

特定空間における人の行動予測モデルに基づく言語化への取り組み

An Approach on to Verbalization based on a Prediction Model of Human Behaviors in a Particular Space

落合 恵理香*¹ 小林 一郎*²
Erika Ochiai Ichiro Kobayashi

*1*²お茶の水女子大学大学院人間文化創成科学研究科理学専攻
Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University

Recently, we can easily obtain quite a few movies, therefore, we need a method to effectively manage those movies. At this moment, it is difficult for us to find a particular scene in a movie. if we want to do so, we have to watch through all the parts of a movie. Considering this problem, in this study, as the first step to retrieve a particular scene of a movie, we try to retrieve a scene where a human interacts with objects in a room. To achieve this, we firstly apply image processing techniques to analyze a movie and identify human motions in the movie, and then verbalize human behaviors in a room by observing how the human interacts with objects with a model of Bayesian network for identifying human actions.

1. はじめに

近年、人が動画を撮影する機会が増加しており、多量の動画の管理が必要とされている。一方で、撮影された多量の動画の中から人の行動の一部分を探し出すには、人が動画の内容を確認しなければならないため、人的負担が大きい。今日までに、人の頭部の位置や姿勢から人の行動を言語化する手法^[1]や一方向のカメラから得られる動画をを用いて言語化する手法^[2]が考えられている。これらの研究により、人の行動を自然言語による検索が可能となれば、人的負担が軽減される。

そこで、本研究では、人の行動を言葉でし、撮影された動画に対して画像処理を施すことにより得られた複数の特徴データを基に、特定空間内での人の動きと物との関わりを観察し、人の行動を言葉で説明する手法を提案する。

2. 言語化システムの構築

2.1 動画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^[3]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。

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)。

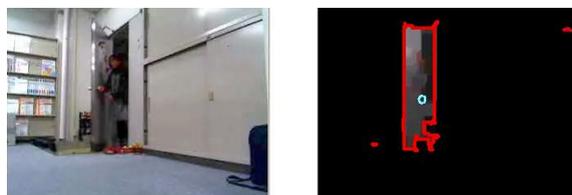


図 1: 入力画像 (左) と輪郭抽出画像 (右)



図 2: 原画像

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 5 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 5 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、本研究では、白の領域である定義物体内のことを、指定した物体の「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 4 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^[4]を基に構成する。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

連絡先: 落合恵理香, お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース小林研究室, 〒112-8610 東京都文京区大塚 2-1-1, 03-5978-5708, ochiai.erika@is.ocha.ac.jp



図 3: 画像上における物体指定



図 5: 膨張処理前 (左) と膨張処理後 (右) のマスク画像

```

*定義設定*
マウスによって4点座標を指定する。
まずDefine1から4点指定してください。
指定できたらキーを押してください。
*カメラ1*
  [x座標] [y座標]
[1] : 102   135
[2] : 100   214
[3] : 131   214
[4] : 131   136
定義物体名を入力してください>ドア
次にDefine2で物体を定義してください。指定後キーを押してください。
*カメラ2*
  [x座標] [y座標]
[1] : 121   205
[2] : 118   71
[3] : 148   69
[4] : 149   222
定義物体名を入力してください>ドア。
    
```

図 4: 物体定義画面

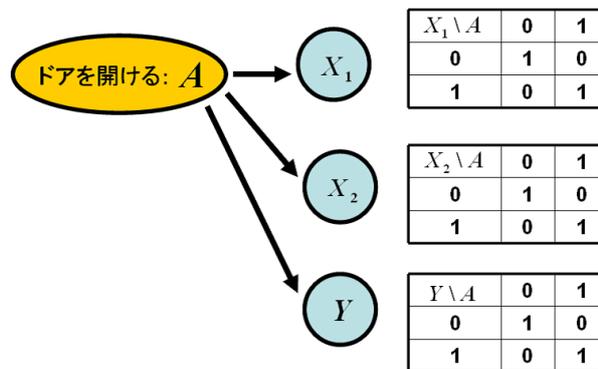


図 6: 行動判定のモデル(「ドアを開ける」)

層格における「動作主格」,「対象格」に限定して,「人が,何を,どうする」という形の文とする。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。

人や物の動作が表現されやすい述部の適切な表現を選ぶため,京都大学で開発された大規模格フレーム^[5]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。これにより,3 個の動詞 { 開ける, 閉める, ノック } を選定した。

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^[6]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 6 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。重心座標が指定した物体の定義域に一定時間入った場合に,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^[1]において課題とされていた一つの動画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,この事象を図 6 のノード A とする。各ノードの条件付確率分布表は予め付与するものとし,「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」という言語が出力される。

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。実験環境は以下の通りである。

OS : Windows XP

CPU : Intel Core Duo

メモリ : 4Gbyte

カメラ : Panasonic 社製 BB-HCM715, 二台

本研究では,言語化の時間的正確性の向上を目的として,二台のネットワークカメラ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ファイルを用いた。先行研究と本研究での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, 表 2 に示す。

3.1 考察

今回用いた動画は,ドアに対する人の振る舞いが映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。

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。

表 1: 先行研究での言語化結果
時刻: 0 出力言語無し



時刻: 91 「人がドアを開けた」



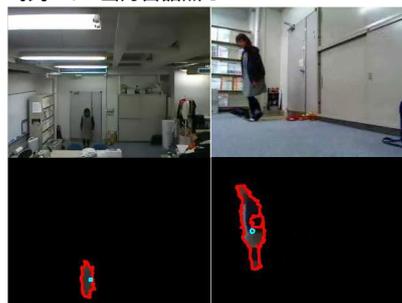
時刻: 148 出力言語無し



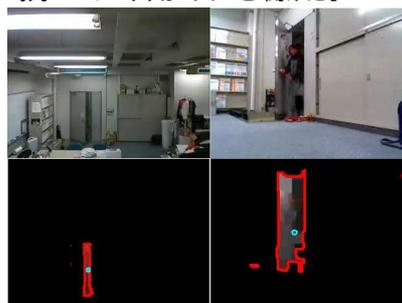
表 2: 本研究での言語化結果
時刻: 0 出力言語無し



時刻: 91 出力言語無し



時刻: 148 「人がドアを開けた」



4. おわりに

本研究では、動画ファイルに対して、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予定である。

参考文献

- [1] 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
- [2] 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
- [3] OpenCV, <http://opencv.jp/>

- [4] Fillmore, CHARLES J., "The Case for Case", In Bach and Harms (Ed.), *Universals in Linguistic Theory*, pp.1-88, New York: Holt, Rinehart, and Winston, 1968.
- [5] 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
- [6] 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
- [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