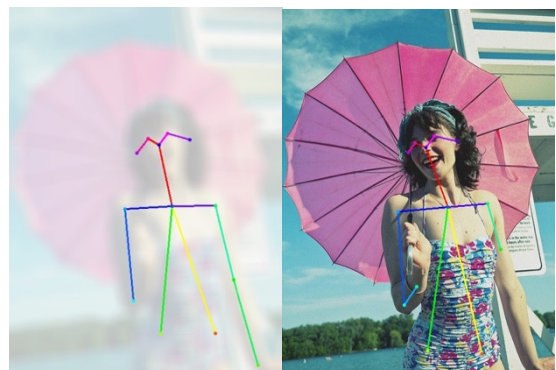


図 4-2 推論結果 (左：ぼかしあり、右：ぼかし無し)

結果よりそれぞれ比較すると、ぼかしありの画像は右手首の推論が誤推論になっており、ぼかし無しの推論は左手の推論が完璧にできていない。

次に先ほど述べた混合のデータセットのぼかしありとぼかし無しのデータの比率を 1:2 にしたデータセットでの推論結果を図 4-3 として示す。

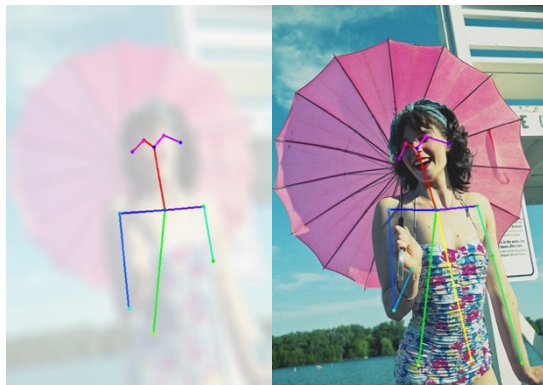


図 4-3 推論結果 (左: ぼかしあり、右: ぼかし無し)

図 4-2 と比較すると、ぼかしあり画像の推論は左手、左胴体が不十分であるため精度が落ちているとわかる。ぼかし無し画像の推論は左手の精度が上がっている。

ここでぼかしあり、ぼかし無し画像の枚数を同数で学習させた時の $\text{loss}, \text{paf}, \text{heat}$ を表 4-1、ぼかしありとぼかし無しの画像を 1:2 の割合に変更した学習時の $\text{loss}, \text{paf}, \text{heat}$ の値を表 4-2 として示す。

表 4-1 数値結果 (枚数比 1:1)

イテレーション数	main/loss	mian/paf	main/heat
20	0.41493106	0.18135078	0.23358023
200000	0.0185049	0.01052695	0.01900773

表 4-2 数値結果 (枚数比 1:2)

イテレーション数	main/loss	mian/paf	main/heat
20	0.49021322	0.18512271	0.3050905
200000	0.02104633	0.01292852	0.00811781

この結果より学習回数を繰り返すことで精度が向上していることが分かる。また枚数比を変化させても図 4-3 と図 4-4 の推論結果に大きな誤差がないことが数値でも見ることができる。

4.2 物理的フィルタ加工画像での評価

前節では OpenCV でのフィルタ画像を使用し学習の評価をしたが、この節では 2 章述べた養生テープによって加工した画像を使用して評価する。

まず、ぼかしありとぼかし無しの画像を同じ枚数で学習させたデータセットでの学習結果を示す。

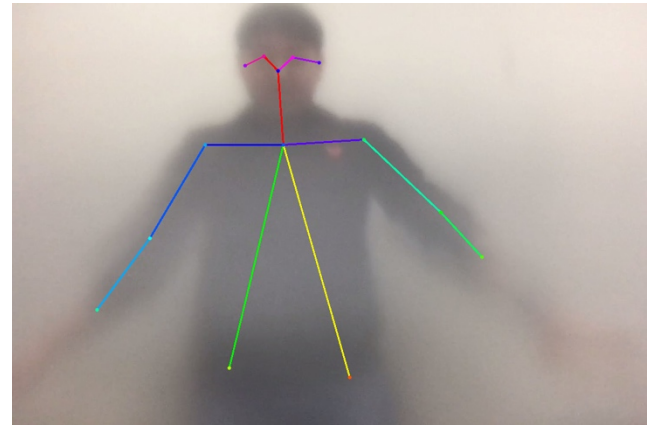


図 4-4 推論結果

次にデータの枚数を 1:2 の割合で学習させた結果を図 4-5 として示す。

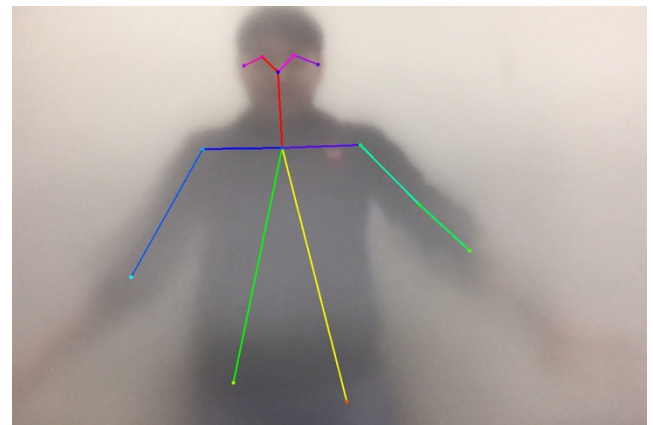


図 4-5 推論結果

図 4-4 と 4-5、テスト画像を比較すると両腕の推論の精度がやや間違いが見られるが胴体部分、顔の特徴点は総合的に正しく推論されている。

4.3 結論

全ての結果画像からデータセットを混合して学習をしているため学習量が不足していることがわかる。

また、図 4-4、図 4-5 から物理的フィルタである養生テープでも精度は 100%までもいかないが推論することができている。

5. むすび

今回の研究ではユーザが自分のスマートフォン及びタブレットから撮影及び加工した映像をネットワークを通じてサーバ側に送信し、その映像に対して姿勢推定をするというものである。そこで映像に対する加工として、養生テープ、セロハンテープのよる物理的フィルタ、OpenCV による加工を使用した。OpenCV での画像透過による加工画像を学習データとして、openpose の再学習及び推論を行った。再学習では、フィルタ処理をした画像と元画像を同数、枚数比率を 1:2 の 2 つの形式で行った。養生テープを使用した物理的フィルタでぼかした画像に対しても精度が完璧とまではいかないが推論することができている。

これらの結果を踏まえ、今後の課題としては学習のイテレーション数を増やすこと、加工の強弱を複数用意しユーザによって異なる可能性のある映像加工に対応させることだと考える。

参考文献

- [1] 川島早紀子他, “OpenPose を用いた回転不変動作の姿勢推定”, https://www.ams.giti.waseda.ac.jp/data/pdf-files/2018 ITEw_13D3_kawashima.pdf, 映像情報メディア学会 冬季大会 13D-3, 2018
- [2] Joseph Redmon 他, “You Only Look Once: Unified, Real-Time Object Detection”, “arXiv:1506.02640”, 2016
- [3] Zhe Cao and Tomas Simon and Shih-En Wei and Yaser Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). p. 7291-7299, 2017.