

1. はじめに

近年、頭部に装着したカメラによる一人称視点映像を解析し、人の行動を支援する First Person Vision の研究が盛んに行われている．本研究では、頭部に装着したステレオカメラ映像から前進、右折、左折、階段の上り、下りの5種類の自己動作の識別を対象とする．提案手法では、マルチクラス識別である Random Forests[1] の分岐関数に、特徴量の共起表現を導入することで、高精度な自己動作識別を実現する．

2. 提案手法

提案手法では、Space-Time Patch(ST-patch) から得られる特徴量を Random Forests を用いて自己動作識別を行う．図1に提案手法の流れを示す．カメラの各シーン映像の大域領域と局所領域から時空間特徴である ST-patch を算出する．マルチクラス識別器である Random Forests により、自己動作として計5種類の動作を識別する．このとき、左右のカメラから得られる特徴量を共起することで高精度化が期待できる．

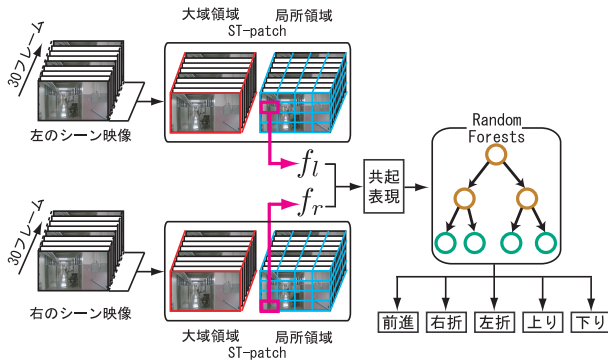


図1：提案手法の流れ

2.1. ST-patch による動きを捉える特徴量

映像中の動きを捉えるために、ST-patchを用いる．ST-patchとは、画像を時間方向に重ねた3次元データである時空間画像の局所領域における x 軸、 y 軸、 t 軸の勾配値から算出する．各軸に対するある画素 i の勾配は、 $\nabla P_i = [P_{xi} \ P_{yi} \ P_{ti}]^T$ となり、この時空間特徴から式(1)に示すST-patchを算出する．

$$M = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y & \sum P_x P_t \\ \sum P_y P_x & \sum P_y^2 & \sum P_y P_t \\ \sum P_t P_x & \sum P_t P_y & \sum P_t^2 \end{bmatrix} \quad (1)$$

式(1)より求められる行列 M が1つのST-patchから得られる特徴となる．行列 $M(3 \times 3)$ には、左上の 2×2 にアピアランス情報、3行目と3列目にはモーションの情報を持つ．本手法では、1つの動作(30フレーム)から10フレーム毎に、画像全体を1つのパッチサイズとした大域的ST-patchと、画像を16分割と64分割にした各領域のST-patchを求める．各ST-patchから時空間特徴として、動きの定常度(0~1)と動きの類似度(0~1)を算出する．

2.2. 特徴量の共起表現

算出したST-patchを用いてマルチクラス識別器である Random Forests を構築する．Random Forests は複数の決定木を構築する識別器であり、決定木は分岐ノードを作成することで学習する．従来の Random Forests では、各分岐ノードに対し、入力データが I_n 、左右に分岐するデータがそれぞれ I_l, I_r とし、特徴量 f と閾値 t を用いて式(2)により左右に分岐する．

$$I_l = \{i \in I_n | f < t\} \\ I_r = I_n \setminus I_l \quad (2)$$

この分岐処理では、左右のカメラのどちらか単一の特徴量を捉えて分岐するものであり、カメラ間の関係性が利用さ

れていない．そこで提案手法では、この分岐関数に共起表現を導入する．左のシーン映像の特徴量と右のシーン映像の特徴量を f_l, f_r とし、直線の一般式 $a f_r + b f_l + c$ の値が、式(3)のように0以下の場合は左、それ以外は右に分岐する．

$$I_l = \{i \in I_n | a f_r + b f_l + c \leq 0\} \\ I_r = I_n \setminus I_l \quad (3)$$

a, b, c はランダムで $-1 \sim 1$ の範囲で選択する変数である．図2に提案手法による分岐関数の分岐図を示す．(I)の場合では、2つの特徴量を用いた分岐、すなわち共起となる．(II)と(III)では、単一の特徴量を用いた分岐となる．直線の一般式 $a f_r + b f_l + c = 0$ を用いることで、2つの特徴量の共起表現から単一の特徴量を対象とした分岐を実現できる．

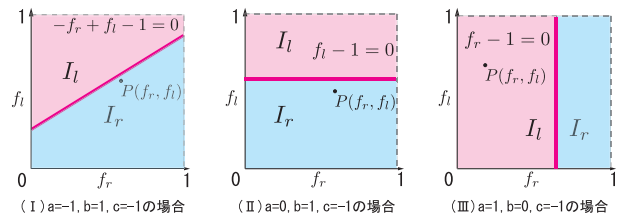


図2：分岐図

3. 評価実験

提案手法の有効性を示すために、頭部に装着したステレオカメラで撮影した各動作シーンにおける識別精度の評価実験を行う．データベースは、屋内における12人分の歩行動作のシーケンスから、学習用サンプルが4144個、評価用サンプルが1263個とする．従来法として、左右のカメラの特徴量をランダムで選択し、従来の Random Forests の分岐関数と同様に分岐をする．学習パラメータは、木の数を30、木の深さを20、サブセットの割合を0.30とする．提案手法では、特徴選択回数を1863回、閾値選択回数を1000回とし、従来法では、特徴選択回数を43回、閾値選択回数を10回とする．図3(a)に従来法、(b)に提案手法の識別結果を示す．F値の平均は、提案手法が91.1%、比較手法が86.2%である．全動作のF値において、提案手法の精度が向上していることがわかる．

	前進	右折	左折	上り	下り		前進	右折	左折	上り	下り
前進	94.2	6.2	4.4	0.0	1.3	前進	96.6	2.5	2.4	0.0	1.3
右折	0.0	82.0	0.0	0.5	0.0	右折	0.0	91.6	0.0	0.9	0.0
左折	0.6	0.0	92.9	0.0	0.7	左折	0.0	0.0	95.9	0.0	1.5
上り	1.1	7.2	0.2	90.9	12.6	上り	0.5	2.8	0.0	94.4	8.4
下り	3.8	7.0	3.1	3.2	71.3	下り	3.9	3.2	2.1	3.9	76.9

(a) 従来法(単一)

(b) 提案手法(共起)

図3：識別結果

分岐ノードにおける直線のパラメータである a, b の値は $a > 0.1$ かつ $b > 0.1$ となるノードが約85%存在した．これは多くの分岐ノードで2つの特徴量を用いた共起が採用されたため、提案手法の精度が向上したと考えられる．また、ST-patchの要素として、 $\sum P_x P_y$ や $\sum P_t^2$ が多く選択されており、アピアランスと動きの情報を捉えていることがわかる．

4. おわりに

本研究では、ステレオカメラを用いた一人称視点映像における自己動作識別法を提案し、その有効性を示した．今後は、Random Forests の分岐関数に2次曲線を用いる等検討し、自己動作識別法の精度向上を目指す．

参考文献

- [1] L.Breiman, "Random Forests", Machine learning, Vol. 45, pp. 5-32, 2001.