1S-06

機械学習を用いた動画からの行動認識手法

関 向然 * 飯島 安恵 * 今野 将 †

* 千葉工業大学大学院 工学研究科 † 千葉工業大学 先進工学部 知能メディア工学科

1 はじめに

近年の情報技術の急速な発展に伴い，コンピュータを用いた人間の行動を認識する手法が注目されてきている．これら行動認識手法の最終的な目標は，未知の動画などから人間の行動を正しく認識し，人間の行動の意味を分析することである．このような人間の行動認識手法の応用範囲は広く，自動サーベイランス，病人サーベイランス，エンターテインメントコンピューティングなど多岐にわたっており注目されている．

既存の人間の行動認識手法の研究では “運動メッセージに基づく動的特徴法” や “optical flow を用いた手法” が提案されている [1]．これら二つの手法は同一エリアにおける同一人物の行動を認識し識別することは可能である．しかしながら，例えば照射される光の加減が異なったり，撮影している画角が異なったり，撮影時の背景が異なったりするなどの空間的・時間的複雑性が生じると認識精度が低下するという問題点がある．

そこで本研究では前述の空間的・時間的複雑性が要因となる認識精度の低下という問題点を解決するために，機械学習を用いた人間の行動認識手法を提案する．

2 機械学習を用いた行動認識手法の提案と設計

2.1 提案手法の概要

本研究で提案する人間の行動認識手法では，まず人間の行動が記録された動画を動画一つにつき一つの行動に分割した後にその一つの動画に対して以下の 4 つの処理を行い人間の行動に関する特徴量を抽出する．

1. 人の行動の動画を一つ読み込みフレームごとに画像に分割する
2. 連続するフレーム画像から背景差分法を用いて行動する人間に該当する部分を抽出する（図 1）
3. 抽出された人間部分の画像から人間の行動に関する特徴量を算出する
4. k-means 法を用いて前の処理で算出した特徴量の中から特に注目すべき特徴を抽出する

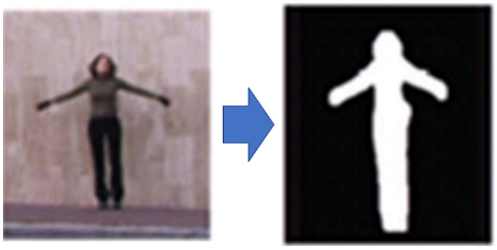

図 1: 背景差分法による人間部分の抽出例

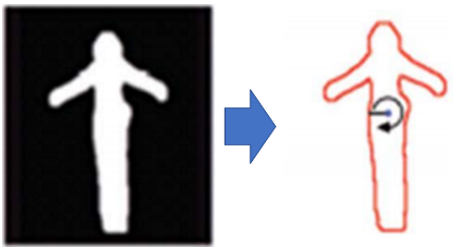

図 2: 人間部分のエッジ抽出例

以上の 4 つの処理により，人の行動が記録された各動画に対して大幅にデータ量を減らし，かつ特徴的なデータを得ることが可能となる．本研究ではこのようにして得られたデータをトレーニングデータとテストデータに分割し提案手法の認識精度の検証を行うが，その際のトレーニングデータとテストデータの比較には動的時間伸縮法：DTW（Dynamic Time Warping）を用いる．

2.2 特徴量の算出手法

人間の行動に関する特徴量を算出するために，まず動画からフレームごとに分割された画像から図 1 に示すように，人間に該当する部分を抽出する．この抽出した人間部分からさらに図 2 に示すような人間のエッジを算出する．次に，エッジ上に n 個の点 $p_i = (x_i, y_i)$ をとり，その点の座標から人間部分の重心点 $C = (x_c, y_c)$ を算出する（式 1）．

$$\begin{aligned} x_c &= \frac{\sum_{i=1}^{n} x_i}{n} \\ y_c &= \frac{\sum_{i=1}^{n} y_i}{n} \end{aligned} \tag{1}$$

次に，エッジ上の各点 p_i から重心点 C までの距離を d_i とし，その集合を DS と定義する（式 2）．

$$d_i = \|C - p_i\|$$

Activity Recognition from Move Data based on Machine Learning

*Guan Xiangran, Yasue Iijima, Graduate School of Engineering , Chiba Institute of Technology.

†Susumu Konno, Department of Advanced Media, Chiba Institute of Technology.

$$DS = \{d_1, d_2, d_3, \cdots, d_n\} \quad (2)$$

このように，一般的に大容量となる動画や画像データをエッジ上の各点から重心までの距離データの集合として表すことでデータ量を大幅に削減することが可能となる．また，エッジ上の点 p_i の個数を各フレーム間で統一することにより，動画に映っている人間のサイズなどに変化があっても特徴量を一定に保つことが可能となり多少の誤差は吸収することが可能となる．DS は各フレーム画像毎に算出するため，一つの人間の行動が撮影された動画に対してはそのフレームの数だけ算出されることになり，その集合を S と定義する．

$$S = \{DS_1, DS_2, DS_3, \cdots, DS_m\} \quad (3)$$

人間の行動が記録された動画のフレームの数は映っている人間の動きの速さや複雑さによって変化する可能性がある．そのため，各動画から得られた S_j 内の DS の個数も動画によって異なるため，比較を容易にするために個数の統一を行う．具体的には，S_j 内の DS に対して k-means 法 [2, 3, 4] を用いて任意の個数のクラスタに分類し，各クラスタの中心座標に最も近い DS を抽出する．これにより，全ての動画において S 内の DS の個数が統一できた $Skey$ を得ることができ，比較が容易になる．

このようにして作成した各動画の $Skey$ をトレーニングデータとテストデータに分類し，トレーニングデータのほうには各動画に映っている人間の行動をラベリングする．トレーニングデータとテストデータの比較には DTW を用いておこなう．

3 試作と実験

提案手法の効果を検証するために，KTH[5] と Weizmann[6] のデータセットを用いて認識精度を計測した．表 1，表 2 に実験結果を示す．

KTH は 25 人が 4 つのエリアで walking，jogging，running，boxing，handwaving，hand clapping の 6 種類の行動を撮影した動画である．KTH を用いて実験を行った結果，表 1 に示すとおりとなり，平均正解率は 94.47% であった．

Weizmann は 9 人が bend，jack，jump，pjump，run，side，skip，walk，wave1，wave2 の 10 種類の行動を撮影した動画である．Weizmann を用いた多くの実験では skip の認識精度が低いため検証対象とされていない場合が多いため，本研究でも skip は検証対象から除外した．Weizmann を用いて実験を行った結果，表 2 に示すとおりとなり，平均正解率は 91.36%であった．

4 おわりに

本研究では，人間の行動認識手法を提案し試作システムを用いてその認識精度の検証をおこなった．現状では

表 1: 実験結果（KTH）

	walking	jogging	running	boxing	handwaving	handclapping
walking	0.97	0.03				
jogging	0.03	0.94	0.03			
running	0.02	0.11	0.87			
boxing	0.01			0.99		
handwaving				0.07	0.93	
handclapping				0.07		0.93

表 2: 実験結果（Weizmann）

	bend	jack	jump	pjump	run	side	walk	wave1	wave2
bend	9/9								
jack		9/9							
jump			9/9						
pjump				9/9					
run					8/10	2/10			
side			2/9			7/9			
walk							10/10		
wave1		1/9						8/9	
wave2		2/9							7/9

90%を超える認識精度を得ているが，幾つかの状況（例えば走る速度が違うなど）で認識精度の低下がみられた．今後はこれらの問題点を解決しさらなる認識精度の向上を目指す．

参考文献

[1] Yiithan Dedeolu, B. Uur Treyin, Uur Gdkbay and A. Enis etin: Silhouette-based method for object classification and human action recognition in video, Computer Vision in Human-Computer Interaction, Springer, Berlin, Heidelberg, Volume 3979 of Lecture Notes in Computer Science, 64–77 (2006).

[2] Chaaraoui, A.A., Climent-Prez, P. and Flrez-Revuelta, F. : Silhouette-based human action recognition using sequences of key poses. Pattern Recognition Letters, 34, 1799–1807 (2013).

[3] Cheema, S., Eweiwi, A., Thurau, C. and Bauckhage, C. : Action recognition by learning discriminative key poses. IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, 6-13 November 2011, 1302–1309 (2011).

[4] Baysal, S., Kurt, M. and Duygulu, P. : Recognizing human actions using key poses. 20th International Conference on Pattern Recognition (ICPR), Istanbul, 23-26 August 2010, 1727–1730 (2010) .

[5] Recognition of human actions (online), ⟨http://www.nada.kth.se/cvap/actions/⟩ (accessed 2017-12-20).

[6] Actions as Space-Time Shapes (online), ⟨http://www.wisdom.weizmann.ac.il/vision/SpaceTimeActions.html⟩ (accessed 2017-12-20).