

1. はじめに

近年、映像中から人の動作を認識する研究が盛んに行われている。しかし、従来法の多くは対象とする人を外側から撮影しているため、人がどのような意図で行動しているのかという情報を知ることが困難である。このような問題に対して、人の行動意図を認識するための First-Person Vision というアプローチ [1] が提案されている。First-Person Vision とは、図 2 のように Inside-Out カメラを人の頭部に装着し、人の視界を観測しつつ、もう一つのカメラで眼球を観測することで、人の行動意図を認識する試みである。本研究では、Inside-Out カメラを用いて撮影したシーン映像からの自己動作の識別法を提案し、その有効性を示す。



図 1：Inside-Out カメラ (文献 [1] より引用)

2. 提案手法

提案手法では、Space-Time Patch(ST-patch)[2]を用いた自己動作の識別を行う。図 2 に提案手法の流れを示す。本手法では、連続した 30 フレームを 1 つの入力として動作を識別する。入力映像の大域領域と局所領域から ST-patch 特徴量を算出し、マルチクラス識別のための Boosting 手法である Joint-Boosting を用い、前進、右折、左折、つまづく、階段の上り、下りの動作を識別する。

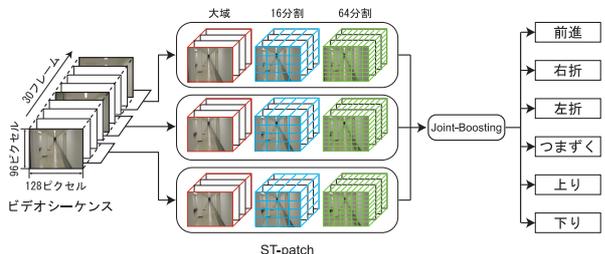


図 2：ST-patch 特徴を用いた Joint-Boosting による識別

2.1 ST-patch による動きの解析

ST-patch 特徴は、画像を時間軸方向に重ねた時空間画像における  $x$  軸、 $y$  軸、 $t$  軸の勾配値より求められる。よって画素  $i$  からは、 $P_{x_i}$ 、 $P_{y_i}$ 、 $P_{t_i}$  の 3 つの勾配値を持つ時空間勾配が得られる。この時空間特徴を局所領域の画素数でスタックした行列  $G$  を算出し、行列  $G$  に転置行列をかけた行列  $M$  が 1 つの ST-patch から得られる ST-patch 特徴である。

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

本手法では、画像を 128pixel×96pixel にダウンサンプリングした映像から、ST-patch 特徴を算出する。図 2 に示すように、一つの動作 (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 特徴と、画像を 16 分割と 64 分割にした各領域 1 つずつをパッチサイズとした局所的 ST-patch 特徴を算出する。さらに、各 ST-patch から動きの定常度 (CRIM) と、前後 2 つの ST-patch から動きの類似度 (MC) 算出する。

2.2 Joint-Boosting による識別器の構築

6 種類の動作識別を行うために、マルチクラス識別のための Boosting 手法である Joint-Boosting を用いた学習により識別器を構築する。1 つの入力映像から得られる特徴

量は膨大になるが、Joint-Boosting は共通する特徴 (弱識別器) を共有するため、識別にかかる計算量を抑えることができる。

3. 評価実験

頭部に装着した Inside-Out カメラを用いて撮影した各動作シーンの識別実験を行う。データベースは、学習用サンプルとして、前進 57 個、右折 68 個、左折 75 個、つまづく 60 個、上り 94 個、下り 95 個の映像を使用する。評価用サンプルとして学習用サンプルとは別に撮影した、前進 39 個、右折 30 個、左折 21 個、つまづく 12 個、上り 20 個、下り 23 個の映像を使用する。表 1(a) に提案手法である ST-patch による識別結果を示し、表 1(b) に比較手法として、オプティカルフローによる識別結果を示す。

表 1：識別実験結果

(a)ST-patchによる識別結果

		出力						再現率[%]
		前進	右折	左折	つまづく	上り	下り	
入力	前進	922	0	0	0	2	24	97.3
	右折	11	108	0	0	39	2	67.5
	左折	8	0	90	1	25	7	68.7
	つまづく	6	0	2	18	2	0	64.3
	上り	2	1	0	0	88	1	95.7
	下り	8	0	0	0	10	31	63.3
	適合率[%]	96.3	99.1	97.8	94.7	53.0	47.7	
F値[%]	96.8	80.3	80.7	76.6	68.2	54.4	76.2	

(b)オプティカルフローによる識別結果

		出力						再現率[%]
		前進	右折	左折	つまづく	上り	下り	
入力	前進	750	23	9	15	112	39	79.1
	右折	10	142	0	4	4	0	88.8
	左折	17	0	106	1	4	3	80.9
	つまづく	11	1	2	5	7	2	17.9
	上り	9	2	2	3	75	1	81.5
	下り	27	2	0	3	8	9	18.4
	適合率[%]	91.0	83.5	89.1	16.1	35.7	16.7	
F値[%]	84.7	86.1	84.8	16.9	49.7	17.5	56.6	

表 1 より、F 値の平均を比較すると提案手法はオプティカルフローより 19.6 % 向上した。オプティカルフローは、ノイズや動きが大きくなると正確なフローを求めることができないため、歩行時の振動が識別率の低下につながったと考えられる。提案手法では、右折、左折を階段の上りと多く誤認識している。これは、映像内の大部分にテクスチャの少ない壁が映っている右折、左折のシーン映像が多いため、勾配情報が得られにくく、識別が困難になったと考えられる。識別例を図 3 に示す。Joint-Boosting により選択された ST-patch 特徴には、横方向と縦方向の勾配が多く、建物内におけるドアや窓などのエッジを多く捉えるということが判明した。提案手法は、30 フレームから算出する 1 つのサンプルに対して、約 50msec で識別が可能である。



図 3：識別例

4. おわりに

本研究では、人の視界を観測した映像から ST-patch 特徴量を用いた自己動作の識別法を提案し、その有効性を示した。今後は、眼球を観測した映像を用いた自己動作の識別精度の向上について検討する予定である。

参考文献

[1] T. Kanade: "First-Person, Inside-Out vision", CVPR First Workshop on Egocentric vision (2009).  
 [2] E. Shechtman, M. Irani: "Space-time behavior based correlation", PAMI, 29, 11, pp. 2045-2056 (2007).