

固有空間法での人物姿勢認識における連続フレームを利用した効率的な探索範囲限定 Effective Constraint of Search Area for Continuous Frame in Human Pose Recognition with Eigen Space Method

松原 文典†
Fuminori Matsubara

河中 治樹‡
Haruki Kawanaka

岩堀 祐之†
Yuji Iwahori

1. まえがき

これまでに提案されている姿勢認識の手法はモデルベースの手法とイメージベースの手法に大別される。モデルベースの方法には、マーカーやグローブを用いたモーションキャプチャによって得られた対象人物の関節座標から姿勢の推定を行う手法[1][2]などがある。しかし、手法[1][2]では座標データを得るためにマーカーが必要であり、認識対象に負担がかかる。マーカーを必要としない手法として、2台のカメラから撮影された画像でのシルエットとマーカーレスモーションキャプチャにより姿勢を推定する手法[3]が提案されている。また、複数視点の画像から作成したボリュームデータを細線化したスケルトンとあらかじめ用意されたモデル群を比較して類似姿勢を選定する手法[4][5]などがある。しかし、これらの手法[3][4][5]ではカメラが複数必要であるため、撮影環境の条件が限定される。

一方、イメージベースの手法としては、マーカーを必要とせず、1枚の入力画像から対象の撮影方向を認識する固有空間法[6]がある。この手法は認識対象を複数視点から撮影した画像群より固有空間を作成し、固有空間内で入力画像の特徴を最も良く表している画像を探索する方法である。さらに、固有空間法を人物の姿勢認識に応用した研究のひとつとして手法[7]が提案されている。[7]では、1方向のカメラ映像から人物の様々な姿勢の認識を実現するために、予め人物の様々な姿勢を様々な方向から観測したシルエット画像より固有空間を作成し、作成した固有空間内で認識対象に最も類似した画像を探索することによってその人物の姿勢を認識する。

しかし、[7]のようにシルエット情報のみによる姿勢認識では、例えば人物を前方や後方から撮影した場合に認識対象が正面向きであるのか後ろ向きであるのか区別がつかない場合がある。また、人物の姿勢によっては手足の左右の区別がつかない場合があり、これにより誤認識することがある。

そこで、[7]の手法をもとに過去のフレームの姿勢情報を利用して、人物の向きを誤りを起こさない姿勢認識の手法を提案する。シルエットだけでは判別できない姿勢への対処法としては、手法[3]のように視点の異なる複数の画像から部位の情報を得るものがあるが、我々の提案手法は時刻の異なる複数の画像から部位の位置関係を含めて姿勢を推定するというものである。本稿では、過去2フレームの身体部位の位置情報を用いて固有空間内での評価対象画像を限定した姿勢認識を行うことで、連続フレームにおける滑らかな姿勢変化を考慮した認識および処理時間の短縮が行えることを確認する。

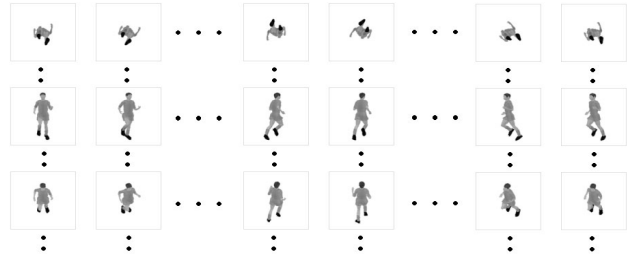


図1: 作成した人物モデルの画像群の一部

2. 固有空間法での姿勢認識

文献[5]では固有空間法による人物の姿勢認識を行うために、固有空間を構築する画像群をCGの人物モデルの画像で準備している。本研究でも同様に固有空間の作成に使用する画像群をCGの人物モデル(身長170cm)を利用して作成する。この画像群は「走る」「歩く」などの動作をそれぞれ3パターンずつ、計650個の姿勢から作成する。各々の姿勢について天頂角を22.5度おき、方位角を22.5度おきにとった144方向の視点から見た画像を作成することで合計93,600枚の画像群で構成する。各画像のサイズは64×64[pixel]とした。作成したCG画像群の一部を図1に示す。

このN枚のCG画像群(N=93600)を用いて固有空間を作成する。固有空間の作成には画像が持つ情報から人物の姿勢のみに注目するために、各々のシルエットを抽出して利用する。シルエット画像の縦横サイズを正規化し、ラスタースキャンをしてP個の変量(P=4096)を持つ一次元の特徴ベクトル $\mathbf{a}_i = \{a_1, a_2, \dots, a_P\}$ (i=1, ..., N)を得る。この処理を全ての画像について行い、得られた特徴ベクトル群 $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N)^T$ に主成分分析を施す。式(1)~(3)によりi番目の変量 \mathbf{a}_j の平均 \bar{a}_j 、分散 s_{ii} およびi, j番目の変量間の共分散 s_{ij} を算出する。

$$\bar{a}_i = \frac{1}{N} \sum_{n=1}^N a_{i,n} \quad (1)$$

$$s_{ii} = \frac{1}{N} \sum_{n=1}^N (a_{i,n} - \bar{a}_i)^2 \quad (2)$$

$$s_{ij} = \frac{1}{N} \sum_{n=1}^N (a_{i,n} - \bar{a}_i)(a_{j,n} - \bar{a}_j) \quad (3)$$

次に、得られた分散 s_{ii} および共分散 s_{ij} から構成される共分散行列 \mathbf{S} を使用して固有値 λ 、固有ベクトル Φ を次の固有値問題を解くことで算出する。

$$\mathbf{S}\Phi = \lambda\Phi \quad (4)$$

この λ および Φ において累積寄与率が90%以上となる主成分により固有空間を作成する。

人物の姿勢認識をする際には、入力の実画像中の人物領域を背景差分などによって抽出し、一次元ベクトル化および

† 中部大学, Chubu University

‡ 愛知県立大学, Aichi Prefectural University

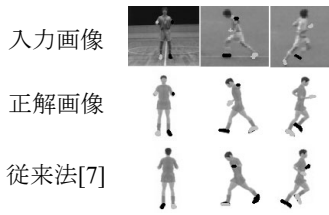


図 2: シルエット情報のみでの認識における失敗例

正規化する. 一般的な固有空間法においては, その特徴ベクトル \mathbf{g} , 画像群の各特徴ベクトル \mathbf{f}_i との固有空間内の距離 d_i を式(5)によって評価し,

$$d_i = \|\mathbf{g} - \mathbf{f}_i\| \quad (5)$$

d_i が最小となる姿勢 i を対象の姿勢として認識する.

しかし, シルエットのみからの姿勢認識では図 2 に示すように認識する画像によっては身体部位 (手足) の左右が入れ替わってしまう場合がある.

3. 部位の位置座標および画像座標による探索範囲の絞り込み

シルエット情報のみでは身体部位の左右の入れ替わりなどの問題が生じることがある. そこで, 過去フレームの身体部位座標 (右手, 左手, 右足, 左足の三次元位置座標および二次元画像座標) の情報を積極的に利用して, 入れ替わりが起きないようにする. そのために, 過去フレームから推定した現フレームの身体部位座標をもとに, 式(5)によって行われる画像群内での距離の評価範囲 (以下, これを単に探索範囲と呼ぶ) を予め絞り込んでおく. 以下にその手順を詳しく述べる.

時刻 t における身体部位 k ($k = 1, \dots, 8$) の速度ベクトル $\mathbf{v}_{t,k}$ を過去 2 フレームの身体部位座標 $\mathbf{x}_{t-1,k}$ および $\mathbf{x}_{t-2,k}$ から算出する.

$$\mathbf{v}_{t,k} = \mathbf{x}_{t-1,k} - \mathbf{x}_{t-2,k} \quad (6)$$

さらに, $\mathbf{v}_{t,k}$ に基づいて時刻 t における部位 k の推定部位座標 $\mathbf{x}'_{t,k}$ を次式によって求める.

$$\mathbf{x}'_{t,k} = \mathbf{x}_{t-1,k} + \mathbf{v}_{t,k} \quad (7)$$

この $\mathbf{x}'_{t,k}$ から距離 r_k 内に部位が存在する姿勢 i だけに探索範囲を限定する. ただし, すべての部位 k で範囲内に入っている姿勢を探索範囲とする.

この時, 探索範囲の距離 r_k は推定部位座標 $\mathbf{x}'_{t,k}$ の推定誤差を考慮して図 3 に示すように部位 k の速度 $\mathbf{v}_{t,k}$ から次式によって求める.

$$r_k = \|\mathbf{v}_{t,k}\| \cdot \tan \theta_k + C_k \quad (8)$$

ただし, θ_k および C_k は探索範囲調節のためのパラメータである. これにより部位の位置推定に生じる誤差の影響を軽減しながら, 探索範囲を設定することができる.

各部位 k の座標がすべて探索範囲に含まれる画像 \mathbf{f}_i と対象画像 \mathbf{g} との固有空間内の距離 d_i を評価し, 姿勢認識を行う. これにより認識精度が向上するとともに, 連続フレ

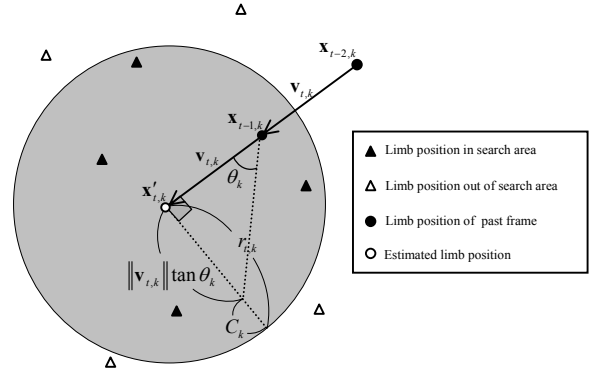


図 3: 探索範囲の絞り込みの概念図

表 1: パラメータ θ_k [度]

部位	右手	左手	右足	左足
三次元	71	54	62	54
二次元	38	37	37	39

表 2: パラメータ C_k

部位	右手	左手	右足	左足
三次元[mm]	208	140	118	148
二次元[pixel]	10	9	7	7

ームでの滑らかな姿勢の変化に対応した認識が可能となる.

式(8)のパラメータ θ_k および C_k は固有空間構築用の CG 画像群における人物モデルの身体部位座標を使用して適切な値に設定する. CG 画像群の人物モデルの姿勢は連続したもので構成しているため, 過去 2 フレームの姿勢から算出した推定部位座標 $\mathbf{x}'_{t,k}$ と現在フレームの姿勢の部位座標 $\mathbf{x}_{t,k}$ の距離が理想的な探索範囲と言える. このため, 全ての姿勢および視点における部位座標が次式の条件を満たすように θ_k および C_k を設定する.

$$\|\mathbf{x}'_{t,k} - \mathbf{x}_{t,k}\| \leq \|\mathbf{v}_{t,k}\| \tan \theta_k + C_k \quad (9)$$

全ての CG 画像において式(9)を満たす θ_k および C_k を線形最小二乗法によって算出する. 実際作成した画像群から算出した θ_k および C_k を表 1 および表 2 に示す.

4. 評価実験

シルエットだけでは姿勢の認識が困難であるような 15 フレームの連続した姿勢変化の画像系列を 5 種類用いて, 姿勢認識精度の評価実験を行う. 認識精度としては, あらかじめ目視で正解画像の選定を行い, 正解画像と認識結果の部位座標の平均誤差を測定する. また, 初期 2 フレームは姿勢認識の対象とはせず, 目視で選定した正解画像の部位座標を与えるものとする.

正解画像の部位座標 $\mathbf{X}_{t,k}$ と認識結果の部位座標 $\mathbf{Y}_{t,k}$ の平均誤差を計測して, 従来法[7]および提案手法の認識精度を比較する. 二次元部位座標の平均誤差 \bar{e}_{2D} および三次元部位座標の平均誤差 \bar{e}_{3D} をそれぞれ式(10)および(11)により算出する.

表 3: 正解画像と認識結果の部位座標の平均誤差

映像		1	2	3	4	5	平均
従来法	\bar{e}_{2D} [pixel]	27.20	8.73	27.22	36.73	13.66	22.71
	\bar{e}_{3D} [mm]	54.48	127.44	490.00	584.02	162.77	283.74
提案手法	\bar{e}_{2D} [pixel]	0.76	5.05	5.89	4.81	15.63	6.43
	\bar{e}_{3D} [mm]	16.49	32.02	56.40	57.76	247.75	82.08

$$\bar{e}_{2D} = \frac{1}{13} \sum_{t=3}^{15} \left(\frac{1}{4} \sum_{k=1}^4 \| \mathbf{X}_{t,k} - \mathbf{Y}_{t,k} \| \right) \quad (10)$$

$$\bar{e}_{3D} = \frac{1}{13} \sum_{t=3}^{15} \left(\frac{1}{4} \sum_{k=5}^8 \| \mathbf{X}_{t,k} - \mathbf{Y}_{t,k} \| \right) \quad (11)$$

ここでの k は, $k = 1 \sim 4$ は二次元の左右手足, $k = 5 \sim 8$ は三次元の左右手足をそれぞれ表す. 実験により得られた各映像における \bar{e}_{2D} および \bar{e}_{3D} を表 3 に示す. また従来法および提案手法での認識結果を図 4 ~ 7 に示す. 図 4 ~ 7 では左手足の部位を右手足と区別できるように濃い黒色に塗って示してある.

表 3 より, 映像 1 ~ 4 においては提案手法の方が正解との誤差が小さく, シルエット情報だけで姿勢認識を行う従来法よりも, 身体部位の位置情報を利用して探索範囲を絞り込む提案手法の方が認識精度が高くなっている. 従来法に比べて二次元座標における平均誤差は約 1/3 に, 三次元座標における平均誤差も約 1/3 に減らすことができた. また, 図 4 の $t = 13 \sim 15$ の時の認識姿勢の画像が示しているように, 従来法では部位の入れ替わりが起きる画像であっても提案手法では入れ替わらずに認識できていることが確認できる. また, 図 5 のように人物の前方から撮影された画像の認識においては, 従来法のようにシルエットだけでは前方および後方のどちらを向いているかの判別ができず, 結果として前方から撮影された画像であるのに後方から見た画像として認識されてしまう. しかし, 提案手法のように探索範囲を絞り込むことで姿勢(体の前後の向き)を正しく認識できていることが確認できる.

しかし図 7 にある映像 5 に対する認識結果が示すように, 連続する認識処理の過程で同じ姿勢の画像が結果として続くような場合においては, 提案手法で得た姿勢は正解と大きく異なるものになり, 表 3 に示すように部位の位置座標の平均誤差も大きくなっている. これは, 図 7 における $t = 6 \sim 9$ ように数フレームが連続して同じ姿勢と認識した場合, 探索範囲に正解が含まれなくなってしまうためであると考えられる. 一旦探索範囲からはずれてしまうとそれ以降のフレームでは誤認識を起こす可能性が高い. そのために, 認識結果画像と入力画像のシルエットから類似度を算出するなどして, 認識失敗を検出する必要がある. 認識を誤った時点でパラメータ θ_k および C_k の補正, 過去 2 フレーム分の部位座標の修正, もしくは左右の推定部位座標の関係から部位の入れ替わりが起らない程度で範囲を広げるなどの対策が考えられる.

次に, 探索範囲の絞り込みを行うことにより処理時間(探索時間)の短縮にもなるため, その効果を検証する. 探索範囲の絞り込みを行わない従来法[7]と絞り込みを行う提案手法のそれぞれの処理時間を計測した. 実験に用いた PC の仕

表 4: 平均処理時間[fps]

映像	1	2	3	4	5	平均
従来法	3.67	3.74	3.66	3.70	3.67	3.69
提案手法	0.78	0.85	0.85	0.84	0.77	0.82

表 5: 絞り込み後の平均候補数[個]

映像 1	映像 2	映像 3	映像 4	映像 5	平均
12.46	55.77	53.31	44.85	9.38	41.6

様は CPU : Pentium(R)4 2.40GHz, Memory : 1024 MB である. 1 ~ 5 の各映像における 1 フレームあたりの平均処理時間を表 4 に示す. 表 4 より従来法に比べて提案手法の方が処理時間を約 1/4 に抑えられることがわかる. また, 表 5 に示すように, 画像群 93600 個から平均で約 42 個の候補に絞り込みを行っていることから処理時間の大幅な短縮ができたことがわかる.

5. まとめ

過去 2 フレームの姿勢情報から固有空間内の探索範囲を限定し, シルエットのみでの認識では困難であった姿勢の認識を行う方法を提案した. 提案手法を実装し, 実験により提案手法の有効性を確認した.

しかし, 過去 2 フレームの部位座標をもとにして探索範囲を限定しているため, 一旦探索範囲の絞り込みに失敗し, 姿勢を誤認識してしまうとそれ以降のフレームにおいて正しい姿勢の認識ができないという問題点がある. これを解決することは今後の課題である.

謝辞

本研究の一部は中部大学研究費のサポートによる. ここに感謝申し上げる.

参考文献

- [1] J. Lee, et al. "Constraint-based hand animation", *Models and Techniques in Computer Animation (Computer Animation Series)*, 110-127, 1993.
- [2] 山根亮, 他 "動作データの時系列相関行列による舞踊動作解析", 信学論, **J88-D2(8)**:1652-1661, 2005.
- [3] T. Tangkuampien et al. "Real-time human pose inference using kernel principal component pre-image approximations", *BMVC2006*.
- [4] 田中秀典, 他 "ボリュウムデータの細線化とグラフマッチングを用いた事例ベース人体姿勢推定", *CVIM158-22*, 2007.
- [5] Chi-Wei Chu, et al. "Markerless Kinematic Model and Motion Capture from Volume Sequences", *Proc. of CVPR 2003*, 2: 475-482, 2003.
- [6] 村瀬洋, 他 "2次元照合による3次元物体認識—パラメトリック固有空間法—", 信学論 D-II, **J77-D-II(11)**:2179-2187, 1994.
- [7] Haruki Kawanaka, et al. "Recognition Shape and Generation of Free Viewpoint Image from Database", *FORMA*, **19(3)**:279-286, 2005.

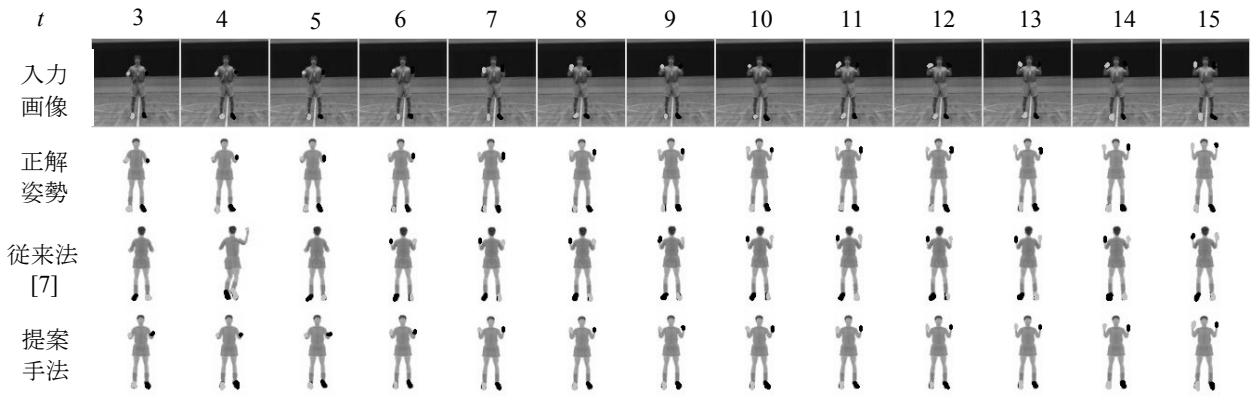


図4: 映像1の入力画像およびその正解姿勢画像と各手法での認識結果

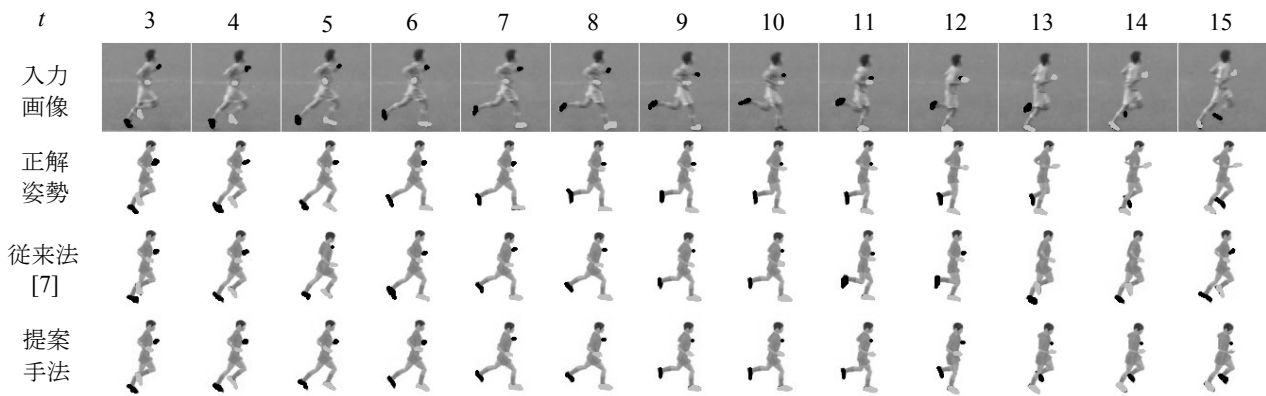


図5: 映像2の入力画像およびその正解姿勢画像と各手法での認識結果

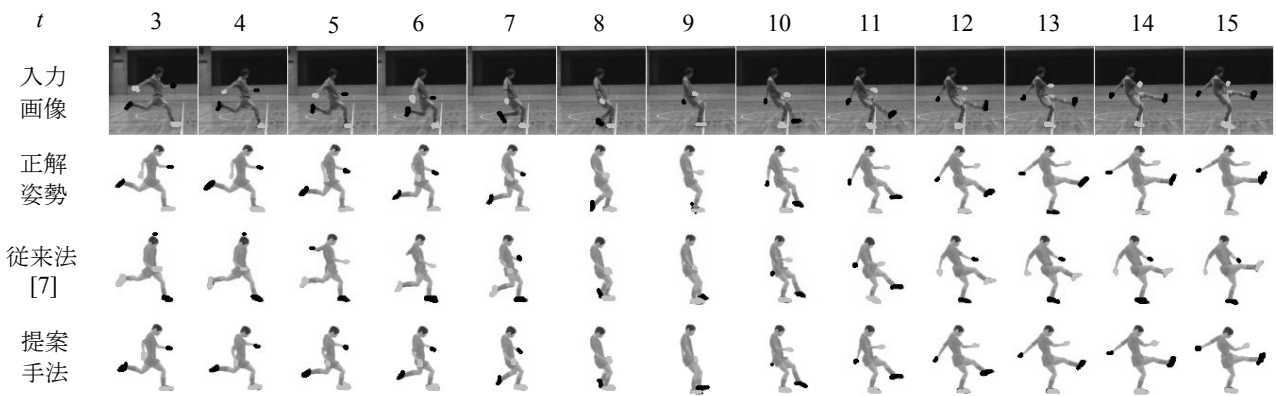


図6: 映像3の入力画像およびその正解姿勢画像と各手法での認識結果

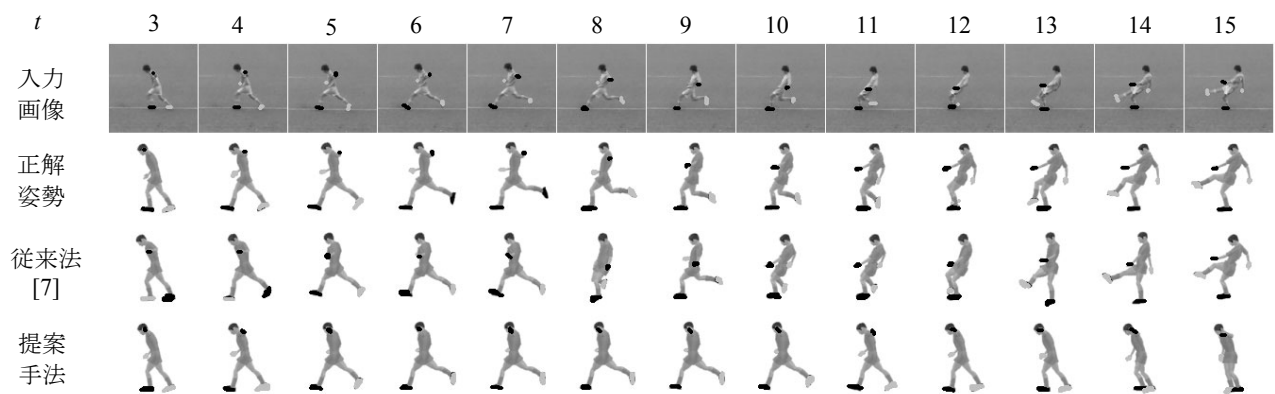


図7: 映像5の入力画像およびその正解姿勢画像と各手法での認識結果