I-057

ウェアラブルカメラによる人物行動の認識と複数メディアを用いた要約表現
Multimedia Summary of Human Behavior Using Wearable Camera

青木 茂樹† 加藤 亮一‡ 小島 篤博‡ 福永 邦雄‡
Shigeki Aoki Ryoichi Kato Atsuhiro Kojima Kunio Fukunaga

1. まえがき

マン・マシンインターフェースの改善や人間とロボットの円滑なコミュニケーションの実現を目指して，人物の動作やジェスチャを認識する研究が盛んに行われている．人物の動作やジェスチャを認識する手法は数多く提案されているが，これまでに提案されている認識手法の多くは，動作や行動の認識に主眼が置かれ，認識結果を一つのシンボルとして提示する手法が採られている．このため，認識結果を人に分かりやすく提示する方法として適切であるとは言えなかった．

本稿では，小型軽量なウェアラブルカメラを用いて，装着者の動作・行動を認識することにより，装着者の位置に関わらず動作や行動を認識し，複数のメディアを用いて認識結果を表現することによって，装着者の行動を人に分かりやすく出力する手法を提案する．まず，ウェアラブルカメラから得られる入力画像から装着者の移動情報，手の動きなどを抽出する．次に，抽出した情報を用いてイベントを検出し，階層化した動作のモデルに基づいて人物の動作を認識する．その後，連続して行われた動作の関連性に注目して装着者の動作を要約し，複数のメディアを組み合わせて要約結果を生成する．

2. 動作特徴の抽出

ウェアラブルカメラから得られる映像を用いて，装着者が行った動作を認識する．本研究では，常に装着者の正面を向いている位置として胸の部分 (図 1 (a) 中の丸印の位置) にカメラを取りつけた．ウェアラブルカメラの拡大図を図 1 (b) に示す．この位置にカメラを取りつけることによって，人物の移動量や手領域の移動方向を安定して抽出することができる [1].

2.1 移動情報の抽出

ウェアラブルカメラから得られる映像は，装着者の動きに対応してカメラが移動するため，映像の変化を計測することで装着者の移動量を推定することができる．本研究では，カメラの動きによるフレーム間の画像の変化をアフィン変換で表せると仮定し，フレーム間のアフィン変換パラメータを統計的な最適解探索手法である CONDENSATION アルゴリズムを用いて逐次求めていくことによって人物の移動量を推定する．

CONDENSATION アルゴリズムを用いた最適なアフィン変換パラメータの探索は，パラメータを乱数で生成して過去に得られた基準フレームとの誤差が小さいパラメータを選択することによって行うが，連続するフレーム間では人物は同様の動きをしていることが多いと考えられるため，パラメータ値が大きく変化することは少なく，前フレームで推定したパラメータと近い値である可能性が大きいと考えられる．そこで，前フレームで推定したパラメータから誤差が小さいものを選択し，誤差が小さいパラメータの周囲で数多くのパラメータを生成することにより，効率よく最適解を探索する．ただし，人物が大きく移動した場合には，単一の基準フレームでは人物の動作に追従できない場合が考えられるため，最小の誤差がしきい値以上になった場合には，人物の位置が変化したものと考えて，基準画像をそのときに得られている入力画像に更新する．以上の処理を連続するフレームに順次適用して，アフィン変換パラメータを求め，人物の移動量を推定する．

(a) 装着位置 (b) 拡大図

図 1: ウェアラブルカメラ

2.2 肌色情報の抽出

入力画像中の肌色領域を抽出することによって，装着者と会話などをしている対象者の顔領域や，装着者の手領域などを抽出する．本研究では，複雑な背景環境下でも安定して肌色画素を抽出することができる手法 [2] を用いて肌色画素を抽出する．

抽出した肌色画素をクラスタリングして得られる各肌色領域に対して，楕円領域との類似度を計測し，楕円に類似していてかつ，輝度値の低い画素がしきい値以上の割合で含まれている領域を対象者の顔領域として抽出する．そして，それ以外の肌色領域を装着者の手領域とする．ただし，面積がしきい値未満の領域はノイズとして除去している．

2.3 対象物体領域の抽出

装着者が物体を手に持って動かした場合，物体も画像上を移動するため，物体領域の背景確率は低くなる．そこで，背景確率と肌色確率の両方が低い画素を抽出して，クラスタリングすることによって前景領域を抽出する．このようにして抽出した前景領域には装着者の腕などの部分が含まれる可能性が考えられるため，抽出した前景領域の中で左手の重心より右上，かつ右手の重心より左上に存在する前景領域を対象物体領域として選択する．

3. 動作の認識

3.1 イベントの検出

2. で抽出した動作特徴の変化をイベントとして検出する．イベントとしては，装着者の「移動／静止」に関

†熊本電波高専, Kumamoto National College of Technology
‡大阪府立大学, Osaka Prefecture University

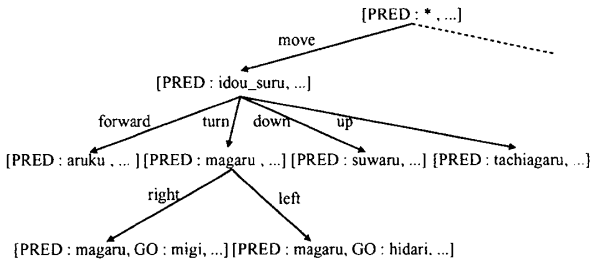

図 2: 動作のモデル

するもの,「移動方向」に関するもの,「手の動き」に関するもの,「対象者」に関するもの,「物体」に関するものの 5 種類に分けて考える.

1. 装着者の移動／静止に関するイベント

装着者が移動している場合，映像の変化が激しくなるためにアフィン変換するときの基準画像の更新頻度が大きくなる．そこで，基準画像の更新頻度がしきい値以上の場合にイベント “move” を出力し，それ以外の場合にイベント “not move” を出力する．

2. 装着者の移動方向に関するイベント

装着者の移動方向はアフィン変換パラメータの変化として検出することができる．パラメータの変化を “増加”, “減少”, “停滞” の三つの状態に分けて考え，“増加” から “減少”, “停滞” など状態が変化した場合にイベントを出力する．

3. 装着者の手の動きに関するイベント

手の動きに関するイベントも，手の移動方向が変化したときに出力する．ここでは，ノイズ除去のため，映像上での手の移動方向を 8 方向に区分してイベントを検出している．

4. 対象者に関するイベント

人物の顔領域を抽出した場合，イベント “meet(man)” を出力する．また，人物領域を一定時間以上抽出できなくなった場合，“part(man)” を出力する．

5. 物体に関するイベント

物体領域を抽出した場合，イベント “have(obj)” を出力する．また，物体領域が画面上から消失した場合，イベント “pick(obj)” を出力する．

3.2 装着者の動作の認識

人物の動作に関わる動作概念との対応を考慮して動作を階層的に表現した図 2 に示すモデルの階層を，前節で検出した人物の移動量や手の動作軌跡などの動作特徴のイベントをもとに下っていくことによって，装着者の動作を認識する [1].

4. 行動の認識

前節までで認識した装着者の動作の時間的な連続性に注目して，行われた動作を要約することによって行動を認識し，認識結果を複数のメディアを用いて表現する．

人物行動の要約を考えるとき，どの程度の時間間隔での行動を要約するかを決定する必要があると考えられる．本研究では，ある一定の位置に留まっている時間に注目して，位置情報の変化が少ない時間間隔での装着者の動作を要約する．

4.1 位置情報の抽出

人物の位置情報は，GPS (Global Positioning System) 受信機から得られる緯度，経度の情報と，あらかじめ与えておいたマップ情報とを比較することにより抽出する．しかし，屋内では GPS 信号を受信できないため，GPS 受信機のみで装着者の位置情報を常に安定して抽出することは難しい．そこで，**2.1** で求めた移動量がしきい値未満のときに得られる画像列の平均を取ることによって得られる，装着者の現在の位置を表す画像と，過去に同様の手法で作成して保存しておいた画像 (代表画像) を比較する [3] ことによって，GPS 受信機からの情報が得られない場合にも，装着者の位置情報を安定して抽出する．ここで，動作を要約する時間間隔を選択できるようにするために，装着者の位置情報を，例えば “位置 C” のように最も詳細な位置情報のみで表すのではなく，“A 大学内の B 図書館の位置 C” のように階層的に表現している．また，屋外などで GPS 受信機が使用できる環境では，装着者の移動量の抽出に GPS 受信機から得られる情報を補助的に用いている．

4.2 装着者の動作の要約と複数メディアによる表現

一定の位置に留まっている時間内に観測された動作を，それぞれの動作に分類して，それらの観測回数を計測する．例えば，図書館で移動に関する動作が多く観測されている場合,「図書館で本を捜した.」のように推論して装着者の動作を要約する．また，食堂で物体領域を検出し，手を縦に移動する動作の回数が多く観測された場合,「食堂で食事をした.」と推論する．ここで，単に食事をしたという記述が得られても,「何を食べたのか」など詳細な情報が分からないため，そのときに得られている物体領域の画像を加えることによって，どのようなものを食べたのかを表現し，更に「食事をしている」時間に得られている映像を加えることによって，どのように食事したかを表現する．

5. 実験

本手法の有効性を確かめるために，実験を行った．実験では，大学生が通常の生活で行う行動を想定して設定したシーンの行動を行い，動作の認識，行動の要約，複数メディアを用いた要約表現が生成できることを確認した．

6. むすび

本稿では，人物に装着したウェアラブルカメラから得られる情報を用いて，人物の動作を認識し，観測された動作の時間的な連続性に注目して行動を要約し，認識結果を複数のメディアを用いて表現する手法を提案した．今後の課題としては，要約する時間間隔の検討などが挙げられる．

参考文献

[1] 青木，高橋，大西，小島，福永，“ウェアラブルカメラによる動作の認識とテキスト表現，” 信学技報，PRMU2003-67, July 2003.

[2] 浅沼，大西，小島，福永，“色情報と領域追跡情報を用いた人物の顔と手の領域の抽出，” 電学論 (C), Vol.119-C, no.11, pp.1867–1875, Nov. 1998.

[3] 加藤，青木，小島，福永，“ウェアラブルカメラ映像に基づく行動パターンの学習と認識，” 第 2 回情報科学技術フォーラム講演論文集，I-041, pp.91-92, Sep. 2003.