2 次元空間での奥行きベクトル場を考慮した人物の 3 次元姿勢推定

小松悠斗 †　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

人物の姿勢推定は対象の関節位置を推定する問題であり，広く研究されている．姿勢推定は大きく 2 次元の姿勢を推定する手法 [1] と 3 次元の姿勢を推定する手法 [2] に分類される．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，動作認識 [3, 4] や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 次元姿勢推定には，オクルージョンがある部位の特徴を捉えることが困難であるという問題点があり，これにより，推定精度が低下することがあげられる．

そこで本研究では，2 次元空間での奥行きベクトル場を考慮した 3 次元姿勢推定を行うことで姿勢推定の高精度化を行う．奥行きベクトル場は，関節間の位置情報とデプス方向の情報を含んだ特徴表現である．提案手法では奥行きベクトル場推定モジュールを従来手法に対して導入する．また，奥行きベクトル場の出力をもとに関節間の中心位置の推定を行う．これにより，デプス方向と関節間の位置の特徴を捉え，複雑な姿勢やオクルージョンに対応した姿勢推定を可能とする．評価実験により，奥行きベクトル場および関節間の中心位置を推定することで，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

姿勢推定は古くから研究されており，CNN を用いた推定手法による精度向上から [1]，近年では研究がより活発に行われるようになっている．姿勢推定には 2 次元姿勢推定と 3 次元姿勢推定が存在する．

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 [5, 6, 7] やボトムアップ型の手法 [8, 9] 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．

3 次元姿勢推定の場合，姿勢推定の手法は多岐にわたるため [10, 11, 12]，単眼からの単一の人物に対する姿勢推定の研究に対して焦点を当てる．単一の人物に対する 3 次元姿勢推定の手法は主に 2 つの種類に分類され，3 次元姿勢を直接推定する手法 [11, 13] と 2 次元姿勢と 3 次元姿勢を分けて推定する手法 [12, 14, 15] がある．3 次元姿勢を直接推定する手法は，直接 3 次元関節位置を回帰する手法 [13, 2] や，3 次元ヒートマップを推定することにより 3 次元関節位置を推定する手法 [11, 16] がある．2 次元姿勢と 3 次元姿勢を分けて推定する手法は，2 次元姿勢推定で獲得した特徴マップをもとに 3 次元関節位置を推定する手法 [12] や，2 次元姿勢推定で獲得した関節位置をもとに 3 次元姿勢推定を行う手法 [15, 17] がある．

単眼からの姿勢推定全般の問題として，手前にある物体が背後にある物体を隠して見えないようにする状態である，オクルージョンの発生している関節の推定が困難であることが挙げられる．単眼からの姿勢推定では一方向から人の姿勢を映すため，発生しやすくなっている．3 次元姿勢推定の手法の一つとして VFNet [18] がある．VFNet は，Zhou らの手法 [12] をベースとした手法であり，3 次元ベクトル場を推定することでオクルージョンに対応した 3 次元姿勢推定を行う．VFNet での 3 次元ベクトル場推定モジュールの導入は，デプスの推定に対して寄与したが，2 次元姿勢の x,y 座標の推定に対する精度向上は得られていない．また，3 次元畳み込み処理によってベクトル場を推定しているためネットワークサイズが非常に大きいという問題点がある．本研究では，VFNet の上記の問題に対処することで，よりオクルージョンに対応した 3 次元姿勢推定を行う．

3 提案手法

本研究では，オクルージョンに対応した 3 次元姿勢を行うために，2 次元空間での奥行きベクトル場を考慮した人物の 3 次元姿勢推定を提案する．提案手法のネットワークはベースとして Zhou らの手法 [12] を用いている．2 次元空間での奥行きベクトル場を推定するモジュールを導入し，奥行きベクトル場推定モジュー

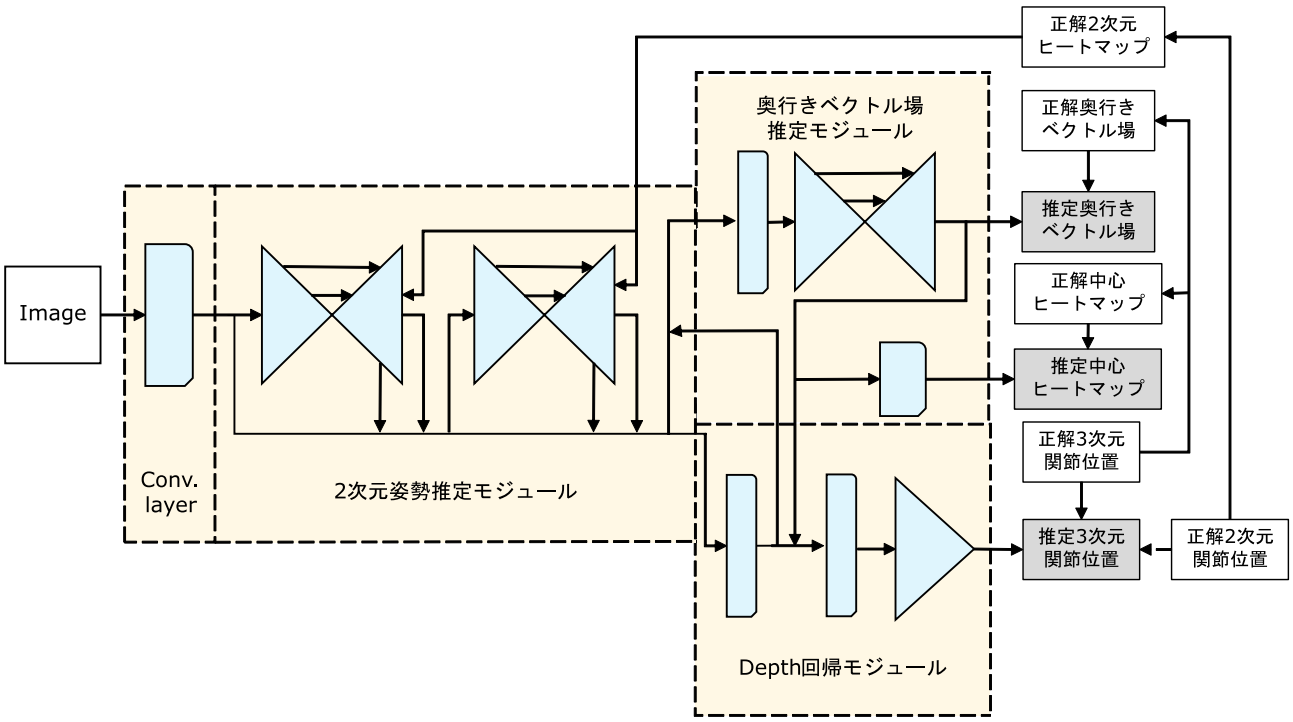

図 1　提案手法のネットワーク構造

ルの出力をデプスの推定に利用する．また，奥行きベクトル場推定モジュールの出力を利用して関節間の中心ヒートマップを推定する．これらの追加モジュールによって，デプス方向と関節間の位置の特徴を捉え，2 次元姿勢および関節のデプスの推定精度を向上させることができる．

3.1　ネットワーク構造

提案手法のネットワーク構造を図 1 に示す．2 次元姿勢推定モジュールは，Hourglass Network [5] をベースにしている．2 次元姿勢推定モジュールは，入力画像を複数の畳み込み層で畳み込み処理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して 2 次元姿勢をヒートマップで推定する．

追加モジュールである奥行きベクトル場推定モジュールおよび関節間の中心ヒートマップ推定モジュールの構造を図 2 に示す．奥行き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と，出力に対して 2 つの Residual ブロックから獲得した特徴を連結した特徴マップを入力とする．奥行きベクトル場推定モジュールは，まず入力した特徴マップに対して畳み込み処理を行い，次元を縮小する．次元を縮小した後に，Hourglass Network に基づく畳み込みネットワークで複数のスケールで畳み込み処理することにより奥行きベクトル場を推定する．

また，奥行きベクトル場推定モジュールの出力を入力とし，複数回畳み込み処理することで，関節間の中心ヒートマップを推定する．デプス回帰モジュールは 2 次元姿勢推定モジュールの出力を畳み込み処理した特徴マップと奥行きベクトル場推定モジュールの出力を連結した特徴マップを入力とする．デプス回帰モジュールは，入力した特徴マップに対して奥行きベクトル場推定モジュールと同様に次元を縮小した後，MaxPooling で特徴マップのサイズを小さくしつつ畳み込み処理し，デプス値を推定する．

3.2　正解奥行きベクトル場の生成

奥行きベクトル場推定モジュールの学習を行うため，正解奥行きベクトル場をデータセットから生成する．奥行きベクトル場は，2 点の関節間の位置情報とデプス方向の情報を含んだ特徴表現である．正解奥行きベクトル場を生成する条件を式 (1) に示す．図 3 に示すように，点 $\mathbf{p}$ が関節点 $\mathrm{J}_{1,xy}$ から関節点 $\mathrm{J}_{2,xy}$ を基準とする関節間 c 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．

$$\mathbf{L}^*_c(\mathbf{p}) = \begin{cases} \mathbf{v} & \text{if } \mathbf{p} \text{ on limb } c \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

このとき，$\mathbf{v} = (\mathbf{J}_{2,depth} - \mathbf{J}_{1,depth})/||\mathbf{J}_2 - \mathbf{J}_1||_2$ は始点 $\mathbf{J}_1$ から終点 $\mathbf{J}_2$ の単位ベクトルを表す．また，単位ベクトル $\mathbf{v}$ を生成する条件を式 (2) および (3) に示す．

$$0 \leq \mathbf{v} \cdot (\mathbf{p} - \mathbf{J}_{1,xy}) \leq l_c \tag{2}$$

$$|\mathbf{v}_\perp \cdot (\mathbf{p} - \mathbf{J}_{j_1,xy})| \leq \sigma_l \tag{3}$$

ここで，σ_l はピクセル単位の奥行きベクトル場の生成範囲を決める閾値であり，$l_c = \|\mathbf{J}_{2,xy} - \mathbf{J}_{1,xy}\|$ は奥

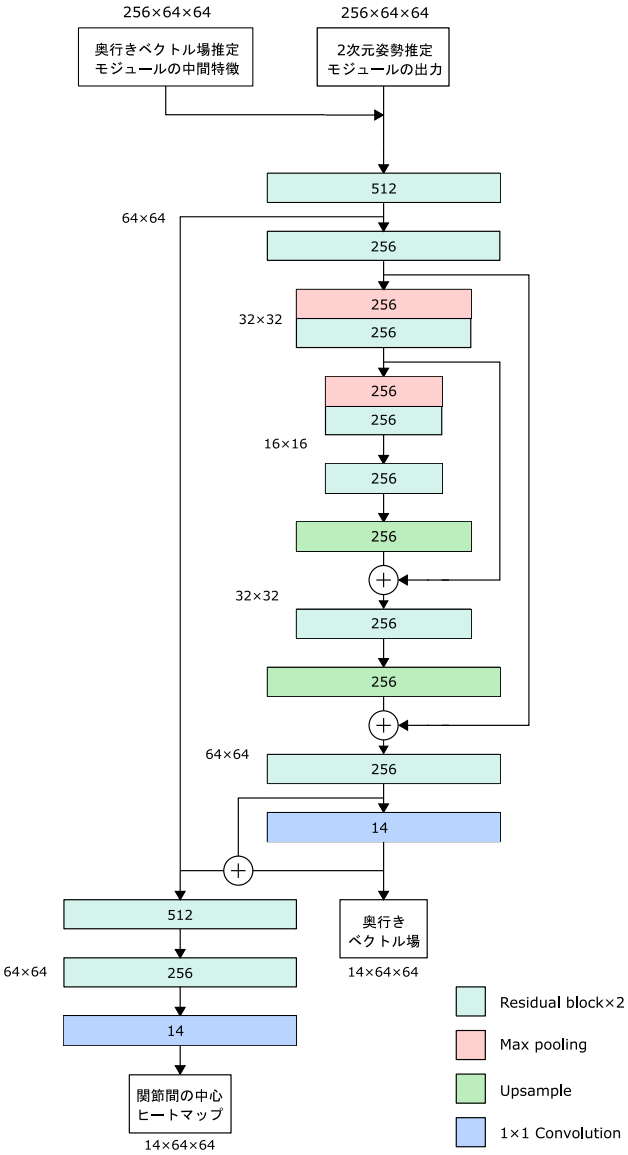

図 2　奥行きベクトル場推定モジュールおよび関節間の中心位置の推定モジュールの構造

行きベクトル場の長さである．また，$\mathbf{v}_\perp$ は $\mathbf{v}$ に垂直なベクトルである．

生成した正解奥行きベクトル場の例を図 4 に示す．正解奥行きベクトル場の関節間ごとの値は同じになっている．デプス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 4 の上の姿勢では，座って構えている姿勢であるため各関節でのデプス位置の差が大きくなっており，左足の大腿部や左腕の前腕部など様々な部位で生成ベクトル場の値が高くなっている．また，図 4 の下段の姿勢では立っている姿勢であるため各関節でのデプス位置の差が小さいことが多く，奥に伸びている右腕の前腕部以外では生成ベクトル場の値が低くなっている．

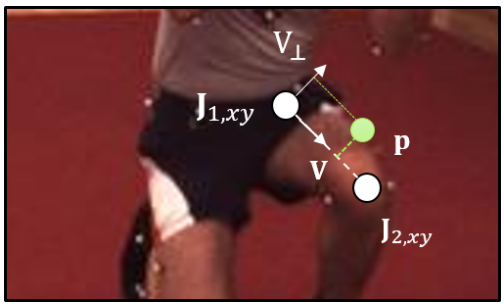

図 3　関節 $\mathbf{x}_{j_1}$，$\mathbf{x}_{j_2}$ と空間上の点 $\mathbf{p}$ の関係

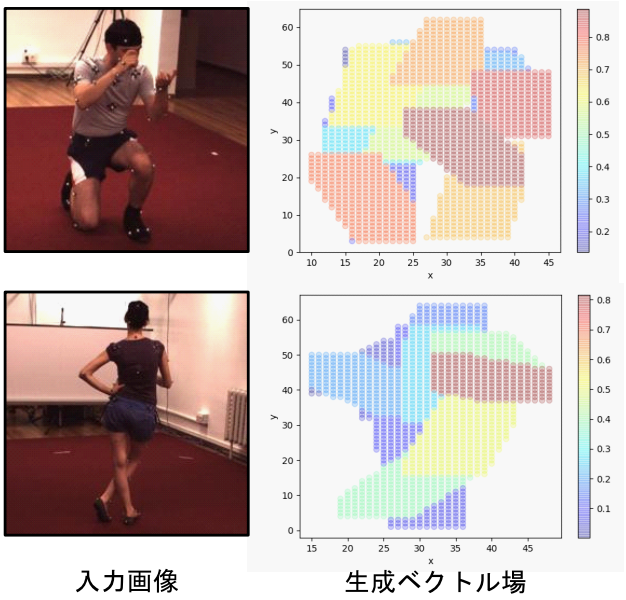

図 4　$\sigma = 6$ で生成時の奥行きベクトル場の例

3.3　損失関数

2 次元姿勢推定モジュールの損失関数には平均二乗誤差を用い，デプス回帰モジュールの損失関数には正解座標データが 2 次元座標データのとき Geometric Loss，3 次元座標データのとき平均二乗誤差を用いる．Geometric Loss は Zhou らの手法 [12] で提案された人の骨の長さの比が一定であることをもとにした損失関数である．Geometric Loss を式 (4) に示す．骨格グループ i の集合を R_i とする．l_c を推定された関節の e 番目の関節間の長さとし，$\bar{l}_c$ を正解データの e 番目の関節間の長さとする．また，$\bar{r}_i$ は分散に用いる各 R_i の比率 $l_c/\bar{l}_c$ の平均である．

$$L_{geo} = \sum_i \frac{1}{|R_i|} \sum_{c \in R_i} \left(\frac{l_c}{\bar{l}_c} - \bar{r}_i\right)^2 \tag{4}$$

$$\bar{r}_i = \frac{1}{|R_i|} \sum_{c \in R_i} \frac{l_c}{\bar{l}_c} \tag{5}$$

奥行きベクトル場推定モジュールの損失関数には，奥行きベクトル場の推定結果と正解との L2 ノルム二乗を用いる．関節間の中心ヒートマップの損失関数には，2 次元姿勢推定モジュールと同様に平均二乗誤差を用いる．

表 1　MPJPE による各関節の 3 次元姿勢推定精度

		右肘	右手首	左肘	左手首	右膝	右足首	左膝	左足首	平均
全体	Zhou ら [12]	80.0	106.5	71.4	96.9	64.0	94.5	61.7	95.5	83.8
	VFNet[18]	81.1	104.5	71.1	100.3	62.1	85.6	60.8	91.3	81.2
	提案手法	**71.9**	**92.6**	**68.3**	**93.1**	**55.4**	**79.3**	**57.6**	**83.6**	**75.2**
MPJPE>120	Zhou ら [12]	94.8	131.7	82.0	116.8	74.3	114.3	71.2	115.0	99.0
	VFNet[18]	93.1	122.3	80.2	116.3	70.7	100.8	68.8	107.6	94.9
	提案手法	**82.1**	**109.1**	**76.7**	**109.1**	**63.3**	**93.5**	**65.5**	**98.9**	**87.2**

表 2　各ステージにおける姿勢推定精度

		ステージ 1	ステージ 2	ステージ 3
PCKh@0.5 (2 次元姿勢)↑	Zhou ら [12]	84.2	91.7	92.0
	VFNet[18]	84.3	91.7	92.0
	提案手法	84.1	92.7	92.7
MPJPE (3 次元姿勢)↓	Zhou ら [12]	-	63.9	63.2
	VFNet[18]	-	61.9	61.8
	提案手法	-	60.1	59.1

3.4　学習方法

学習は 3 ステージに分けて行う．ステージ 1 では 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．ステージ 2，3 では，全モジュールを 2 次元および 3 次元座標データで学習する．デプス回帰モジュールの損失関数として，ステージ 2 は平均二乗誤差のみ，ステージ 3 は平均二乗誤差と Geometric Loss の両方を使用する．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，デプス回帰モジュールの平均二乗誤差の損失関数に 0.1，ステージ 3 は，デプス回帰モジュールの平均二乗誤差の損失関数に 0.1，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提案手法のベースラインである Zhou らの手法 [12] および VFNet[18] 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 [19] 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 Human3.6M データセット [20] を使用する．Human3.6M 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 Human3.6M 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 Human3.6M 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 (6) で表すことができる．

$$M = \frac{1}{J}\sum_{i=1}^{J} ||P_i^* - P_i||_2 \tag{6}$$

4.3　従来手法との評価比較

MPJPE による各関節の精度を表 1 に示す．表 1 における全体は，全体のデータで評価比較を行った結果を示し，表 1 における MPJPE $>$ 120 は，提案手法のベースラインである Zhou らの手法 [12] の推定時に，MPJPE が 120 を超えた関節を含む姿勢のみでの評価比較を行った結果を示す．提案手法は，全体のデータおよび MPJPE $>$ 120 のデータのいずれにおいても，すべての関節で従来手法に比べて誤差が減少している．ま

表 3　中心位置の推定の有無による姿勢推定精度

	中心推定なし	中心推定あり
右肘	76.0	**71.9**
右手首	98.3	**92.6**
左肘	70.8	**68.3**
左手首	**91.8**	93.1
右膝	58.5	**55.4**
右足首	84.8	**79.3**
左膝	**56.7**	57.6
左足首	91.6	**83.6**
平均	78.5	**75.2**

表 4　生成ベクトル場のサイズ変更時の推定精度

	サイズ	ステージ 1	ステージ 2
PCKh@0.5 (2 次元姿勢) ↑	4	84.1	**92.7**
	5	84.3	92.4
	6	84.1	92.3
MPJPE (3 次元姿勢) ↓	4	-	**60.1**
	5	-	64.7
	6	-	71.0

た，各ステージにおける 2 次元姿勢推定及び 3 次元姿勢推定の精度を表 2 に示す．提案手法の PCKh@0.5 は，従来手法と比べステージ 2 以降で精度が向上している．提案手法の MPJPE はステージ 2 で 60.1mm，ステージ 3 で 59.1mm であり，Zhou らの手法 [12] に比べ約 4.1mm，VFNet に比べ約 2.7mm 誤差が減少している．したがって，奥行きベクトル場推定モジュールおよび関節間の中心位置の推定が 3 次元姿勢推定に寄与していると考えられる．

4.4　関節間の中心位置の推定の有無による評価比較

関節間の中心位置の推定が精度改善に貢献しているかを確認するために，中心位置の推定モジュールおよび損失関数を設定した場合と設定していない場合で実験を行う．実験では，評価指標に MPJPE を用い，ステージ 3 まで学習したモデルで各関節の評価比較を行う．中心位置の推定の有無による姿勢推定精度を表 3 に示す．中心位置の推定モジュールを導入した場合では，右手首と左膝を除くすべての関節において，中心位置の推定モジュールがない場合に比べて誤差が減少している．この結果から，関節間の中心位置の推定が精度改善に貢献していることがわかる．

4.5　生成ベクトル場のサイズ変更時の評価比較

正解奥行きベクトル場は σ_l の値によりサイズの変更が可能である．正解ベクトル場のサイズを変更して実験を行う．実験では奥行きベクトル場推定モジュールの学習を行うステージ 2 までの精度で評価比較を行う．生成ベクトル場のサイズ変更時の推定精度を図 4 に示す．ステージ 2 では，2 次元姿勢および 3 次元姿勢のいずれの場合においてもサイズ 4 で学習した場合の推定精度が最も高いことが確認できる．この結果の要因として，正解ベクトル場のサイズに比例して生成ベクトル場のサイズも大きくなるため，関節間の中心位置の推定が困難になり，関節位置の推定に悪影響を及ぼしていると考える．

4.6　推定結果例の可視化

提案手法による 2 次元，3 次元姿勢および奥行きベクトル場の推定結果を図 5 に示す．推定奥行きベクトル場は，正解奥行きベクトル場と同様にデプス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．いずれの推定画像において画像奥に伸びている部位において推定奥行きベクトル場の値が高くなっていることが確認できる．また，3 次元姿勢推定結果も関節位置が奥に伸びていることから奥行きベクトル場が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，2 次元空間での奥行きベクトル場を考慮した人物の 3 次元姿勢推定手法を提案した．提案手法では，2 次元空間での奥行きベクトル場を推定モジュールを導入し，関節間の中心ヒートマップを推定することで，奥行き方向と関節間の特徴を獲得している．評価実験により，提案手法が従来手法に比べ，3 次元姿勢の推定精度が向上したことを確認した．また，推定奥行きベクトル場が関節間の奥行き方向の特徴を捉えていることを確認した．今後の課題としては，他の手法へ提案するモジュールを導入した際の効果の検証などが挙げられる．

参考文献

[1] A. Toshev and C. Szegedy: “Deeppose: Human pose estimation via deep neural networks”, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1653–1660 (2014).

[2] M. S. V. L. Bugra Tekin, Isinsu Katircioglu and P. Fua: “Structured prediction of 3d human pose with deep neural networks”, Proceedings of the British Machine Vision Conference, pp. 130.1–130.11 (2016).

[3] Y. Sijie, X. Yuanjun and L. Dahua: “Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition ”, Associa-

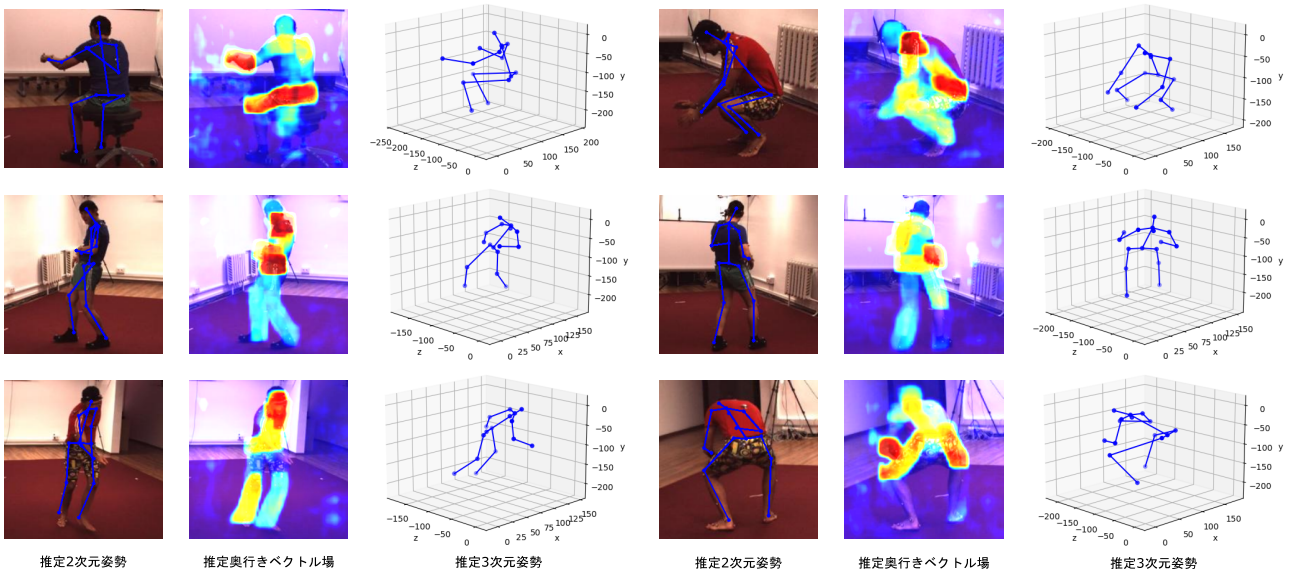

図 5　提案手法での各推定結果

tion for the Advancement of Artificial Intelligence (2018).

[4] L. Chaolong, C. Zhen, Z. Wenming, X. Chunyan, J. Rongrong and Y. Jian: “Action-Attending Graphic Neural Network ”, Institute of Electrical and Electronics Engineers, pp. 3657–3670 (2018).

[5] J. Yang, Q. Liu and K. Zhang: “Stacked Hourglass Network for Robust Facial Landmark Localisation”, The IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 79–87 (2017).

[6] 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).

[7] K. Sun, B. Xiao, D. Liu and J. Wang: “Deep high-resolution representation learning for human pose estimation”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019).

[8] 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).

[9] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang and L. Zhang: “Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5386–5395 (2020).

[10] 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).

[11] 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).

[12] 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).

[13] 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).

[14] 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.

[15] 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 Pat-

tern Recognition, pp. 3425–3435 (2019).

[16] X. Sun, B. Xiao, F. Wei, S. Liang and Y. Wei: “Integral human pose regression”, Proceedings of the European Conference on Computer Vision, pp. 529–545 (2018).

[17] S. Li, L. Ke, K. Pratama, Y.-W. Tai, C.-K. Tang and K.-T. Cheng: “Cascaded deep monocular 3d human pose estimation with evolutionary training data”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6173–6183 (2020).

[18] 小松悠斗, 白木克俊, 平川翼, 山下隆義, 藤吉弘亘：“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”, 画像センシングシンポジウム (2020).

[19] 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).

[20] 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).