社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# 物体と動き特徴を用いた行動認識

勝手　美紗†　　内海ゆづ子†　　黄瀬　浩一†

† 大阪府立大学 大学院工学研究科　〒 599–8531 堺市中区学園町 1–1
E-mail: †katte@m.cs.osakafu-u.ac.jp, ††{yuzuko,kise}@cs.osakafu-u.ac.jp

**あらまし**　高齢化が急速に進む中で，家に設置したカメラを用いて遠距離から高齢者を見守るシステムが注目を集めている．本稿では，見守りシステムでの高齢者の行動の詳細を獲得，提示することを目的とし，動画像からの人物の行動を検出する．人物の行動は物体などの周りの環境を伴って行われるものが大半である．そこで，物体検出と人物の動きを用いた行動認識手法を提案し，人物の動きだけでなく，人物の周りの物体を検出し，動きと周りの環境の相互の情報を用いて行動の認識を行う．人物の動きは，動画中における局所特徴量の特徴点の配置情報を特徴量として，周辺の環境は，人の周りで検出した物体の位置や大きさの比の情報を特徴量とする．認識には，動き，物体の特徴量を結合し，最近傍探索により行う．

**キーワード**　行動認識，物体検出，Spatio-temporal 特徴量

## 1. はじめに

今日の核家族化による一人世帯の高齢者の増加に伴い，高齢者を遠方から見守るシステムへのニーズが高まっている．実際に，カメラ [1] を用いた見守りシステムが実用化されている例もある．カメラを用いた見守りシステムは，ユーザが必要な時間に，部屋に設置したカメラにより撮影された高齢者の生活の様子を観察することができるため，高齢者の室内での具体的な行動を知ることができる．しかし，カメラで撮影した一日分の動画像から高齢者の様子を確認するためには，ユーザが撮影したすべての動画像に目を通す必要があるため，多くの手間と時間が必要である．この問題の解決策の一つに，動画像中の見守る対象者の一日の様子を，短時間で知ることが出来るように要約し報告することが考えられる．この機能は，動画像中の人物の行動を認識し，その結果を時系列毎にまとめることで実現できる．そこで，本研究では動画像中からの人物の行動を認識し，提示することを目的とする．

人物の行動は，物体を用いて行われるものが多く，その動作で用いられている物体によって動作の目的は大きく異なる．例えば，人物が椅子に座り腕を動かしている時に，本を持っていることが分かっていればその人物は読書をしているといえるが，箸を持っていれば食事をしているといえる．このように，人の行動を認識するためには，人の動作だけではなく，その動作で使われている物体や周りの環境も考慮に入れることが重要である．

これまでの行動認識の研究では，様々な手法で行動の情報を抽出していた．オプティカルフローを用いて，人全体の動きや形の情報を特徴量として用いるものや [2]，特徴点周りから輝度値を利用し特徴量を抽出するものがある [3]．しかし，これらの研究では人の動きの情報のみが使われ，物体の情報は反映されていなかった．そこで，本稿では人物の動きの特徴だけでなく，その行動に用いられる物体や周りの環境の情報を用いて，行動を詳細に認識する手法を提案する．提案手法では，人の動きと周辺の環境を用いて行動を認識するために，動画像から人の動き，周辺の物体特徴を抽出し，最近傍探索を用いて行動の認識を行う．以降，2 章で特徴抽出手法，3 章で認識手法について説明する．4 章で実験について述べ，5 章で本稿をまとめる．

## 2. 動きと物体の特徴量

本稿では，動きの特徴量を Ryoo [3] らの手法により抽出し，人物の周辺にある物体の特徴量は Sadeghi [5] らの手法により抽出する．行動の認識に用いる特徴量は，これら二つの特徴を結合したものとする．

### 2.1 動きの特徴量

動きの特徴抽出は，まず，動画像から spatio-temporal 特徴 [4] と呼ばれる局所特徴量を抽出する．特徴量に対してクラスタリングを行い，クラス間の時間軸方向，画像平面上での位置関係を求める．この位置関係を特徴量とする．

#### 2.1.1 Spatio-temporal 特徴の抽出

spatio-temporal 特徴 $\boldsymbol{f} = (\boldsymbol{f^{des}}, \boldsymbol{f^{loc}})$ は，画像特徴量 $\boldsymbol{f^{des}}$ と画像平面上の位置 $x, y$ と時刻 $t$ を表す $\boldsymbol{f^{loc}} = (x, y, t)$ の二つの要素からなる．図 1 に特徴点の抽出の流れをまとめたものを示す．特徴点 $f^{loc}$ は，式 (1) で示すように，動画像 $I(x, y, t)$ に対して 2D Gaussian spatial kernel $g(x, y; \sigma)$ と 1D Gabor temporal filter $h_{ev}(t; \tau, \omega) = -\cos(2\pi t\omega)e^{-t^2/\tau^2}$, $h_{od}(t; \tau, \omega) = -\sin(2\pi t\omega)e^{-\tau^2/\tau^2}$ の畳み込み積分の極値として得られる．

$$R = (I * g * h_{ev})^2 + (I * g * h_{od})^2 \tag{1}$$

ただし，$\sigma$ は画像平面上でのスケールを決めるパラメータ，$\omega$ は $1/\tau$ とし，$\tau$ は時間軸に対するスケールを決めるパラメータである．画像特徴量 $\boldsymbol{f^{des}}$ は，検出された特徴点の 3 次元の

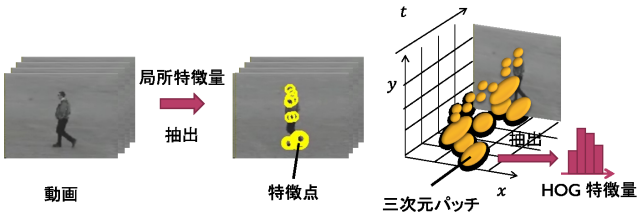


図 1　動画像から局所特徴量の抽出

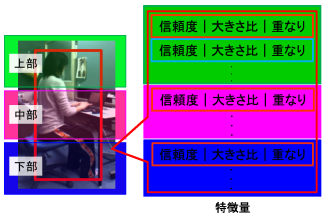


図 2　人物の周辺からの特徴抽出

パッチから HOG 特徴量を抽出することで得られる．

### 2.1.2　関係の構築

抽出された特徴量 $\boldsymbol{f_1}, \ldots, \boldsymbol{f_n}$ を，$\boldsymbol{f^{des}}$ を用いて k-means クラスタリングにより $k$ 種類に分類する．次に各クラスタ間で，特徴点 $\boldsymbol{f_1^{loc}}, \ldots, \boldsymbol{f_n^{loc}}$ の 3 次元パッチの時間軸での前後関係や画像平面上での位置関係を求めて頻度のヒストグラムで表す．時間軸方向では，3 次元パッチ同士の時間軸方向における位置関係を 13 通り定義し，画像平面上では，特徴点の 2 点間の距離の関係を 4 通り定義し，ヒストグラムを作成する．

### 2.2　環境の特徴量

環境の特徴量は，人物のまわりで物体検出を行い，検出器の出力値や物体と人との相対位置，人物領域との面積比，人物領域と物体領域の重複した部分の面積比を用いて表す．物体検出には，Felzenszwalb ら [6] の手法により学習した検出器を用いる．特徴抽出の手順は，まず，人物の動きの特徴量が抽出された領域を求め，その周辺の領域を，図 2 のように上部，中部，下部の 3 つに分割する．この 3 つの領域から使用する全ての物体検出器を用いて物体を検出する．そして，各検出器で最も出力値が高い領域を求め，その領域と人物領域との面積比，人物領域と物体領域が重なっている割合を求める．

## 3.　認 識 処 理

認識は，最近傍探索により行う．学習用の動画像から特徴量を抽出し，データベースに登録する．クエリの動画像からも同様に特徴量を抽出する．クエリの特徴量と，データベースに登録した特徴量とのユークリッド距離を求め，最近傍となる特徴点の行動を認識結果とする．

## 4.　実　　験

提案手法が行動認識に有効であるか評価するために実験を行った．実験に用いた動画像は，床から 1 mの高さにカメラを固定し，撮影を行った．動画像は，1024×768 pixels, RGB カラー画像で，フレームレートは 15fps である．1 人の人物の 5 つの行動（掃除機をかける，パソコンを使用する，お茶を飲む，読書，食器洗い）をそれぞれ 15 秒間，4 セット撮影した．データベースには，2 セット動画を登録し，クエリには，残りの 2 セットを用いた．

表 1　行動認識の結果

| 行動 | 認識率 | |
|---|---|---|
| | 動き特徴量のみ | 提案手法 |
| 掃除機をかける | 2/2 | 2/2 |
| パソコンを使用する | 1/2 | 1/2 |
| お茶を飲む | 1/2 | 2/2 |
| 読書 | 1/2 | 2/2 |
| 食器洗い | 2/2 | 2/2 |

動き特徴量を抽出したときのクラスタ数 $k$ を 50 とし，物体検出器は Pascal VOC2008 dataset [7] で学習したもの 20 個を用いた．また動画中の人物の領域は既知であるとし，周辺の物体特徴量は図 2 で示したように人物領域の近傍から抽出した．

動き特徴量のみを用いて認識した結果と，提案手法により認識した結果を表 1 に示す．掃除機をかけるや食器を洗うなどの動作の規模が大きい行動は，ほかの 3 つの行動と比較して，動きが異なるため，動き特徴量のみで認識できたものと考えられる．一方で，お茶を飲んだり，パソコンを使用したり，読書などの動作が似ている行動は，動き特徴量のみでは認識が困難である．これらの動きは，周辺の物体の情報を付加することで認識率が向上した．

## 5.　ま　と　め

本稿では，動画像から人物の行動を認識することを目的とし，人物の動きと人物のまわりの環境の情報を用いて行動認識する手法を提案した．認識実験では，動きのみの特徴量で認識した結果と比較し，提案手法により認識率が向上したことが示され，周辺の物体を考慮し行動を認識することが有効であることが明らかとなった．今後の課題として，データベースの規模を拡大し，提案手法の有効性の評価を行うことが挙げられる．また，動画像中の人物領域を自動で検出することも今後の課題である．

## 文　　献

[1] http://www.secom.co.jp/lp/hs/afv1/.
[2] A. Fathi and G. Mori, “Action recognition by learning mid-level motion features,” Proc. of CVPR, pp.1–8, 2008.
[3] M.S. Ryoo and J.K. Aggarwal, “Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities,” Procc. of ICCV, pp.1593–1600, 2009.
[4] H. Wang, M.M. Ullah, A. Klaser, I. Laptev, and C. Schmid, “Evaluation of local spatio-temporal features for action recognition,” Proc. of BMVC’09, 2009.
[5] Sadeghi, M. Amin, and A. Farhadi, “Recognition using visual phrases,” Proc. of CVPR, pp.1745–1752, 2011.
[6] P.F. Felzenszwalb, R.B. Girshick, and D. McAllester, “Discriminatively trained deformable part models, release 4,” http://people.cs.uchicago.edu/ pff/latent-release4/.
[7] Everingham, M., V. Gool, L., Williams, C.K. I., J. Winn, and A. Zisserman, “The PASCAL Visual Object Classes Challenge 2011 (VOC2011) Results,” http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html.