1X-06

人物の行動と周囲の特徴を用いた行動認識手法の提案

國藤　貴則 †　　真部　雄介 †　　菅原　研次 †
† 千葉工業大学情報科学部

1　はじめに

近年の画像処理分野の進歩により，行動認識に関する研究が盛んに行われている．文献 [1] によると人間の行動認識にはビデオの保存や検索，インテリジェントビデオ監視，家庭環境モニタリングなど幅広い用途がある．しかしながら正確で効率的な人間の行動認識は難しいテーマとなっている．

既存の行動認識手法は，人物の行動特徴に着目した手法と行動特徴以外に着目した手法の 2 つに分けられる．前者の手法として，ディープラーニングを用いる手法や骨格の軌跡を用いる手法，時空間特徴を用いる手法などが提案されている [1]．後者の手法として，空間内に存在する物体を用いて人物の行動を特定する手法 [2] や，人物の位置情報と家電につけられた消費電力センサを用いる行動認識手法 [3] などが提案されている．

以上を踏まえると，より正確な行動認識を実現するためには，人物の行動特徴に着目するだけでなく，行動が行われる周辺の情報にも着目した行動認識アプローチが必要だと考えられる．

そこで本研究では，人物の行動と周囲の特徴を用いた行動認識手法の提案を行う．本研究では周囲の特徴として，行動を行っている人物周辺に存在する物体と Activity Space（AS）を用いる．AS とは，物体と人間の行動の関係を用いた活動空間の概念であり，複数の行動が行われる空間と定義する．また，人物の行動は CNN ベースの手法を利用する．

2　提案手法

提案手法の概要と詳細について述べる．

2.1　概要

本研究における提案手法の流れを図 1に示す．大きく分けて CNN を用いた行動認識，AS の認識，人物最近傍物体認識を行う 3 つの流れがある．

CNN を用いた行動認識では，既存の畳み込みニューラルネットワーク [4] を用いて行動認識を行う．

AS の認識では，物体認識器を用いて物体認識を行い，認識器を用いて AS 認識を行う．

最近傍物体認識では，物体認識器を用いて物体認識を行い最近傍物体を求める．

最後に全ての出力統合した多次元ベクトルを作成し，識別器であるランダムフォレストの入力にすることで，最終的な行動認識を行う．

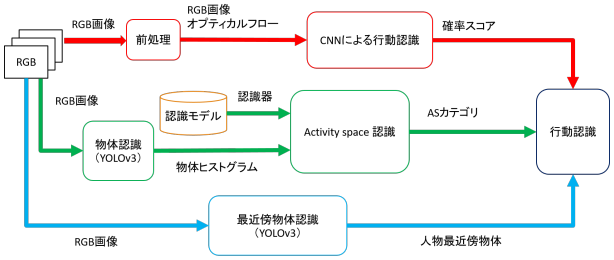

図1　提案手法の流れ

2.2　CNN による行動認識手法

入力画像に対して行動認識を行うため，CNN の学習を行う．学習する CNN として Simonyan らの Two stream CNN[4] を用いる．これは静止画像を学習し，空間的特徴を学習する Spatial ConvNet と，複数枚のオプティカルフローを学習し，時間的特徴を学習する Temporal ConvNet の 2 つの CNN を持ち，それぞれ畳み込んだ結果を統合することによる行動認識を行う手法を提案している．それぞれの CNN は 5 層の畳み込み層と 2 層の全結合層からなる CNN となっている．本研究では，使用するオプティカルフローとして，Bi-directional optical flow を用いる．

2.3　Activity Space 認識手法

AS とは，物体を伴った特定の行動が発生する部分空間のことを指し，人物の周辺に存在する物体によって特徴づけられる．例えば，「キーボード」や「マウス」，「PC」が存在する空間では，「文字を打つ」，「マウスを操作する」，「ディスプレイの電源を入れる」といった行動が発生すると考えられる．したがって，本研究では，行動周辺に分布する物体の傾向によって行動をクラスタリングし，AS を作成する．

AS の認識を行うために，人物の行動を，クラスタに分割する必要がある．

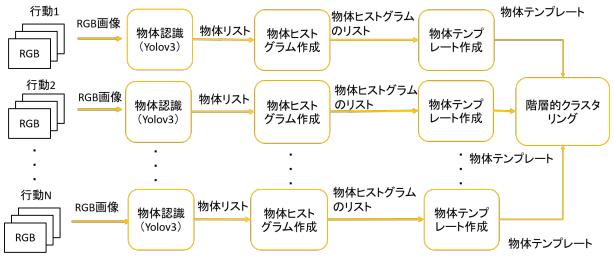

図2　AS の作成の流れ

図 2に AS の作成の流れを示す．まず，物体認識器を用いて行動ごとに物体認識を行う．物体認識は，1 つの行動を記録した動画の各フレームに対して行い，物体ヒストグラムを得る．その際に，人物領域の中心座標から左上の座標との距離より短い範囲にある物体を用いてヒストグラムを作成する．物体ヒストグラムの全フレームに対する平均を 1 つの行動に対するテンプレートとす

Activity recognition method by using human activity and its ambient features
†Takanori KUNITO　†Yusuke MANABE　†Kenji SUGAWARA
†Faculty of Information and Computer Science, Chiba Institute of Technology

る．次に，全ての行動から得られたテンプレートを階層的クラスタリングを用いて分類する．そして，分類の結果得られたクラスタを AS とする．本研究では，階層的クラスタリング手法として Ward 法を用い，クラスタ数を 5（AS は 5 種類）とした．

学習では，入力として物体ヒストグラムを与える．教師データとしてこれに AS ラベルを付与する．認識では，入力された動画に対してフレーム単位で物体認識を行い，物体ヒストグラムを作成し，認識器に入力することで AS の認識を行う．本研究では，AS 認識器として，ランダムフォレストを用いる．

2.4 最近傍物体認識手法

最近傍物体の認識方法は，物体認識を行い，人物領域を求め，人物の上半身の重心を求める．求めた重心と検出された物体の中心座標との距離を求め，距離が一番近いものを最近傍物体とする．

2.5 最終的な行動認識手法

CNN を用いた行動認識，AS の認識，人物最近傍物体の出力結果を 1 つのベクトルにし，ランダムフォレストの入力とする．教師データは，用意したデータセットの入力ベクトルに対してデータセットの正解ラベルを付与する．認識では，CNN を用いた行動認識，AS の認識，人物最近傍物体認識の 3 つの出力を 1 つのベクトルにし，認識を行う．

3 評価実験

データセットと実験，結果の概要について述べる．

3.1 データセットの概要

データセットとして Watch-n-Patch[5] を用いて評価実験を行う．これは Kinectv2 カメラで記録され，21 クラスの行動を含み，合計約 230 分で 458 本のビデオで構成されているデータセットである．

3.2 実験概要

提案手法の有効性を示すため，データセットを学習データとテストデータとして 7 対 3 で分割し，テストデータに対する精度を評価する．その際に，CNN を用いた行動認識のみ用いる場合と人物行動特徴と人物周辺の情報を用いる場合で実験を行い，精度の比較を行う．評価基準として，適合率，再現率を用いる．本研究で用いる物体認識器は 80 クラスを検出可能な YOLOv3[6] を用いて実験を行なった．

3.3 実験結果

表 1 に CNN を用いた行動認識のみ用いる場合と人物行動特徴と人物周辺の情報を用いる場合の精度の比較を示す．提案手法の全体的な認識率は，72%と CNN のみの場合と比較し，6 ポイント上昇した．また再現率を見ると，「microwaving」,「pouring」,「leave-kitchen」,「leave-office」,「play-computer」は下がったが，他の行動は改善され，最大で「turn-off-monitor」が 37 ポイントと大きく上昇している．そして，CNN では「fill-kettle」など一部のキッチンで行われる行動が「leave-office」と認識される場合があったが，提案手法では，kitchen で行われる行動が office で行われる行動と間違えることがなくなり，AS を用いることで精度が改善されることを示した．

表1　実験結果

	CNN		提案手法	
	適合率	再現率	適合率	再現率
fetch-from-fridge	0.60	0.56	0.61	0.73
put-back-to-fridge	0.46	0.32	0.44	0.33
prepare-food	0.79	0.68	0.75	0.80
microwave	0.60	0.67	0.62	0.62
fetch-from-oven	0.63	0.50	0.61	0.64
pouring	0.78	0.91	0.86	0.88
drinking	0.83	0.35	0.77	0.53
leave-kitchen	0.29	0.73	0.68	0.64
fill-kettle	0.93	0.77	0.91	0.80
plug-in-kettle	0.90	0.83	0.93	0.92
move-kettle	0.82	0.63	0.76	0.85
reading	0.92	0.94	0.91	0.97
walking	0.79	0.77	0.78	0.85
leave-office	0.56	0.89	0.79	0.85
fetch-book	0.51	0.54	0.48	0.67
put-back-book	0.00	0.00	0.36	0.29
put-down-item	0.50	0.49	0.55	0.55
take-item	0.56	0.03	0.53	0.16
play-computer	0.84	0.93	0.92	0.92
turn-on-monitor	0.72	0.18	0.69	0.44
turn-off-monitor	0.00	0.00	0.42	0.37
平均	0.66	0.56	0.72	0.66

4 おわりに

提案手法の全体的な認識率は，72%と CNN のみの場合と比較し，6 ポイント上昇し，行動認識を行う際に，周囲の特徴を付与することの有効性を示すことができた．今後の展望として，精度のさらなる向上があげられる．より多くの物体を検出可能な物体認識器を用いることや，人物がどのような姿勢をしているかなど人物の特徴量を増やすこと，物体を保持しているかしていないかなど物体の位置関係の情報を増やすなど情報を増やすことで，より高精度な行動認識ができるようになると考えられる．

参考文献

[1] H. Zhang et al.: A comprehensive survey of vision-based human action recognition methods. Sensors, vol.19 issue.:5, 1005, 2019.
[2] N. Yamada et al.: Applying ontology and probabilistic model to human activity recognition from surrounding things. IPSJ Digital Courier 3, pp. 506–517, 2007.
[3] 上田他: ユーザ位置情報と家電消費電力に基づいた宅内生活行動認識システム. 情報処理学会論文誌, vol.57.2, pp.416–425, 2016.
[4] K. Simonyan et al.: Two-stream convolutional networks for action recognition in videos. Advances in neural information processing systems, pp. 568–576, 2014.
[5] C. Wu et al.: Watch-n-patch: Unsupervised understanding of actions and relations. Proc. IEEE Conference on CVPR, pp. 4362–4370, 2015.
[6] J. Redmon et al.: YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.