3次元ベクトル場を導入した CNN による人体の 3次元姿勢推定

小松悠斗† 白木克俊† 平川翼† 山下隆義† 藤吉弘亘† †中部大学

E-mail: u1370@mprg.cs.chubu.ac.jp

1 はじめに

人体の姿勢推定は対象の関節位置を推定する問題であり、動作認識やモーションキャプチャ等に用いられるため、広く研究されている.

姿勢推定は2次元の姿勢を推定する手法と3次元の 姿勢を推定する手法がある. 2次元姿勢推定の場合, 姿 勢推定の手法としてトップダウン型の手法 [1][2][3] や ボトムアップ型の手法 [4] がある. トップダウン型の手 法は,人物を検知した後,姿勢推定を行う.ボトムアッ プ型の手法は,画像内の人物の関節位置を推定した後, 推定した関節位置を人物ごとにマッチングさせていく. トップダウン型の手法は,単一の人物に対する姿勢推 定が中心であるのに対して, ボトムアップ型の手法は, 複数の人物に対する姿勢推定が中心である. トップダ ウン型の手法には, 計算コストが高くなりやすいとい う問題があり、ボトムアップ型の手法には、関節位置の マッチングが必要であるため、トップダウン型の手法に 比べ精度が低くなるという問題がある. 近年, 高精度な 2次元の姿勢推定手法が提案されているが、モーション キャプチャなどに活用するためには、奥行きを含めた3 次元の姿勢推定が必要である. 3次元姿勢推定の場合, 単一の人物に対する姿勢推定の手法が広く研究されて おり、3次元姿勢を直接推定する手法 [5] や2次元姿勢 と 3 次元姿勢を分けて推定する手法 [6][7][8][9] がある. また, リアルタイムで姿勢推定を行う手法 [10][11] や, 複数の人物に対する姿勢推定の手法 [12] も研究されて いる. 3次元姿勢推定の問題として、屋内データセット での学習による汎化性の確保や隠れの存在する姿勢で は姿勢推定が困難であること等が挙げられる. 3次元 の姿勢推定手法として, Zhou 等は Weakly-supervised Approach[7] の手法を提案している. 本手法は、複数の ステージでの学習により汎化性を確保している. また, depth に対する損失関数として Geometric Loss を提案 し、これにより2次元データから3次元姿勢推定の学 習を行うことができる. しかしながら, オクルージョ ンがある部位の特徴を捉えることは難しいため、depth 推定精度が低下することがある.

そこで本研究では、2次元 RGB 画像から3次元ベクトル場を考慮した3次元姿勢推定を行うことで depth

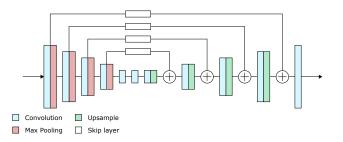


図1 砂時計モジュール

推定の高精度化を行う. 3次元ベクトル場は, 関節間の位置情報と方向情報を含んだ特徴表現である. 3次元ベクトル場を推定するために, 提案手法では3次元ベクトル場推定モジュールを導入する. 3次元ベクトル場推定モジュールは, 2次元姿勢推定モジュールによって抽出した2次元特徴マップおよびDepth回帰モジュールからの出力を3次元特徴マップに拡張する. 拡張した特徴マップに対して複数のスケールで畳み込み処理を行い, 3次元ベクトル場を推定する. このとき, 3次元ベクトル場推定モジュールからの逆伝播によって, 3次元姿勢推定の精度が向上することを示す.

2 関連手法

単一の人物に対する姿勢推定の手法として、2次元姿勢推定の手法ではHourglassNet[1],3次元姿勢推定の手法ではMehta等、Pavlakos等、Zhou等の手法[6][5][7]がある。以下にこれらの手法について説明する。

2.1 HourglassNet

本手法 [1] は、単一の RGB 画像から人物の 2 次元姿勢を推定する手法である。画像中の様々なスケールの特徴を抽出可能な砂時計モジュールを提案しており、砂時計モジュールを複数繋げた砂時計ネットワークを構築し2 次元姿勢推定を行っている。HourglasNet の砂時計モジュールを図 1 に示す。砂時計モジュールは畳み込み層と Max Pooling層によって最低解像度まで畳み込んだ後、アップサンプリングを行いながら Max Pooling前の各スケールの情報を skip layer を用いることで連結している。

2.2 Coarse-to-Fine Volumetric Prediction for Single-Image 3D Human Pose

本手法 [5] は、単一の RGB 画像から人物の 3 次元姿勢を推定する手法である。ネットワーク構造は HourglassNet をもとに構築しており、3 次元ヒートマップを推定することで、3 次元姿勢推定を行う。また、砂時計間で 2 次元ヒートマップを推定することにより、画像と 2 次元関節位置の両方からの情報を効率的に伝達している。損失関数は、推定関節位置に対して L2 ノルムの損失、推定 3 次元ヒートマップに平均二乗誤差を適用している。3 次元ヒートマップは、式 (1) の 3 次元ガウス分布によって生成する。

$$G_{i,j,k}(\boldsymbol{x}_{gt}^n) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-i)^2 + (y-j)^2 + (z-k)^2}{2\sigma^2}} \tag{1}$$

このとき、 $\boldsymbol{x}_{gt}^n = (x,y,z)$ は 3 次元正解間接位置であり、 $\sigma = 2$ で 3 次元ヒートマップを生成する.

2.3 Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision

本手法 [6] は、単一の RGB 画像から人物の 3 次元姿 勢を推定する手法である. 本手法のネットワーク構造 を図2に示す. ResNet をベースに構築しており,2次 元姿勢推定ネットワークで学習後、3次元姿勢推定ネッ トワークに転移学習している.また,本手法では3次元 姿勢推定ネットワークで補助タスクとして, 関節位置 の2次元ヒートマップを推定する. 推定されたヒート マップの最大値を2次元関節位置 K とし、関節位置の グローバル座標の導出に利用する. 本手法では, 運動学 に基づいて対象の関節 P だけではなく対応する 2 つの 関節 O1, O2 の関節位置も推定する. 推定した関節位置 をもとに、最終的な3次元関節位置 P_{fused} を推定する. P_{fused} および K からグローバル座標を式 (2) で導出す る. また,本手法は学習時 Corrective Skip Connections を追加し、 P_{sum} , $O1_{sum}$, $O2_{sum}$ を推定することで精度 を向上させている. Corrective Skip Connections は学 習後に削除する.

$$T = \frac{\sqrt{\sum_{i} \parallel P_{|xy|}^{i} - \bar{P}_{|xy|} \parallel^{2}}}{\sqrt{\sum_{i} \parallel K^{i} - \bar{K} \parallel^{2}}} \begin{pmatrix} \frac{\bar{K}_{|x|}}{\bar{K}_{|y|}} \\ f \end{pmatrix} - \begin{pmatrix} \frac{\bar{P}_{|x|}}{\bar{P}_{|y|}} \\ 0 \end{pmatrix} \quad (2)$$

 $P_{|xy|}$ は P_{fused} の x, y 部分である.

2.4 Weakly-Supervised Approach

本手法 [7] は、様々な姿勢の人物に対する 3 次元姿勢推定手法である。本手法は depth に対する損失関数として、人の骨の長さの比が一定であることをもとにした Geometric Loss を提案している。 Geometric Loss は計算コストが小さく、2 次元データから depth に対する学習を行うことができる。本手法のネットワーク構

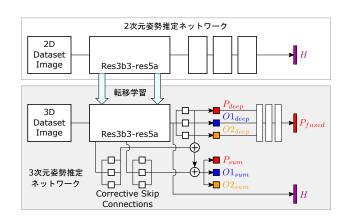


図 2 文献 [6] のネットワーク構造

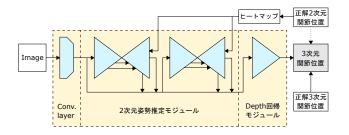


図 3 文献 [7] のネットワーク構造

造を図3に示す。HourglassNet をもとにした2次元姿勢推定モジュールから出力された特徴マップおよび中間の画像特徴,また2次元姿勢推定モジュールに入力される前の特徴マップを連結したものを depth 回帰モジュールの入力とする。その後,depth 回帰モジュールは入力をもとに depth を回帰推定することで,3次元姿勢推定を行う。

2.4.1 2次元姿勢推定モジュール

モジュール出力は J 個の低解像度ヒートマップであり,各マップ $\hat{Y}_{HM} \in \mathcal{R}^{H \times W}$ は,2 次元確率分布を表す.2 次元姿勢推定モジュールで予測された部位 \hat{Y}_{2D} はヒートマップ上のピーク位置である.2 次元姿勢推定モジュールの損失関数を式 (3) に示す.予測されたヒートマップ \hat{Y}_{HM} と正解座標 Y_{2D} から求めたヒートマップ $G(Y_{2D})$ との平均二乗誤差を使用する.ここで, $G(\cdot)$ はガウスカーネルである.

$$L_{2D}(\hat{Y}_{HM}, Y_{2D}) = \sum_{h}^{H} \sum_{w}^{W} (\hat{Y}_{HM}^{(h,w)} - G(Y_{2D})^{(h,w)})^{2}$$
 (3)

2.4.2 Depth 回帰モジュール

Depth 回帰モジュールは、複数の Residual ブロックと 1 つの全結合層から構成されている。 Depth 回帰モジュールの損失関数を式 (4) に示す。

$$L_{dep}(\hat{Y}_{HM}|I, Y_{2D}) = \begin{cases} \lambda_{reg} \|Y_{dep} - \hat{Y}_{dep}\|^2 & \text{if } I \in I_{13D} \\ \lambda_{geo} L_{geo}(\hat{Y}_{dep}|Y_{2D}) & \text{if } I \in I_{32D} \end{cases}$$
(4)

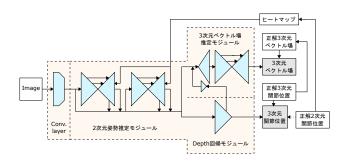


図 4 提案手法のネットワーク構造

 \hat{Y}_{dep} は予測 depth, Y_{dep} は正解 depth である。 λ_{reg} と λ_{geo} は対応する誤差の重みである。I は入力画像である。また $L_{geo}(\hat{Y}_{dep}|Y_{2D})$ は 2 次元データセットにおける depth 推定誤差であり,Geometric Loss と呼ぶ。Geometric Loss を式 (5) に示す。骨格グループi の集合を R_i とする。 l_e を推定された関節の e 番目の関節間の長さとし, $\bar{l_e}$ を正解データの e 番目の関節間の長さとする。

$$L_{geo}(\hat{Y}_{dep}|Y_{2D}) = \sum_{i} \frac{1}{|R_i|} \sum_{e \in R} (\frac{l_e}{l_e} - \bar{r}_i)^2$$
 (5)

$$\bar{r}_i = \frac{1}{|R_i|} \sum_{e \in R_i} \frac{l_e}{\bar{l}_e} \tag{6}$$

 \bar{r}_i は分散に用いる各 R_i の比率 l_e/\bar{l}_e の平均である.

2.4.3 学習方法

本手法の学習は3ステージに分けて行う.ステージ1は,2次元姿勢推定モジュールを2次元座標データで学習する.損失関数は,式(3)の損失関数のみを用いる.ステージ2,3は,両モジュールを2次元,3次元座標データで学習する.ステージ2は式(4)の3次元データセットの場合の損失関数を追加し,ステージ3はGeometric Lossである式(4)の2次元データセットの場合の損失関数を追加する.

3 提案手法

前章で述べた3次元姿勢推定手法は、オクルージョンを含む姿勢や画像内での関節間が非常に短くなっている姿勢等でdepth 推定精度が低下することがある.したがって、本研究では、3次元ベクトル場推定モジュールを従来手法に導入することで、各方向に対する関節間の特徴を抽出し、3次元における姿勢推定を強化することを目的とする.提案するネットワーク構造を図4に示す.ネットワーク構造は、従来手法のネットワーク構造をもとに構成し、3次元ベクトル場推定モジュールを追加している.2次元姿勢推定モジュールは、複数の

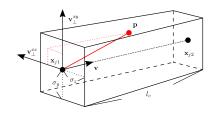


図 5 関節 \mathbf{x}_{j_1} , \mathbf{x}_{j_2} と空間上の点 \mathbf{p} の関係

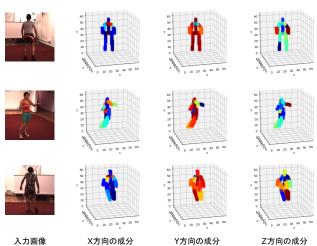


図 6 3次元ベクトル場の例

畳み込み層で入力画像から抽出された特徴マップをもとに2次元ヒートマップを推定する.2次元姿勢推定モジュールからの出力は、2次元姿勢推定モジュールの中間特徴および入力された特徴マップと連結され、Depth回帰モジュールに入力される.Depth回帰モジュールは2次元姿勢推定モジュールからの連結された特徴マップをもとにdepthを推定する.3次元ベクトル場推定モジュールは、2次元姿勢推定モジュールからの出力値およびdepth回帰モジュールの中間特徴から、3次元ベクトル場を推定する.

3.1 正解 3 次元ベクトル場の生成

提案する損失関数を適用するため,正解 3 次元ベクトル場を生成する。3 次元ベクトル場は,2 点の関節間の位置情報と方向情報を含んだ特徴表現である。正解 3 次元ベクトル場を生成する条件を式 (7) に示す。図 5 に示すように,点 \mathbf{p} が関節点 \mathbf{x}_{j_1} から関節点 \mathbf{x}_{j_2} を基準とする関節間にある場合, $\mathbf{F}_c^*(\mathbf{p})$ の値は関節 j_1 から関節 j_2 への単位ベクトル \mathbf{v} になり,それ以外の場合では 0 となる。

$$\mathbf{L}_{c}^{*}(\mathbf{p}) = \begin{cases} \mathbf{v} & \text{if } \mathbf{p} \text{ on limb } c \\ 0 & \text{otherwise} \end{cases}$$
 (7)

このとき、 $\mathbf{v} = (\mathbf{x}_{j_2} - \mathbf{x}_{j_1})/||\mathbf{x}_{j_2} - \mathbf{x}_{j_1}||_2$ は始点 \mathbf{x}_{j_1} から終点 \mathbf{x}_{j_2} の単位ベクトルを表す.また,単位ベクト

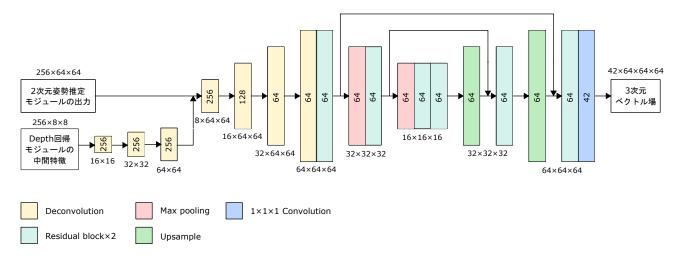


図 7 3 次元ベクトル場推定モジュールの構造

ル v を生成する条件を式 (8)(9)(10) に示す.

$$0 \le \mathbf{v} \cdot (\mathbf{p} - \mathbf{x}_{j_1}) \le l_c \tag{8}$$

$$|\mathbf{v}_{\perp}^{xy} \cdot (\mathbf{p} - \mathbf{x}_{j_1})| \le \sigma_y \tag{9}$$

$$|\mathbf{v}_{\perp}^{xz} \cdot (\mathbf{p} - \mathbf{x}_{j_1})| \le \sigma_z \tag{10}$$

ここで、 σ_y , σ_z は y 方向および z 方向のピクセル単位の 3 次元ベクトル場の幅であり、 $l_{c,k} = \|\mathbf{x}_{j_2,k} - \mathbf{x}_{j_1,k}\|$ は 3 次元ベクトル場の長さである。また、 $\mathbf{v}_{\perp}^{xy}, \mathbf{v}_{\perp}^{xz}$ は \mathbf{v} に垂直な y 方向および z 方向のベクトルである。

生成した正解 3 次元ベクトル場の例を図 6 に示す.正解 3 次元ベクトル場は各方向の成分ごとに分かれており,関節間ごとの値は同じになっている. 各方向に対応する成分が強いほど値が高くなり,値の高さに応じて色が赤くなる. 図 6 の 3 つの姿勢はいずれも立った姿勢であるため,Y 方向の成分の値が高くなっている関節間が多い. また,図 6 の中段の姿勢では左腕が手前に曲がっているため Z 方向の成分の値が高くなっている.

3.2 3次元ベクトル場推定モジュール

3次元ベクトル場は、2点の関節間の位置情報と方向情報を含んだ特徴表現である。3次元ベクトル場を推定することで関節間の位置情報と方向情報を考慮し、オクルージョンに対応した3次元姿勢推定を可能とする。3次元ベクトル場推定モジュールの構造を図7に示す。3次元ベクトル場推定モジュールの構造を図7に示す。3次元ベクトル場推定モジュールは、2次元姿勢推定モジュールの出力値および Deconvolution で拡張した Depth 回帰モジュールの中間特徴を、複数の Deconvolution 層を用いて特徴マップを3次元に拡張する。モジュールの構造は HourglassNet をもとに構成しており、Max Poolingと Upsamplingを用いて複数のスケールで特徴マップに畳み込み処理している。2次元姿勢推定モジュールの特徴マップに拡張した Depth 回帰モジュールからの中間特徴を追加することで、depth に対する特徴を考慮した3次元ベクトル場を推定できるため、より正確な

3次元ベクトル場を推定できると考える.

3次元ベクトル場に対する損失関数 \mathbf{L}_{vec} を式 (11) に示す。このとき, \mathbf{F}_c は推定 3次元ベクトル場, \mathbf{F}_c^* は正解 3次元ベクトル場を示す。また,c は各関節間を示す.

$$\mathbf{L}_{vec} = ||\mathbf{F}_c - \mathbf{F}_c^*||_2^2 \tag{11}$$

3.3 学習方法

提案手法の学習は 3 ステージに分けて行う. ステージ 1 は, 2 次元姿勢推定モジュールを 2 次元座標データで学習する. 損失関数は、式 (3) の損失関数のみを用いる. ステージ 2 、3 は,全モジュールを 2 次元,3 次元座標データで学習する. ステージ 2 は式 (4) の 3 次元データで学習する. ステージ 2 は式 (4) の 3 次元データセットの場合の損失関数および提案した損失関数である式 (11) を追加する. ステージ 3 は,式 (4) のGeometric Loss を追加する. また,ステージごとに一部の損失関数に係数を掛ける. ステージ 2 は,式 (4) の 3 次元データセットの場合の損失関数に 0.1,ステージ 3 は,式 (4) の 3 次元データセットの場合の損失関数に 0.1,式 (4) の Geometric Loss に 0.01 を重みとして乗算する.

4 評価実験

提案手法の有効性を確認するために評価実験を行う. 2次元,3次元の姿勢推定において従来手法との評価比較を行う.

4.1 データセット

データセットは、2次元データセットと3次元データセットの2つを使用する。2次元データセットは MPII データセット [13] を使用する。 MPII データセットは 大規模な in-the-wild の人間の姿勢推定のためのデータセットであり、オンラインビデオから収集された画像内にいる人物に対して 16 個の2次元関節位置のアノテーションが付けられている。また,MPII データセッ

表 1 各ステージにおける精度

		ステージ 1	ステージ 2	ステージ3
従来手法 [7]	PCKh@0.5[%]	84.2	91.7	92.0
	MPJPE[mm]	-	63.9	63.2
提案手法	PCKh@0.5[%]	84.2	91.7	91.9
	MPJPE[mm]	-	63.0	61.8

トは学習に使用するデータが 22,246 枚,検証に使用するデータが 2.958 枚ある.

3次元データセットは Human 3.6M データセット [14] を使用する. Human 3.6M データセットは,人間の 3次元姿勢推定に多く使用されているデータセットであり,モーションキャプチャで記録された屋内環境の RGB 画像および 3次元の関節座標がある.学習に使用するデータは Human 3.6M 内の 5 人の被験者のデータであり,総数は 312,188 枚ある.検証に使用するデータは Human 3.6M 内の 2 人の被験者のデータであり,総数は 2.874 枚ある.

4.2 評価方法

評価方法は、2次元姿勢推定に対する評価として PCKh@0.5、3次元姿勢推定に対する評価として MPJPE を使用する。PCKh@0.5 は、推定関節位置と 正解関節位置の距離が対象の頭部の 50%の大きさの閾値以下に存在する割合である。MPJPE は、正解 3次元関節位置と推定 3次元関節位置の誤差の平均である。 MPJPE を M,関節数を J,正解 3次元関節位置を P^* ,推定 3次元関節位置を P とすると、式 (12) で表すことができる。

$$M = \frac{1}{J} \sum_{i=1}^{J} ||P_i^* - P_i||_2$$
 (12)

4.3 従来手法との評価比較

提案手法の有効性を確認するため、2次元、3次元の 姿勢推定精度を [7] と比較する. 2次元, 3次元姿勢推 定における各ステージの精度を表1に示す. ステージ 1は、学習方法が同一であるため、提案手法と従来手 法 [7] の精度が同一になっている. 2次元, 3次元とも にステージごとに推定精度が向上していることが確認 できる. PCKh@0.5 において提案手法はステージ 2, 3 での精度が従来手法と同程度の精度であることが確認 できる. MPJPE は、提案手法がステージ 2、3 のいず れにおいても従来手法に比べ精度が向上しており, ス テージ3では、誤差が1.4mm 小さくなっていることが 確認できる.2次元姿勢推定精度が同程度の精度である のに対し、3次元姿勢推定精度が向上していることから depth の推定精度が向上していると考えられる. また, 他の従来手法との3次元姿勢推定精度の評価比較を表 2に示す. 提案手法は、比較手法以外の手法と比較して も精度が上回っていることを確認できる.

また, 提案手法の有効性を確認するため, 特定の条

表 2 3次元姿勢推定精度の評価比較

	MPJPE [mm]
Mehta[6]	74.1
Pavlakos[5]	66.9
Zhou[7]	63.2
Ours	61.8

表 3 オクルージョンを含む姿勢に対する精度

	MPJPE>120		Occlusion	
	従来手法 [7]	提案手法	従来手法 [7]	提案手法
右肘	94.8	93.1	81.1	80.5
右手首	131.7	122.3	106.4	102.4
左肘	82.0	80.2	71.4	71.7
左手首	116.8	116.3	98.1	103.8
右膝	74.3	70.7	68.6	67.6
右足首	114.3	100.8	99.9	92.9
左膝	71.2	68.8	66.6	65.3
左足首	115.0	107.6	102.6	98.9
平均	73.0	69.6	65.4	64.1

件下における評価比較を行った. 従来手法において, い ずれかの関節の MPJPE が 120.0mm を超えた姿勢およ びオクルージョンを含む姿勢における MPJPE の精度 を表3に示す. オクルージョンの有無は関節間と関節 位置から判断する. 関節間は2次元ベクトル場を生成 して表現する.2次元ベクトル場の生成は,3次元ベク トル場から Z 方向を除いたものになり、胴体部分のみ 両肩の半分の長さを閾値としている. また, オクルー ジョンの判断条件は,対象の関節を含まない関節間の 範囲に対象の関節があり、対象の関節の depth が関節 間の平均 depth より奥であるときとする. MPJPE が 120.0mm を超えた姿勢では、提案手法は手足のいずれ の関節においても従来手法に比べ精度が向上しており, 誤差が平均で3.4mm 小さくなっていることが確認でき る. このことから、提案手法は、従来手法では推定が難 しい姿勢を推定できていることがわかる. オクルージョ ンを含む姿勢では、提案手法は左腕を除くすべての関 節位置で従来手法に比べ精度が向上しており、誤差が 平均で 1.3mm 小さくなっていることが確認できる. 左 腕における精度の低下は,推定精度が全体的に大きく 向上している MPJPE が 120.0mm を超えた姿勢でも精 度の向上が小さいことから他の関節に比べ学習がうま くいかなかったと考えられる.

4.4 推定結果例の可視化

提案手法と従来手法における2次元,3次元姿勢推定 結果の例を図8に示す.図8の最上段の姿勢では,胴体によって隠れている左肘から左手首にかけての3次元姿勢推定結果において,従来手法では奥に伸びていないのに対して,提案手法では奥に伸びており,従来手法よりも正解関節位置との差が少ない.図8の中段

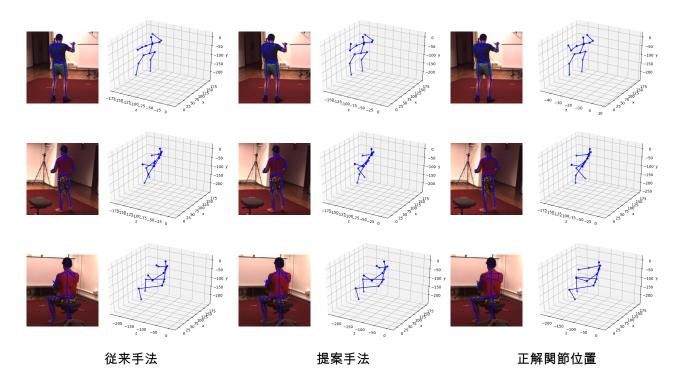


図8 2次元,3次元姿勢推定結果の比較



図 9 3次元ベクトル場の推定結果

の姿勢では、右膝から右足首にかけての3次元姿勢推定結果において、従来手法に比べ提案手法の方が正解関節位置との差が小さくなっている。図8の下段の姿勢では、2次元空間において関節間が短くなっている左肘から左手首の3次元姿勢推定結果において、従来手法に比べ提案手法の方が腕がより奥に伸びており、正解関節位置との差が小さくなっている。

3次元ベクトル場の推定結果を図9に示す。図9の上段および中段の姿勢は、直立した姿勢であるため、胴体から脚部のY方向の値が高い。また、下段の姿勢では、椅子に座っている姿勢であるため、脚部は大腿ではZ方向の値、下腿ではY方向の値が高い。また、特

定の方向の成分で値が高い関節間は、他の方向成分で値が低いことがわかる。例えば、図9の上段および中段の姿勢における胴体のベクトル場は、X方向の成分は値が低いため描画されていない。したがって、各姿勢における3次元ベクトル場の推定が可能であると考えられる。図8において推定結果が改善された関節に着目した時、上段の姿勢では左前腕、中断の姿勢では右脚、下段の姿勢では左前腕のZ方向の値が高くなっている。これらのことから、3次元ベクトル場の導入が姿勢推定に寄与していると考えられる。

5 おわりに

本研究では、3次元ベクトル場を考慮した3次元姿勢推定を提案した.提案手法では、従来手法のネットワークに3次元ベクトル場推定モジュールを導入することにより、3次元ベクトル場を推定し、関節間の方向に対する特徴を獲得する.これにより、関節間の方向に対する特徴を考慮した3次元姿勢推定を可能とする.評価実験により、提案手法は、従来手法より3次元姿勢推定の精度が向上した.また、推定3次元ベクトル場は方向の成分に応じた値を獲得することを確認した.今後の課題として、3次元ベクトル場推定モジュールの見直しによる精度向上が挙げられる.

参考文献

[1] J. Yang, Q. Liu, and K. Zhang, "Stacked Hourglass Network for Robust Facial Landmark Local-

- isation", The IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.79-87, 2017.
- [2] X. Bin, W. Haiping, and W. Yicgen, "Simple Baselines for Human Pose Estimation and Tracking", The European Conference on Computer Vision, pp.472–487, 2018.
- [3] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation", The IEEE Conference on Computer Vision and Pattern Recognition, pp.5693-5703, 2019.
- [4] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields", The IEEE Conference on Computer Vision and Pattern Recognition, pp.1302—1310, 2017.
- [5] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-Fine Volumetric Prediction for Single-Image 3D Human Pose", The IEEE International Conference on Computer Vision, pp.7025-7034, 2017.
- [6] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt, "Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision", 3D Vision, pp.506-516, 2017.
- [7] X. Zhou, Q. Huang, X. Sun, X. Xue, and Y. Wei, "Towards 3D Human Pose Estimation in the Wild: A Weakly-Supervised Approach", The IEEE International Conference on Computer Vision, pp.398-407, 2017.
- [8] I. Habibie, W. Xu, D. Mehta, G. Pons-Moll, and C. Theobalt, "In the Wild Human Pose Estimation Using Explicit 2D Features and Intermediate 3D Representations", The IEEE Conference on Computer Vision and Pattern Recognition, pp.10905–10914.
- [9] L. Zhao, X. Peng, Y. Tian, M. Kapadia, and D. N. Metaxas, "Semantic Graph Convolutional Networks for 3D Human Pose Regression", The IEEE Conference on Computer Vision and Pattern Recognition, pp.3425–3435, 2019.
- [10] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H. P. Seidel, W. Xu, D. Casas, and C. Theobalt, "Vnect: Real-time 3d human pose estimation with a single rgb camera", Association for Computing Machinery, vol.36, no.4, 2017.

- [11] R. Darbral, A. Mundhada, U. Kusupati, S. Afaque, A. Sharma, and A. Jain, "Learning 3D Human Pose from Structure and Motion", The European Conference on Computer Vision, pp.600–615, 2018.
- [12] G. Moon, J. Y. Chang, and K. M. Lee, "Camera Distance-Aware Top-Down Approach for 3D Multi-Person Pose Estimation From a Single RGB Image", The IEEE International Conference on Computer Vision, pp.10133-10142, 2019.
- [13] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of The Art Analysis", The IEEE Conference on Computer Vision and Pattern Recognition, pp.3686-3693, 2014.
- [14] C. Ionescu, P. Dragos, V. Olaru, and C. Sminchisescu, "Human3.6m:Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments", The IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.36, pp.1325–1339, 2014.