特定空間における人の行動の言語化手法の提案

落合 恵理香　(指導教員：小林 一郎)

1　はじめに

近年，人が動画像を撮影する機会が増加しており，多量の動画像の管理が必要とされている．動画像内に現れる人の振る舞いを検索する場合，現状では撮影された内容を人が確認しながら探すことしかできないため，人的負担が大きい．そこで，自然言語により人の行動を検索することができれば，人的負担が軽減される．

本研究では，カメラから取得された動画像に対して画像処理を施し得られた特徴データと物体に取り付けられたセンサから取得されるデータを用い，人と物の関わりに着目することにより特定空間内での物体に対する人の振る舞いを観察する．これを基に，人の行動を言葉で説明する手法を提案する．

この手法をもって動画像内の人の振る舞いを検索できるシステムの構築を目指す．

2　言語化システムの構築

2.1　動画像からの特徴データ抽出

本研究では，画像処理に Intel 社が公開している画像処理ライブラリである OpenCV[1] を用いる．

OpenCV に用意されている背景差分法，輪郭抽出を用いて人を認識する．背景差分法を用いることにより，動画像ファイルの初期画像と入力画像の差分によって得られた領域を人の領域として捉え，輪郭抽出を行う．人として捉えた領域の面積の重心を求め (図 1 右)，人を表す特徴データとして扱う．この特徴データと空間内の物体との関係を観察することにより，人の行動の言語化を行う．

図 1: 入力画像 (左) と輪郭抽出画像 (右)

2.2　物体に関する知識作成

人の振る舞いを画像内の物体との位置関係により捉えるため，空間内の物体を定義する．

本研究では，撮影された動画像ファイルの初期画像を「原画像」と呼び（図 2），原画像内に存在する物体の座標値をマウスで指定することにより物体を定義する（図 3）．

次に，取得された物体の座標値を用いて，定義物体のマスク画像を作成する（図 4 左）．人と物の関わりは物体付近で生じることを考慮して，物体の存在領域を意図的に広げるため膨張処理を施し（図 4 右），実際の物体より大きい領域を作成することにより，人と物との関わりを正確に捉える可能性を広げる．

白の領域が定義物体内，黒の領域が定義物体外に相当し，以下，白の領域である定義物体内のことを，指

図 2: 原画像

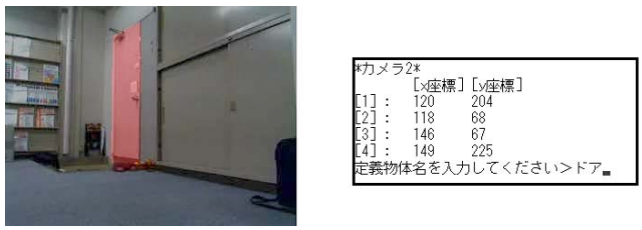


図 3: 物体指定画面

定した物体の「定義域」と呼ぶ．

このマスク画像は，特徴データ抽出の際に人の動作として取得される領域の重心座標が，定義された物体の領域内に含まれるかを判断するために使用する．マスク画像名と定義物体名をファイルに保存することにより知識を作成する（図 3 右：入力画面）．

2.3　出力文の構成

本研究では，出力文を Fillmore の格文法 [2] を基に構築する．

物体との関わりに対する人の振る舞いを言語化するため，深層格における「動作主格」,「対象格」に限定して考え,「人が，何を，どうする」という形をとることとする．本研究では，物体として「ドア」を対象にし，ドアと関わる人の振る舞いに対する文を出力する．

京都大学で開発された大規模格フレーム [3] を用い,「ドア」に関する動詞を選定する．名詞「ドア」から検索することにより，頻度 3000 以上で，ヲ格の最上位に「ドア」が存在し，ガ格の上位 3 個以内に「人」が存在するという条件を満たした動詞を出力文に使用される動詞として定めた．結果，3 個の動詞 { 開ける，閉める，ノック } を選定した．

図 4: 膨張処理前 (左) と膨張処理後 (右) のマスク画像

2.4　ベイジアンネットワークを用いた人の行動判定

本研究では，グラフ構造を持つ確率モデルの一つであるベイジアンネットワーク [4] を用いて，人の振る舞いを判定するモデルを作成する.「ドアを開ける」とい

う人の振る舞いについてのモデル例を図 5 に示す．

まず，画像処理から人の行動として捉えられた画像領域の重心座標に着目する．一定時間，重心座標が指定した物体の定義域に入った場合に，状態を「1」とし，それ以外の場合もしくは定義域内に重心が入る時間に開きが生じた場合は状態「0」とする．先行研究[5] において課題とされていた，一つの動画像ファイルのみを用いた場合での人と物体との関わりの誤認識による言語化の不正確さを改善するために，異なる角度から撮影された二つの動画像ファイルを用いて人の動作を判断する．それぞれのファイルから得られる事象を図 6 のノード X_1，X_2 とする．

次に，センサからのデータについて考える．センサをドアに取り付け，取得された値が閾値を超えた場合に状態を「1」とし，それ以外の場合もしくは値が閾値を超えた時間と開きが生じた場合は状態「0」とする．この事象を図 6 のノード Y とする．

「ドアを開けた」と判定される場合を状態「1」，判定されない場合を状態「0」とし，各ノードの条件付確率分布表は予め付与するものとする．

「ドアを開けた」と判定される場合の確率が，出力されない場合の確率よりも高い場合に「人 が ドア を開けた」と言語化が行われる．

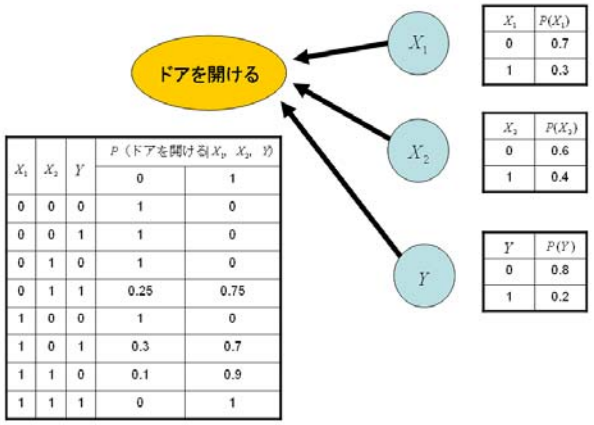


図 5: 行動判定のモデル（「ドアを開ける」）

3　実験と考察

このシステムを用いて，特定空間内での人の行動を言語化する実験を行った．

実験環境は，二台のネットワークカメラ[1]を用い，これらを別々の位置に固定し，異なる角度から人の行動を撮影した．物体として,「ドア」を定義し，人がドアを開けて入室するという行動を撮影した動画像ファイルを用いた．言語の出力結果とその時間における入力画像と輪郭抽出後の画像を表 1 に示す．

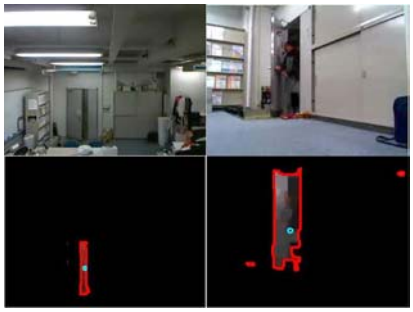
今回用いた動画像は，ドアに対する人の振る舞いが写されているものであり，それに対する言語化を行った．画像とセンサから得られる異なる 2 種類のデータから人の行動を推定するモデルを構築し，これを用いて言語化することができた．

[1]Panasonic 社 BB-HCM715

表 1: 言語化結果

時刻：0

時刻：148 「人 が ドア を 開けた」



さらに，人の振る舞いの前後関係をモデル化していくことで，より詳細に人の行動を言語化することが可能になると考える．

4　おわりに

本研究では, 動画像ファイルに対して，画像処理技術を施し，特定空間内に存在する物体に対する人の振る舞いを言葉で説明する手法を提案した．具体的には，動画像から取得されたデータとセンサから取得されたデータを，ベイジアンネットワークを用いることにより，異なる 2 種類のデータから言語を推定するモデルを構築した．今後は，人の振る舞いの前後関係をモデル化していくことにより，動画像に対して，より詳細な言語表現の付与を行っていく．

参考文献

[1] OpenCV，http://opencv.jp/

[2] Fillmore, CHARLES J.，”The Case for Case”，In Bach and Harms(Ed.)，Universals in Linguistic Theory，pp.1-88，New York：Holt，Rinehat，and Winston，1968.

[3] 河原大輔，黒橋禎夫，高性能計算環境を用いた Web からの大規模格フレーム構築，情報処理学会研究報告. 自然言語処理研究会報告，2006(1)，pp.67-73，2006

[4] 本村陽一，佐藤泰介，ベイジアンネットワーク - 不確定性のモデリング技術 -，人工知能学会誌，15(4)，pp.575-582，2000

[5] 能見麻未，小林一郎，特定空間における人と物のインタラクションの言語化，第 1 回データ工学と情報マネジメントに関するフォーラム（DEIM2009），E3-1，2009

[6] 小島篤博，田原典枝，田村武志，福永邦雄，動画像における人物行動の自然言語による説明の生成，電子情報通信学会論文誌（D-II），Vol.J81-D-II，No.8，pp.1867-1875，1998

[7] 亀井剛次，柳沢豊，前川卓也，岸野泰恵，櫻井保志，須山敬之，岡留剛，実世界イベント理解に向けた語彙集合の構築と評価，情報処理学会研究報告，Vol.2009-UBI-22，No.15，2009