広角映像の歪みに頑健な注目点検出手法の開発と人物動作解析への応用

Development of Robust Keypoint Detector for Distorted Wide-Angle Images and Application to Human Motion Analysis

東京大学, 都産技研	三木　大輔	Daisuke　MIKI	Member
都産技研	阿部　真也	Shinya　Abe	Non-member
東京大学	陳　実	Shi　CHEN	Student Menber
東京大学	出町　和之	Kazuyuki　Demachi	Member

Abstract

Tracking human motion from video sequences is a notable technique that is used to detect anomalies in individual human behavior. Several commercially available motion capture devices are based on the use of depth cameras. However, there are a couple of problems with the use of a depth camera. Firstly, a complicated camera system is required, and secondly, the optical field of view is limited. To overcome these problems, we need a technique that can recognize human motion from wide-angle images. In this study, we will devise a method for tracking human motion that is robust toward the distortion of wide-angle images. The main contribution of this study is the development of a methodology that can automatically estimate the transformation parameters that are required to improve the accuracy of human motion recognition. We propose a new architecture of a multi-layered convolutional neural network that can estimate the location of human joints in images and transformation parameters simultaneously. We confirmed its applicability to human motion analysis by comparing the results of the application for both natural and unnatural human motion data.

Keywords: Human motion recognition, Wide-angle image, Convolutional neural network, Worker Safety, Video surveillance

1．はじめに

近年，多くの保全作業現場において安全衛生管理体制の確立が求められている．厚生労働省の報告[1]では，労働災害による死傷者数は年間約 900 人に上り，その多くが墜落・転落事故，はさまれ・巻き込まれ事故などによるものである．そのような事故の発生を低減するために，作業員に対して過去の事故事例などについて注意喚起を行い，再発防止を図る取組みが行われている．しかし，作業員の従事する業務や環境は多岐に渡るため，このような受動的対策は必ずしも有効とは限らない．一方で事故を未然に防ぐための能動的対策として，作業現場に潜む危険を把握し，取り除くことが有効である．具体的には，人物の危険な動作や，通常動作からの逸脱を検出・評価し，労働環境を改善する取り組みが考えられる．このような人物の動作解析に有効な方法の一つにモーションキャプチャ機器の利用が挙げられる．現在，市場に広く普及しているモーションキャプチャ機器は，RGB-D カメラなどによって撮像された距離画像を解析[2]することで，人物の動作を認識する．しかし，RGB-D カメラに用いられるステレオカメラや赤外線カメラなどを利用した方法は，特殊な撮像機器を必要とすることや，画角が狭く限られるといった課題があった．特に画角が狭く限られることは広範囲または近距離に存在する人物の姿勢を認識する上で不利であり，作業現場の監視等の用途には適さなかった．そこで本研究では，保全作業現場での労働者安全を目的とした広角カメラ映像からの人物動作の解析手法を開発する．

提案手法では，まず広角カメラ映像内の人物の関節位置を多層の畳み込みニューラルネットワーク

〒113-00334 東京都江東区青海 2-4-10
地方独立行政法人東京都立産業技術研究センター
開発本部開発第三部情報技術グループ
E-mail: miki.daisuke@iri-tokyo.jp

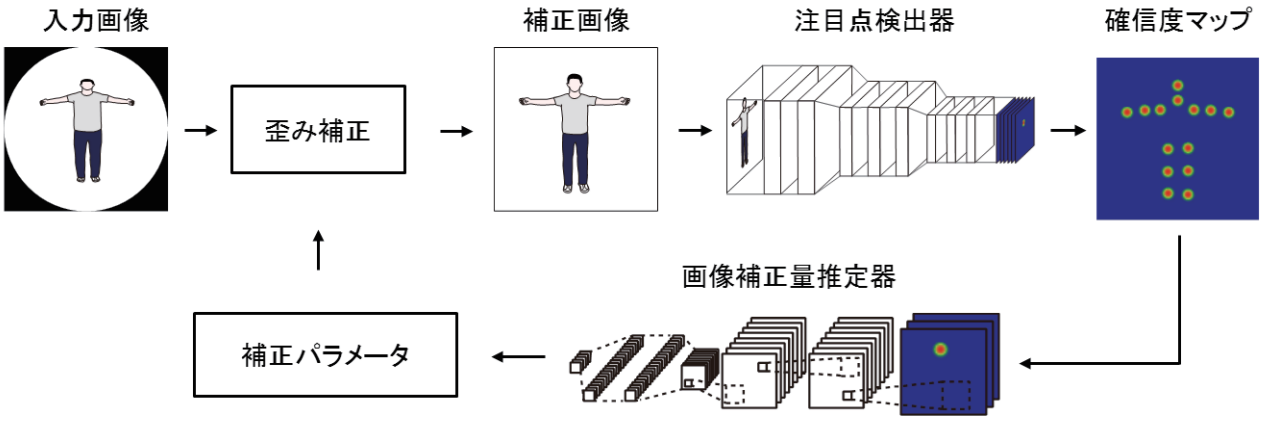


図１　広角映像からの注目点検出手法の概要

(Convolutional Neural Network, CNN)を用いて認識する．ここで，広角画像の歪みに頑健とするためCNNの構造を工夫し，映像の補正に必要なパラメータを推定する画像補正量推定器を設けた．さらに，得られた画像中の注目点情報を利用し，実空間上での人物の 3 次元的な姿勢および位置を推定することで，人物動作の特徴を抽出した．

本稿では，注目点の検出および画像補正量推定のための CNN の構造およびそのパラメータの最適化方法について説明した後，得られた注目点情報を利用して人物の動作を解析する方法およびそれらの評価実験と結果について述べる．

２．画像の歪みに頑健な注目点検出

本研究で提案する画像の歪みに頑健な注目点検出手法の概要を図１に示す．提案手法では画像中の注目点位置を推定する注目点位置推定器と，画像の歪みに頑健とするための画像補正量推定器から成る．

2.1 注目点検出器

画像中の人物における注目点位置の推定のためのCNNには，広角画像（256×256 px）を入力し，画像の畳み込みおよび，プーリングを繰り返すことで，出力層では注目点位置を示す確信度マップ（32×32 px）を回帰する構造とした．ここで，畳み込み層をより多層とすることが認識精度の向上のために効果的であるが，畳み込み層をある程度以上に多層とすると，誤差逆伝搬が上手く行われずに，パラメータの学習が停滞すること（勾配消失）[3]が問題となる．そこで，ネットワーク中で段階的に誤差の計算を行うような構造[4]を採用することで，CNN を多層にできるようにした．

CNN の学習のため，学習用データとして MPII Human pose dataset [xxx]を利用した．データセットに含まれるラベルをもとに，j番目の関節位置を示す確信度マップを

$$\mathbf{S}_j^* = \exp\left(-\frac{\|\mathbf{p} - x_j^*\|^2}{\sigma}\right) \tag{1}$$

とした．ここで，x_j^*は人物の関節位置が存在する座標の真値を示し，$\mathbf{p}$は確信度マップにおける各画素の座標である．パラメータの最適化では，注目点検出器で推定された CNN の出力 $\mathbf{S}_j(\mathbf{p})$および真値 $\mathbf{S}_j^*(\mathbf{p})$との間における誤差

$$\mathcal{L}_1 = \sum_j \|\mathbf{S}_j(\mathbf{p}) - \mathbf{S}_j^*(\mathbf{p})\|_2 \tag{2}$$

を最小化するようにネットワークを最適化した．

2.2 画像補正量推定器

画像補正量推定器は，注目点推定器で得られた確信度マップを入力とし，広角カメラの焦点距離 f及び水平方向の回転量θ_h，垂直方向の回転量 θ_v，画像の原点を中心とした画像平面内での回転量 θ_hの 4 次元の値を出力する構造とした．画像補正量推定器の学習では，まず，図２に示すような人物の画像と同様に 3 次元コンピュータグラフィックス（3D computer graphics, 3DCG）によって，焦点距離 $0.84\ \text{mm} \le f \le 1.41\ \text{mm}$および，光軸を中心に水平方向 $-45° \le \theta_h \le 45°$，垂直方向 $-20° \le$

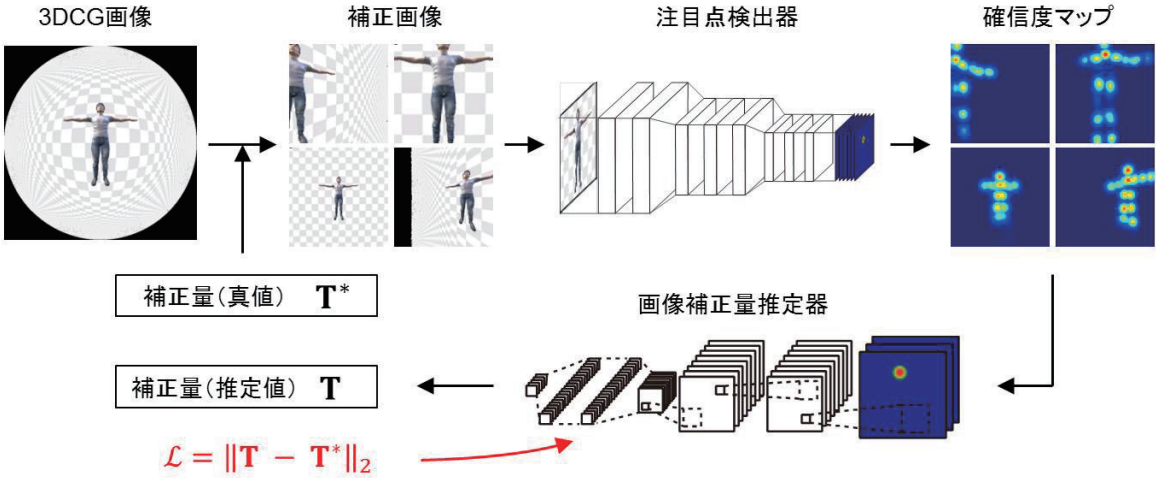


図２　画像補正量推定のための CNN の学習方法

$\theta_r \leq 20°$，画像の原点を中心とした画像平面内で $-30° \leq \theta_r \leq 30°$の範囲で変形させた画像を 1 万枚用意した．次に，先の注目点位置推定器にこれらの画像を入力し，出力として確信度マップを得た．さらにこの確信度マップを入力とし，画像の変形に利用したパラメータを推定すべき真値 $\mathbf{T}^*$とした．最後に，画像補正量推定器における CNN の出力 $\mathbf{T}$および真値 $\mathbf{T}^*$を用いた損失関数

$$\mathcal{L}_2 = \|\mathbf{T} - \mathbf{T}^*\|_2 \qquad (3)$$

を最小化するように CNN のパラメータの最適化を行った．

３．人物動作解析への応用

3.1 人物姿勢情報の再構築

注目点情報から人物の 3 次元姿勢の推定では，まず，あらかじめ用意された 1 万通りの 3 次元姿勢モデルに対して t-SNE による次元削減を行った後，EM アルゴリズムによるクラスタリングを行い，平均姿勢 $\mathbf{Y}_n^*$および正規直交行列 $\mathbf{e}$を求めた．次に，平均姿勢 $\mathbf{Y}_n^*$を 2 次元平面への射影

$$\mathbf{Y}_n(a) = \mathbf{\Pi}(\mathbf{T})(\mathbf{Y}_n^* + a \cdot \boldsymbol{e}) \qquad (4)$$

を求めた．ここで，$\mathbf{\Pi}(\mathbf{T})$は画像の歪みを考慮した射影行列である．さらに注目点検出器によって推定された人物の画像中での 2 次元姿勢 $\mathbf{y}$を用いて得られる損失関数

$$E(\mathbf{y}, \mathbf{Y}_n(a)) = \sum_{n \in N} \|\mathbf{y} - \mathbf{Y}_n(a)\|_2^2 + \|\sigma \cdot a\|_2^2 \qquad (5)$$

を最小化する正規直交行列 aを求め，対応する 3 次元での人物姿勢

$$\mathbf{Y} = s\mathbf{\Pi}(\mathbf{T})(\mathbf{Y}_n^* + a \cdot \mathbf{e}) \qquad (6)$$

を得た．ここで，sはスケーリング係数である．

3.2 人物位置の推定

図３に人物位置の推定手法を示す．先の注目点検出器によって推定された画像中における注目点を入力とし，カメラ空間上での人物の位置を推定する全結合ニューラルネットワーク（neural network, NN）へと入力する．この NN ではカメラから人物の距離を推定し，この距離情報と広角カメラ映像の射影方式を利用して実空間上の 3 次元位置を推定する．これにより得られた 2 次元姿勢と人物位置を利用して，人物の 3 次元姿勢が推定される．

距離推定 NN の学習では，カメラ空間上の位置 $\mathbf{P}^*$に投影された 3DCG 人物モデル画像を生成し，学習用画像とした．注目点推定 CNN および距離推定 NN を用いて 2 次元人物姿勢と人物位置を推定した後，人物位置 $\mathbf{P}$を推定し，$\mathbf{P}^*$および $\mathbf{P}$に関する損失関数

$$\mathcal{L}_3 = \|\mathbf{P} - \mathbf{P}^*\|_2 \qquad (7)$$

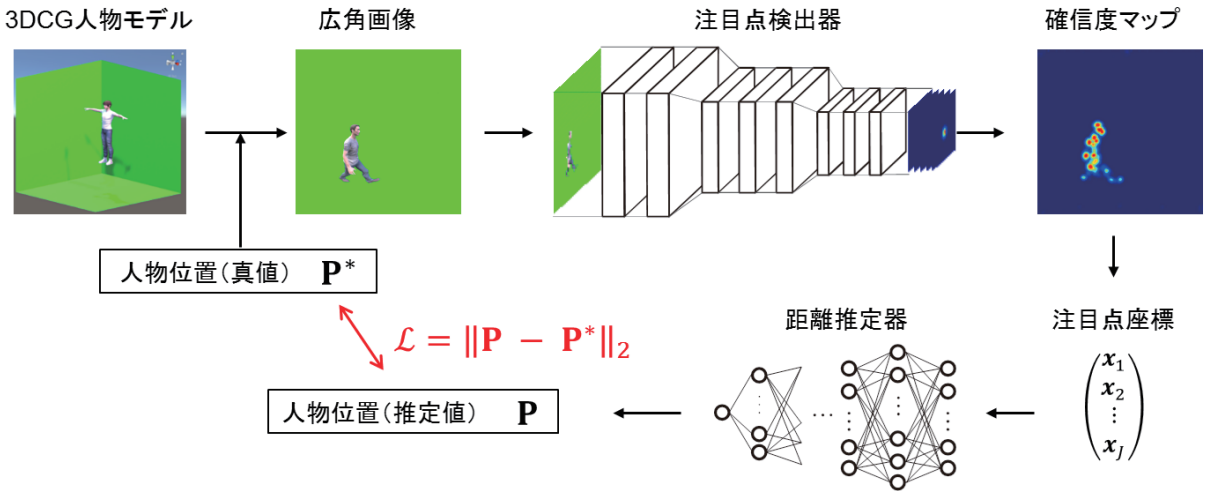


図３　人物位置推定のための NN の学習方法

を最小化するように距離推定 NN の学習を行った.

４．実験および考察

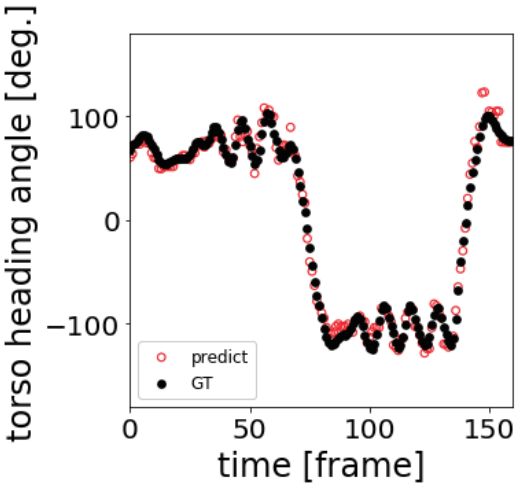
4.1 実験方法

各 CNN の学習および評価実験には GPU（Geforce GTX1070）を搭載した PC（Intel core i7-6700, 3.4GHz）を利用した．注目点位置推定 CNN の学習では，7 万回のパラメータの更新を行った. 同様に画像補正量推定 CNN および距離推定 NN の学習ではそれぞれ 1 万回のパラメータの更新を行った.

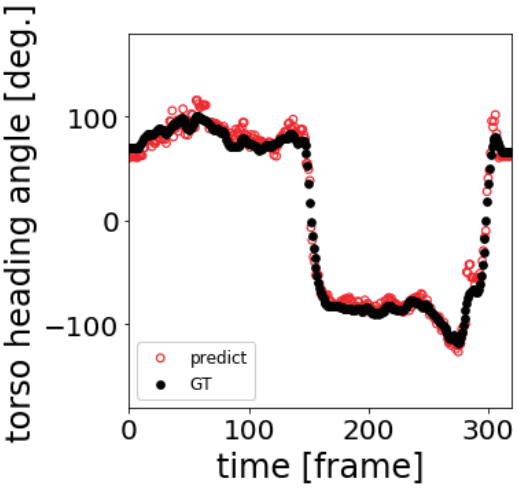
評価では 3DCG を用いて人物が歩行する広角映像を生成し，人物の姿勢および位置に関する真値と推定値の比較を行った．3DCG データには様々な体型をシミュレートできる人物の 3DCG モデル[6]を使用し，モーションキャプチャデータとして CMU Mocap データ[7]を利用した. 特に 3 次元動作の認識のため, CMU Mocap データから，自然な歩行動作（Subject 91, trial 2）および不自然な歩行動作（Subject 91, trial 18）の人物姿勢の認識およびそれらから得られる体の向き，視線方向，移動速度の推定を行い，真値と比較した.

4.2 実験結果

推定された注目点情報から 3 次元姿勢の復元および人物位置の推定を行い，人物動作の特徴を抽出した結果を図４，５，６に示す．自然な歩行動作および，不自然な歩行動作に関する特徴をそれぞれ可視化することが可能であり，真値と比較して良好な認識精度が確認された. 特に不自然な歩行動作において体の向きや視線方向の変動，移動速度の変動に対して特徴的な傾向を把握することが可能であり，真値に対して良好な精度で認識が可能であることが確認できた.

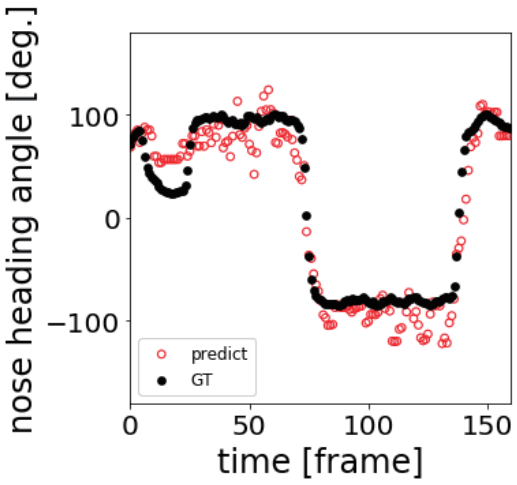


(a) CMU Mocap Dataset Subject 91, trial 2

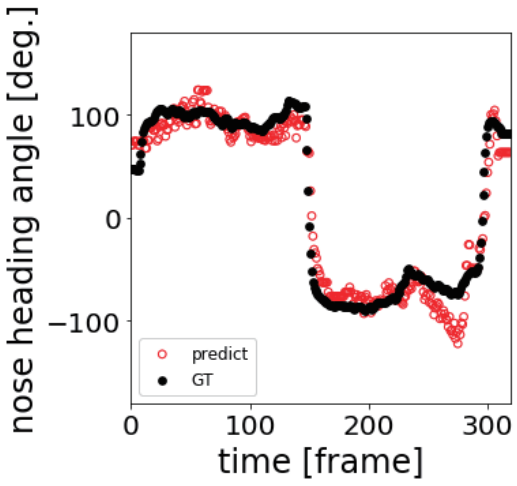


(b) CMU Mocap Dataset Subject 91, trial 18

図４　人物の体の向きの推定結果

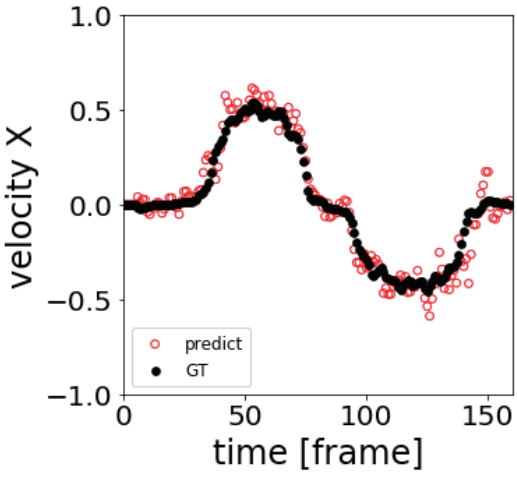


(a) CMU Mocap Dataset Subject 91, trial 2

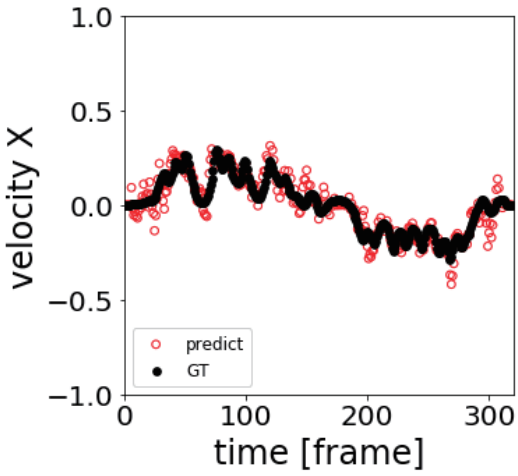


(b) CMU Mocap Dataset Subject 91, trial 18

図５ 視線方向の推定結果



(a) CMU Mocap Dataset Subject 91, trial 2



(b) CMU Mocap Dataset Subject 91, trial 18

図６ 移動速度の推定結果

５．まとめ

本研究では，広角カメラ映像の歪みに頑健な注目点の検出手法およびそれらを利用した人物動作の解析技術を開発した. 注目点検出器および, 画像補正量推定器をCNNで実装し，組み合わせることで画像の歪みに頑健な注目点の検出が可能であることを確認した．また，推定された 2 次元の注目点位置情報を利用することで，人物の 3 次元的な姿勢および位置の推定が可能であった．3DCGを利用した動作解析に関する実験では，広角での人物姿勢認識が可能であり，人物の自然および不自然な歩行動作に関する特徴を抽出することが可能であることを確認した．本手法を応用し，保全作業に従事する作業員の危険な動作や，通常動作からの逸脱を検出できれば，労働災害の事前防止につながる可能性がある. 今後はCNNおよび NN の構造を改善することで認識精度の向上や，動作データからの異常検知，行動認識等に適用できるように改良を行う.

参考文献

[1] 厚生労働省　労働災害発生状況 https://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei11/rousai-hassei/index.html

[2] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake, "Efficient human pose estimation from single depth images" IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(12), 2821–2840,

2013

[3] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult" IEEE Transactions on Neural Networks 5, 2, 157–166, 1994.

[4] S. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines", IEEE Conference on Computer Vision and Pattern Recognition, 4724–4732, 2016

[4] M. Andriluka, L. Pishchulin1, P. Gehler, B. Schiele "2D Human Pose Estimation: New Benchmark and State of the Art Analysis", IEEE Conference on Computer Vision and Pattern Recognition, 2014

[6] https://www.adobe.com/jp/products/fuse.html

[7] http://mocap.cs.cmu.edu/