

First Person VisionのためのST-patch特徴を用いた自己動作識別

古川 拓弥[†] 藤吉 弘亘[†]

[†] 中部大学大学院 工学研究科 〒487-8501 愛知県春日井市松本町 1200

E-mail: †takuya@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

あらまし 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、従来法の多くは人を外側から撮影しているため、行動意図の認識は困難である。このような問題に対し、First Person Vision というアプローチが提案されている。これは、Inside-Out カメラと呼ばれる 2 つのカメラが搭載されている装置を人の頭部に装着し、人の視界を観測するシーン映像と、眼球を観測した眼球映像から人の行動意図を認識する試みである。本稿では、この First Person Vision を実現するために、Inside-Out カメラを用いて撮影したシーン映像のみから大域的 ST-patch と局所的 ST-patch を用いて特徴を算出し、マルチクラスの Boosting 手法である Joint Boosting を用いて歩行時における 6 種類の動作識別を行う。評価実験の結果、提案手法はオプティカルフローを用いた比較手法よりも識別精度が 27.3% 向上した。

キーワード First Person Vision, Inside-Out カメラ, ST-patch, Joint Boosting

Action Recognition using ST-patch Features for First Person Vision

Takuya FURUKAWA[†] and Hironobu FUJIYOSHI[†]

[†] Dept. of Computer Science, Chubu University Matsumototy 1200, kasugai-shi, Aichi, 487-8501 Japan

E-mail: †takuya@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract Much research has been devoted in recent years to recognizing human action from video images. Most existing methods, however, take video of people from the outside making it difficult to understand behavioral intention. The First Person Vision approach has been proposed in response to this problem. In this approach, a device consisting of two cameras collectively called an “inside-out camera” is attached to the head of a person to obtain “scene video” that captures the person’s visual field and “eyeball video” that observes one of the person’s eyeballs. These video streams are used as a basis for understanding that person’s behavioral intention. With the aim of realizing First Person Vision, we here calculate features using a global ST-patch and a local ST-patch from scene video and attempt to distinguish six types of actions while walking using Joint Boosting. Results of a comparison experiment revealed that the proposed method improved the accuracy of distinguishing actions by 27.3% compared to an optical-flow method.

Key words First Person Vision, Inside-Out camera, ST-patch, Joint Boosting

1. ま え が き

オフィス、家、公共施設などの空間内における人の意図を理解し、行動を支援する研究が数多く行われている。ジョージア工科大学の Aware Home プロジェクトでは、生活空間である家にカメラをはじめとする数多くのセンサ群を埋め込み、24 時間を通して生活空間における人の動きをセンシングする研究に取り組んでいる [1]。また、リビングルームを対象とした Microsoft Research の Easy Living [2] のように、センシングにより得られた情報を基に、ユーザである人に対して快適

な空間をアシストする活力生活技術は QoLT (Quality of Life Technology) と呼ばれる技術の一環で、QoLT のためのセンシング技術は、刻々と変化する人の動作や状態を実時間で認識する必要がある。人の動作を認識する手法は数多く提案されており、イベント検出 [3] やジェスチャ認識 [4]、異常行動検知 [5] [6] や動作識別 [7] などの研究が行われている。しかし、これらの動作認識は人を外側から撮影した映像を用いているため、人がどのような意図で行動しているのかを正確に知ることは困難である。

このような問題に対して、より人の行動意図を認識できるよ

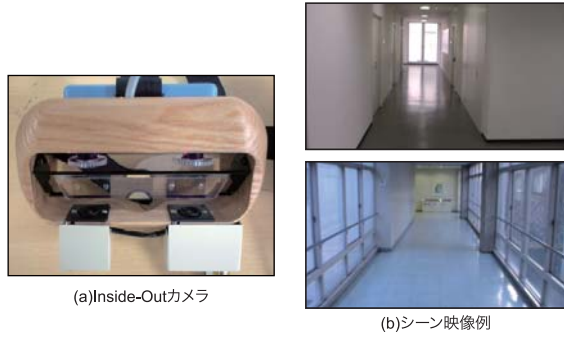


図 1 Inside-Out カメラを用いたシーン映像例

うに, First Person Vision [8] というアプローチが提案されている. これは, Inside-Out カメラと呼ばれる装置を対象である人の頭部に装着し, 人の視界を観測した一人称視点の映像 (シーン映像) を得ると同時に, もう 1 つのカメラで眼球の動きを捉えた映像 (眼球映像) を取得し, 行動意図を認識しようとする試みである. 眼球を撮影したカメラ画像を用いて瞳孔を抽出し視線を求め, シーンを撮影したカメラ画像上の注視点を算出することで, 何を見ているのかを知ることができる [9]. また, シーンを撮影した画像からは画像分類技術を用いた自己位置同定をすることが可能となる [10]. しかしながら, これらの技術により求めた視線方向や位置情報に応じた情報提供を行う際に, シーン映像から対象者の行動や動作を正しく認識していなければ適切な情報提供は行えない.

一人称視点映像における動作認識として, 興枡らは頭部に装着したウェアラブルカメラを用いて, 撮影したシーン映像の画像と事前に登録された画像データベースの照合結果から絶対位置と頭部方位を取得し, 腰部に装着された慣性センサ (加速度, ジャイロ, 磁気センサ) の情報から歩行動作を検出することで, 歩行者の位置推定を行っている [11]. しかし, シーン映像は絶対位置と頭部方位の取得に使用しており, 動作認識を行っていない. 青木らはカメラの動きによるフレーム間の画像変化をアフィン変換で表すことができると仮定し, フレーム間のアフィン変換パラメータから人物の移動量を推定している [12]. しかし, この手法ではウェアラブルカメラを胸部に設置しているため, 人の視点から見た映像と異なるという問題がある.

そこで, 本研究では, 対象者の頭部に Inside-Out カメラを装着して撮影した屋内におけるシーン映像のみを用いて, 前進, 右折, 左折, つまづく, 階段の昇り, 降りの 6 種類の動作を識別する手法を提案する. シーン映像中からアピアランスと見えを同時に捉える ST-patch 特徴を大域領域と局所領域から算出し, マルチクラス識別のための Boosting 手法である Joint Boosting を用いて各動作の識別を行う.

2. ST-patch 特徴を用いた自己動作識別

本研究で扱う映像は, 図 1(a) の Inside-Out カメラを対象者の頭部に装着し, 屋内における歩行時の状況を撮影したシーン映像 (図 1(b)) を対象とする. 提案手法の流れを図 2 に示す. 提案手法は連続した 30 フレームから大域領域と局所領域の

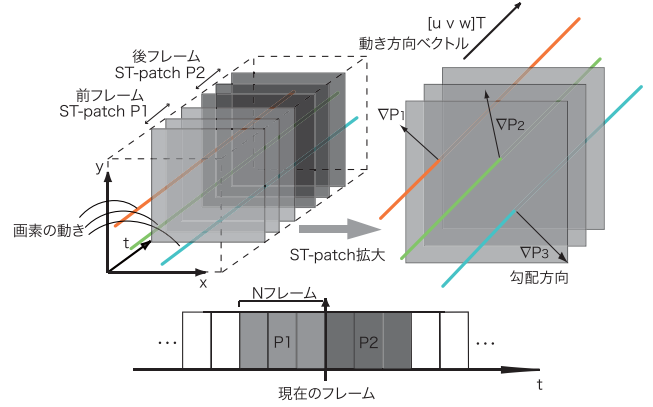


図 3 ST-patch の概要図

ST-patch から ST-patch 特徴を抽出し, Joint Boosting を用いて各動作の識別を行う. 本章では, 映像中の動きの解析を行う ST-patch についてと提案手法における ST-patch 特徴の算出方法, さらに, マルチクラス識別のための Boosting 手法である Joint Boosting について説明する.

2.1 ST-patch

画像上の動きの解析を行う手法として, 物体の移動量をベクトルデータとして表現したオプティカルフロー [13][14] や, 時系列 2 値化画像に対して変位パターンを当てはめ, 時間方向に積算した CHLAC 特徴 [15] が挙げられる. しかし, Inside-Out カメラで撮影されたシーン映像は背景が動的であるため, オプティカルフローや CHLAC 特徴では歩行動作以外の動作に対してノイズが発生しやすいという問題がある.

そこで本稿では, 映像中における物体の「アピアランス」と「モーション」の 2 つの情報を持つ ST-patch 特徴を利用する. ST-patch の相関関係は, テクスチャが異なる対象物体の動きや, 非剛体の物体の様に複雑に動く物体に対しても動きの評価を行うことができる. 図 3 に ST-patch の概要図を示す. 異なる色で示したフレームをそれぞれ ST-patch P1 と ST-patch P2 とする. x, y は画像の座標軸, t は時間軸, 3 本の線は個々の画像の動き $[u \ v \ w]^T$ は ST-patch 内の動きベクトル, ∇P_i は個々の画素の勾配方向ベクトルを表している. この ST-patch 特徴を得るために, 時空間画像においての x 軸, y 軸の勾配を求める. 映像中の動きが一定の場合, 各軸に対するある画素 i の勾配 $\nabla P_i = [P_{xi} \ P_{yi} \ P_{ti}]^T$ は, 画素の動き方向ベクトル $[u \ v \ w]^T$ に対して垂直となるため, 式 (1) の関係が成り立つ.

$$\nabla P_i^T \begin{bmatrix} u \\ v \\ w \end{bmatrix} = 0 \quad (1)$$

画素数が n の場合, 式 (1) は式 (2) となる.

$$\begin{bmatrix} P_{x1} & P_{y1} & P_{t1} \\ P_{x2} & P_{y2} & P_{t2} \\ \vdots & \vdots & \vdots \\ P_{xn} & P_{yn} & P_{tn} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2)$$

ST-patch 中の n 画素の ∇P_i からなる $n \times 3$ の行列を G とし,

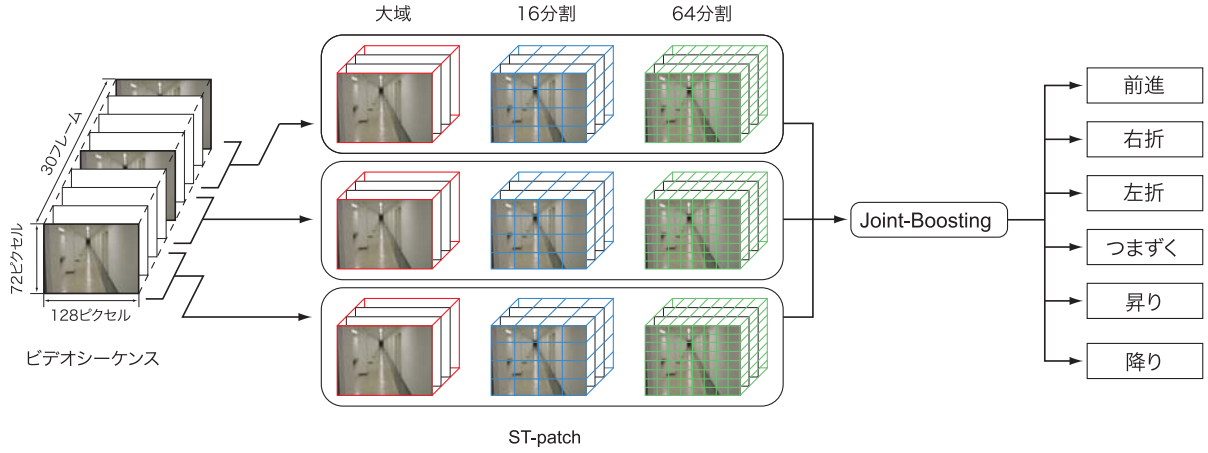


図 2 ST-patch 特徴を用いた動作識別の流れ

行列 G^T を掛けると式 (3) となる .

$$G^T G \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (3)$$

ここで、行列 $G^T G$ を M とすると式 (4) のように表すことができる .

$$M = G^T G = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y & \sum P_x P_t \\ \sum P_y P_x & \sum P_y^2 & \sum P_y P_t \\ \sum P_t P_x & \sum P_t P_y & \sum P_t^2 \end{bmatrix} \quad (4)$$

式 (4) 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 の行列にアピランスの情報、3 行目と 3 列目にはモーションの情報を持つ .

2.2 2 つの ST-patch の一致

2 つの ST-patch P_1, P_2 間に共通のベクトル $u = [u \ v \ w]^T$ が存在する場合、 P_1, P_2 間の動きは一致し、各 ST-patch における行列 G_1 と G_2 は $G_1 u = 0$ かつ $G_2 u = 0$ となる . ここで、 $G_1 u = 0$ と $G_2 u = 0$ をまとめると式 (5) となる .

$$G_{12} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}, \quad \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (5)$$

行列 G_{12} は ST-patch の P_1 と P_2 の両方の勾配を含む . 式 (3)、式 (4) と同様に、式 (5) に G_{12}^T を掛け、 $M_{12} = G_{12}^T G_{12}$ とすると式 (6) となる .

$$M_{12} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (6)$$

また、 M_{12} は式 (7) により求めることもできる .

$$M_{12} = M_1 + M_2 = G_1^T G_1 + G_2^T G_2 \quad (7)$$

2.3 ランク制約による状態判定

ST-patch の時空間特性と画像空間特性を用いて、ランク制約による動きの状態判定を行う . ここで、時空間特性とは式 (4) で示す行列 M を表し、画像空間特性とは以下に示す行列 M° を表す .

$$M^\circ = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y \\ \sum P_y P_x & \sum P_y^2 \end{bmatrix} \quad (8)$$

ST-patch の時空間特性から、単一の動きの ST-patch は $\text{rank}(M) = \text{rank}(M^\circ)$ のようなランクとなる . そして、複数の動きを含む ST-patch では、1 つの行と列だけが行列 M から M° へ変換されているため、ランクの違いが 1 となる . よって式 (9) のような関係が成り立つ .

$$\Delta r = \text{rank}(M) - \text{rank}(M^\circ) = \begin{cases} 0 : \text{単一の動き} \\ 1 : \text{複数の動き} \end{cases} \quad (9)$$

同様に、2 つの異なる ST-patch においても、動きが互いに一致しているかを式 (9) により求めることができる .

2.3.1 CRIM (Continues Rank-Increase Measure)

ランクの増加量 Δr は、動きが単一か複数の 2 通りしか求めることができず、動きがどれだけ類似しているかを求めることはできない . そこで、行列 M と行列 M° より決められる固有値を用いて Continues Rank-Increase Measure (CRIM) Δr を算出する . $\lambda_1 \geq \lambda_2 \geq \lambda_3$ を行列 M の固有値、 $\lambda_1^\circ \geq \lambda_2^\circ$ を行列 M° の固有値とする . 行列 M 、行列 M° は共に対象行列であり、行列 M 、行列 M° の固有値はすべて実数となる . また、それぞれの固有値の関係は $\lambda_1 \geq \lambda_1^\circ \geq \lambda_2 \geq \lambda_2^\circ \geq \lambda_3$ となる . このような関係から、 $\text{CRIM} \Delta r$ を式 (10) のように定義する .

$$\Delta r = \frac{\lambda_2 \cdot \lambda_3}{\lambda_1^\circ \cdot \lambda_2^\circ} \quad (0 \leq \Delta r \leq 1) \quad (10)$$

ここで求められた Δr の値は、ST-patch 内の動きベクトルが一定の時に $\Delta r \approx 0$ となり、動きベクトルの変化が激しいときに $\Delta r \approx 1$ となる . これにより、ST-patch 内の動きベクトルの定常度を求めることができる .

2.3.2 MC (Motion Correlation)

式 (10) から求められた $\text{CRIM} \Delta r$ より、2 つの ST-patch P_1

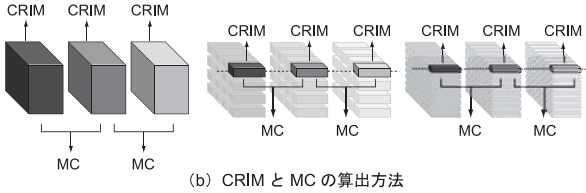
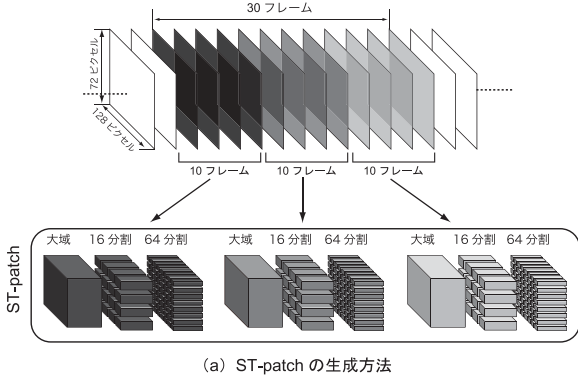


図 4 大域的 ST-patch と局所的 ST-patch を用いた特徴量

と P2 の動きの類似度を求めることができる．P1 の CRIM を Δr_1 , P2 の CRIM を Δr_2 , P1 と P2 をまとめた CRIM を Δr_{12} とすると, 式 (11) により Motion Correlation (MC) m_{12} を求められる．

$$m_{12} = \frac{\min(\Delta r_1, \Delta r_2)}{\Delta r_{12}} \quad (11)$$

ここで求められた m_{12} は, $m \approx 1$ の場合は動きの類似度が高く, $m \approx 0$ の場合は動きの類似度が低いことを意味する．

2.4 大域的 ST-patch と局所的 ST-patch を用いた特徴算出方法

提案手法では, 図 4(a) に示すように, 画像サイズを 128×72 ピクセルにダウンサンプリングした映像から, 連続した 30 フレームの動作を入力とし特徴抽出を行う．30 フレームから 10 フレーム毎に画像全体をパッチサイズとした大域的 ST-patch と, 画像サイズを 16 分割と 64 分割にした領域 1 つずつをパッチサイズとした局所的 ST-patch を生成し, それぞれから ST-patch 特徴を算出する．さらに, これらの ST-patch から CRIM と MC も算出する．図 4(b) に示すように, CRIM は 1 つの ST-patch から算出し, MC は同サイズの同じ位置にある前後 2 つの ST-patch を用いて算出する．よって, 30 フレームからは 243 個の ST-patch が生成され, ST-patch 特徴が 1458 次元, CRIM が 243 次元, MC が 162 次元の計 1863 次元の特徴量となる．ST-patch のサイズを複数組み合わせることにより, 画像全体の動きと局所的な動きを同時に捉えることができる．

2.5 マルチクラス識別機の構築

提案手法では, 歩行時における 6 種類の動作識別を行うために Joint Boosting を用いる．Joint Boosting により前進, 右折, 左折, つまづく, 階段の昇り, 降りの動作をポジティブクラス, 停止している動作をネガティブクラスとして学習を行い識別器を構築する．

Joint Boosting [16] は, マルチクラス識別のための Boosting

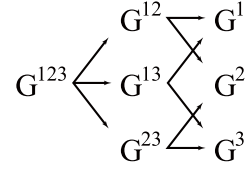


図 5 3 クラスにおける弱識別器集合の共有

手法であり, 共通する弱識別器集合を共有しながら学習を行う手法である．Joint Boosting における弱識別器を $h_m(v, c)$ とすると, 強識別器 $H(v, c)$ は次式となる．

$$H(v, c) = \sum_{m=1}^M h_m(v, c) \quad (12)$$

ここで c はクラスラベルである．Joint Boosting では, クラス集合 $S(n)$ の識別に対して用いられる弱識別器集合を $G^{S(n)}(v)$ とすると, 以下のように表される．

$$G^{S(n)}(v) = \sum_{m=1}^M h_m^n(v) \quad (13)$$

例として, 3 クラスについてのマルチクラス識別器を考える場合, それぞれのクラスを識別する強識別器は $G^{S(n)}(v)$ を用いて以下のように表される．

$$\begin{aligned} H(v, 1) &= G^{1,2,3}(v) + G^{1,2}(v) + G^{1,3}(v) + G^1(v) \\ H(v, 2) &= G^{1,2,3}(v) + G^{1,2}(v) + G^{2,3}(v) + G^2(v) \\ H(v, 3) &= G^{1,2,3}(v) + G^{1,3}(v) + G^{2,3}(v) + G^3(v) \end{aligned} \quad (14)$$

このとき $G^{1,2,3}$ は検出対象 1~3 すべてと背景を, $G^{1,3}$ は検出対象 1 と 3 すべてと背景を識別するのに有効な弱識別器集合である．Joint Boosting は図 5 に示すように, 強識別器間で共通する弱識別器集合を共有する．3 クラスの識別では, 各クラスの強識別器が 4 個の弱識別器集合を必要とするため, 合計で 12 個の弱識別器集合が必要となる．しかし, 識別の際は強識別器間で弱識別器集合を共有しているため, 7 個の弱識別器集合を算出するだけでよい．

本稿では, 前進, 右折, 左折, つまづく, 階段の昇り, 降りの 6 クラスの識別を対象としているため, 弱識別器集合は 63 個となる．

3. 実 験

提案手法の有効性を示すために, 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精度の評価実験を行う．

3.1 データベース

データベースは, Inside-Out カメラを用いて屋内における 10 人分の歩行動作を撮影したシーケンスから作成する．各動作毎に連続した 30 フレームを 1 つのサンプルとして, 学習用 9218 個, 評価用 3956 個のサンプルを用いる．各動作毎の学習用サンプル数と評価用サンプル数を表 1 に示す．今回の実験では, 対象者が停止している際の映像はその他とした．

表 1 データベース

	学習用サンプル	評価用サンプル
前進	3162	1356
右折	660	284
左折	609	262
つまずく	947	406
昇り	1258	540
降り	1075	462
その他	1507	646

表 2 識別結果のコンフュージョンマトリクス
(a) 提案手法

出力 入力	前進	右折	左折	つまずく	昇り	降り	その他	再現率[%]
前進	1180	1	3	5	21	34	112	87.0
右折	0	246	2	2	4	19	11	86.6
左折	0	0	231	6	4	4	17	88.2
つまずく	2	1	1	374	1	2	25	92.1
昇り	2	0	0	1	510	13	14	94.4
降り	7	0	0	4	29	401	21	86.8
その他	10	0	1	0	0	0	635	98.3
適合率[%]	98.3	99.2	97.1	95.4	89.6	84.8	76.0	

(b) 比較手法

出力 入力	前進	右折	左折	つまずく	昇り	降り	その他	再現率[%]
前進	959	5	9	10	14	32	327	70.7
右折	2	157	4	30	7	13	71	55.3
左折	1	10	132	20	12	2	85	50.4
つまずく	2	12	17	226	22	2	125	55.7
昇り	5	9	19	22	307	14	164	56.9
降り	44	1	1	2	54	209	151	45.2
その他	24	1	0	0	2	18	601	93.0
適合率[%]	92.5	80.5	72.5	72.9	73.4	72.1	39.4	

3.2 実験概要

評価用データベースを用いて識別精度の評価実験を行う。比較手法では、オプティカルフローから得られる特徴量を、Joint Boosting により識別する。提案手法と同様に、連続する 30 フレームを入力として、オプティカルフローを 10 フレーム毎に 3 つ算出する。算出したオプティカルフローから、大域領域と局所領域 (16 分割, 64 分割) 毎にフロー方向ヒストグラムを作成し特徴量とする。フロー方向の量子化は 20 度間隔とする。それぞれの特徴量を用いて識別器を構築する学習回数は 1000 回とし、識別精度は再現率、適合率、F 値を用いて評価を行う。

3.3 実験結果

表 2 にそれぞれの特徴を用いた識別結果のコンフュージョンマトリクスを示し、図 6 に各動作毎の F 値を示す。また、提案手法における動作識別例を図 7 に示す。表 2 より提案手法は比較手法に比べ、全ての動作において識別精度が向上している。図 6 の F 値の平均では提案手法は 90.6%、比較手法は 63.3% となり、27.3% 向上した。

また、表 2(b) に示すように、比較手法では、つまずくや階段の昇り、降りの識別率は低く、その他に誤識別されているものが多い。表 2(a) の提案手法においては、つまずくや階段の昇り、降りにおけるその他への誤識別が減少しており、識別精度が向上していることがわかる。

3.4 学習により選択された特徴量

図 8 に、学習により選択された弱識別器集合の特徴量の頻度を示す。 $G^1, G^2, G^3, G^4, G^5, G^6$ はそれぞれ前進、右折、左折、つまずく、階段の昇り、降りに対応している。右折、左折

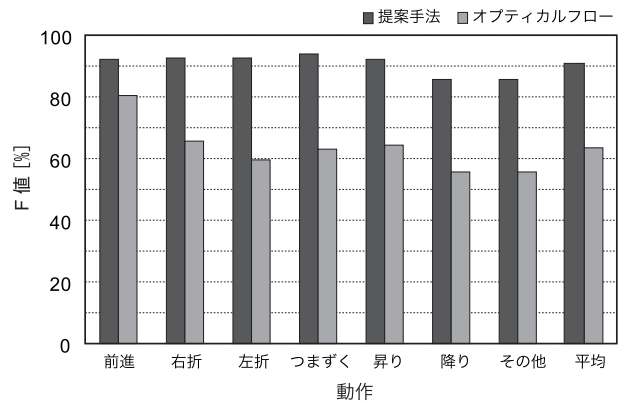


図 6 動作識別結果

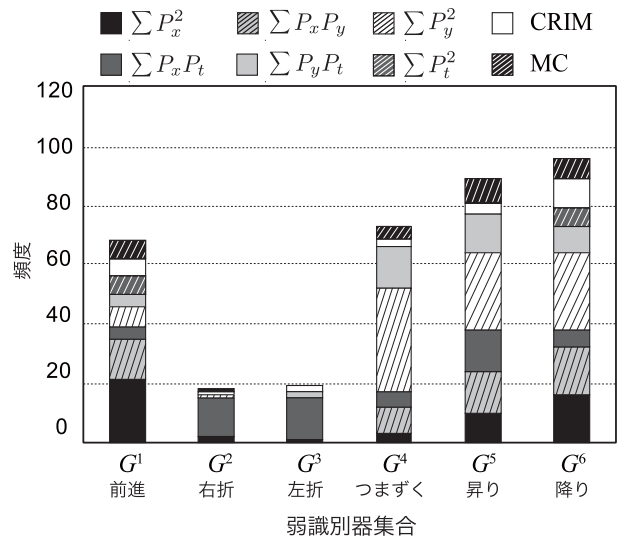


図 8 学習により選択された特徴量の頻度

の識別に必要な弱識別器の数は少なく、モーション情報を含んだ $\sum P_y P_t$ の特徴が多く選択されている。階段の昇り、降りの識別にはアピランスの情報が多く選択されており、特に、y 方向の勾配である $\sum P_y^2$ の特徴が多いことから階段の水平方向のエッジを捉えて識別をしていると考えられる。

4. まとめ

本稿では、First Person Vision を実現するために、Inside-Out カメラを用いて撮影したシーン映像から 6 種類の自己動作を識別する手法を提案した。提案手法は大域的 ST-patch と局所的 ST-patch を組み合わせ特徴量を算出し、Joint Boosting により識別を行った。評価実験の結果、提案手法における F 値の平均値は 90.6% であり、オプティカルフローを用いた手法に比べ、27.3% 向上した。提案手法は、首振りなどによる頭部の動きを考慮していないため、頭部の動きが入った映像では誤識別をする可能性があると考えられる。今後は、加速度センサやジャイロセンサを用いて、頭部の動きに対応した自己動作識別を行う予定である。

文献

- [1] "Aware home," <http://awarehome.imtc.gatech.edu/>.
- [2] B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer, "Easyliving: Technologies for intelligent environments,"



図 7 提案手法における動作識別例

Proc.of International Symposium on Handheld and Ubiquitous Computing, pp.12–29, 2000.

- [3] Y. Ke, R. Sukthankar, and M. Hebert, “Event detection in crowded videos,” In Proc. IEEE International Conference on Computer Vision, pp.8–15, 2007.
- [4] 飯田直也, 戸澤慶昭, 中村明生, “ジェスチャ認識を用いた家電機器操作インターフェースの検討 -操作開始合図としての手ぶり動作認識及び操作機器特定のための指差し方向推定-” 映像情報メディア学会技術報告, vol.33, pp.81–84, Aug. 2009.
- [5] Y. Murai, H. Fujiyoshi, and M. Kazui, “Incident detection based on dynamic background modeling and statistical learning using spatio-temporal features,” Machine Vision Applications 2009(MVA2009), vol.5-3, pp.156–161, 2009.
- [6] 南里卓也, 大津展之, “複数人数からの異常動作検出” 情報処理学会論文誌, コンピュータビジョンとイメージメディア (CVIM), vol.45, pp.43–50, 2005.
- [7] 佐藤竜太, 亀田能成, 大田友一, “CHLAC 特徴量と部分空間法による複数行動の分類” 画像の認識・理解シンポジウム MIRU, 2009.
- [8] T. Kanade, “First-person, inside-out vision,” IEEE Workshop on Egocentric Vision, CVPR, 2009.
- [9] 西尾和晃, 木村誠, 藤吉弘巨, 平田豊, “自由空間における注視点の 3 次元位置を取得する inside-out カメラの提案” pp.IS1-13, 2009.
- [10] H. Kang, A.A. Efros, M. Hebert, and T. Kanade, “Image matching in large scale indoor environment,” IEEE Workshop on Egocentric Vision, CVPR, 2009.
- [11] 興梠正克, 蔵田武志, “慣性センサ群とウェアラブルカメラを用いた歩行動作解析に基づくパーソナルポジショニング手法” 信学技報 PRMU2004, vol.103, pp.25–30, 2004.
- [12] 青木茂樹, 高橋昌史, 大西正輝, 小島篤博, 福永邦雄, “ウェアラブルカメラによる動作認識とテキスト表現” 信学技報

PRMU2003, vol.103, pp.59–64, 2003.

- [13] B. Horn, and B. Dchunck, “Determining optical flow,” Artificial Intelligence, vol.17, pp.185–203, 1981.
- [14] B.D. Lucas, and T. Kanade, “An iterative image registration technique with an application to stereo vision,” In Proc. 7th International Joint Conference on Artificial Intelligence, pp.674–679, 1981.
- [15] T. Kobayashi, and N. Otsu, “Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation,” International Conference on Pattern Recognition, 2004.
- [16] A. Torralba, K.P. Murphy, and W.T. Freeman, “Sharing visual features for multiclass and multiview object detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29, no.5, pp.854–869, 2007.