特定空間における人と物のインタラクションの言語化

能見 麻未　(指導教員：小林 一郎)

1　はじめに

近年，デジタルカメラや Web カメラの普及により，動画像の使用が容易になってきた．しかし，撮影された多量の動画像の中から特定の一部分だけを探し出すことは困難であり，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．

このことから，本研究では取得された動画像に対して画像処理を施し，特定空間内での人の動きとその空間に存在する物体とのインタラクションを観察することにより，人の行為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

本研究では，動画像ファイルの初期画像を「元画像」と呼ぶ．元画像からその画像を示す空間内に存在する物体を定義する知識を作成し，人の振る舞いは，元画像と入力画像の背景差分を用いて捉え，定義された物体の知識を用いて画像理解結果を言語化するシステムを構築する．ここで，本システムにおける具体的な処理の流れについての説明を行う．

2.1　空間知識の作成

多様な背景（元画像）に柔軟に対応するため，元画像が表示されているウインドウ上で空間に存在する物体の四隅をクリックすること（図 1）で，対象となる物体の画像内における座標値を取得する．取得された物体の座標値を用いて，定義物体それぞれの二値化画像を作成する（図 2）．この二値化画像は，特定空間内に存在するどの定義物体に関連した動作を行っているのかを見つけ出すために使用する．また、周辺での動作を考慮するために、膨張処理を施す．白い領域である定義物体内を，指定した物体の「定義域」と呼ぶ．

図 1: マウス指定画像

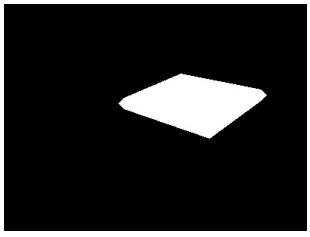


図 2: 二値化画像

2.2　動画像からの特徴データ抽出

画像認識には，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．提供される各種画像処理法の中の，背景差分，輪郭検出を用いて人の振る舞いを認識する．

本研究では、元画像と入力画像の差分をとり，状況の変化に柔軟に対応できる背景差分法を用いた．しかし，背景差分法だけでは，不要である光の変化や影まで差分として捉えてしまう．このため，各画素の RGB 値に注目し，元画像と入力画像，両方の (x, y) 座標について RGB 値の比較を行い，元画像と入力画像の R,G,B の値の差が全て閾値以下であれば，光や影とみなし，差分から排除する処理を施している．

背景差分法を用いて得られた画像から特徴データの抽出を行う．本研究では，差分で得られた画素を一つの集まりとして捉える事を可能とする輪郭検出を用いる．この捉えた領域の位置情報から重心を計算し，特徴データとして扱う．輪郭線で囲まれる領域にある点の重心を座標値の相加平均として計算し，動画像から得られるすべての画像に対してデータの抽出を行う．

このデータを用いて，画像理解，言語化を行う．

2.3　事前知識を利用した言語化

画像理解結果の言語化は，予めシステムに付与される，モノと人の行動のインタラクションに関する事前知識を用いることにより行う．この事前知識は，ロシアの心理学者 Vygotsky によって提案された Activity Theory[4] を用いる．

Activity Theory を参考にし, 空間内で行われる人の行動に関して，動作，目的，対象物体の分析を行った．本研究で用いる事前知識の一部を表にまとめ，図 3 に示す．

Scenario	Type	Artifact	Purpose
部屋に入る・出る ・ドアを開ける ・人が入る ・ドアを閉める	Action Operation	ドア	入退室するため

図 3: 事前知識 (一部)

この作成した事前知識に基づき，画像理解結果とそれを説明する適切な言葉を選択する知識をシステムに付与する．

2.4　言語化システム

画像処理を施し得た特徴データの重心座標が，連続した一定時間，特定空間内で定義した物体領域の中に存在する条件を満たした際に言語化を行う．この定義域と重心の関係は，物体定義時に領域を保存しておいた画像と照らし合わせることによってどの物体の定義域内に重心が存在するかを見つけ出す方法をとる．

ここで,“ ドア ”に関する言語では,「開ける」と「閉める」の両方が同時に言語化されてしまうという問題点が考えられる.「閉める」という動作には,「開ける」という前提条件が存在するため，そのような前提条件を考慮に入れて言語化を行う．

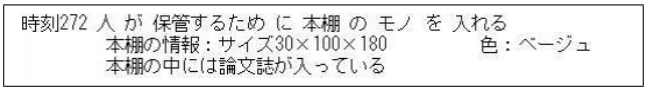


図 4: 言語化出力結果

出力される言語表現には，Activity Theory を基に，基本的には「誰が何のために何をどうする」という形で表示される．

また，インタラクションされる物体の詳細情報にも問い合わせを行い，物体の大きさなどの外部情報，入っているものなどの内部情報の表示を行う．

3　実験

システムを用いて，特定空間内での人の行動を画像理解結果から説明する実験を行った．今回，用いた空間における環境は，ドア，机，椅子，冷蔵庫などがある環境であり，カメラを上方に設置し，固定された一方向から録画された動画像を用いた．

本実験では，人がドアから入室してきて，机を拭くという作業を行った場合の動画像を使用し，空間内に定義した物体は，ドア，机，冷蔵庫の 3 点である．

3.1　実験結果

```
時刻74 人 が 入退室のため に ドア の 扉 を 開ける
        ドアの情報：サイズ100×200　　色：グレー
時刻98 人 が 入退室のため に ドア の 扉 を 閉める
時刻165 人 が 作業するため に 机 で 作業している
        机の情報：サイズ170×150×60　　色：ホワイト
時刻326 人 が 作業するため に 机 で 作業している
時刻476 人 が 作業するため に 机 で 作業している
時刻604 人 が 作業するため に 机 で 作業している
```

図 5: 言語化出力結果

「開ける」,「閉める」という条件付き表現が同時刻で出力されていないことから，条件が考慮されて出力されていることがわかる．

今回はドアと机にインタラクションする動画像を使用したため，両方に関係する言語化の結果を得ることができた．

4　考察

今回用いた動画像は，ドアと机の 2 つの物体にインタラクションする人の動作を捉えたものであり，複数の画像処理技術を駆使し，動画像内での人の振る舞いを認識し，それに対する言語化を行った．元画像内に定義された物体の領域内に特徴データが合致している場合については，言語化することができた．

さらに，リアルタイムで行った言語化に，グラフから得られる結果を付与してゆくことで，人の振る舞いを捉えたより詳細な言語化を行うことが可能になると考える．

5　まとめと今後の課題

本研究では，取得された動画像に対して，画像処理を施し，特定空間内での人の動きとその空間に存在する物体とのインタラクションを観察することにより，人の行為を言葉で説明する手法を提案した．具体的には，空間内の物体定義手法の提案，背景差分画像認識の精度向上を行った．また，言語化に向けて，Activity Theory による人の行動と物とのインタラクションの分析，それに基づく言語化のための知識作成，それらを用いた画像理解結果の言語化システムの構築を行った．

今後は，さまざまな動画像の言語化が行えるように提案手法の汎用性および頑健性を拡張し，言語化された言葉の正確性を向上させると共に，より詳細な言語表現の付与を行っていきたいと考えている．

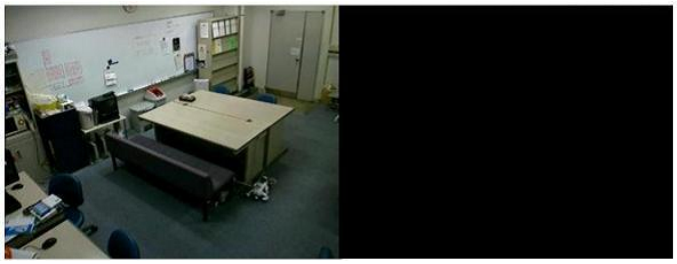
参考文献

[1] 檜山　敦子, 小林一郎：“ 生活空間内における人物行動の画像理解による言語での説明 ”，情報処理学会全国大会，4P-3，Mar.2007.

[2] Mirai Higuchi，Shigeki Aoki，Atsuhiro Kojima，Kunio Fukunaga：“ Scene Recognition Based on Relationship between Human Actions and Objects ”，Proceedings of the 17th International Conference on Pattern Recognition (ICPR ’04)，2004.

[3] “ OpenCV ”，http://opencv.jp/

[4] Vygotsky：“ Acting With Technology:Activity Theory And Interaction Design(Acting With Technology Series) ”, Oct.2006.



時刻 0

時刻 70　「人が入退室のためにドアを開ける」

時刻 88　「人が入退室のためにドアを閉める」

時刻 534 「人が作業するために机で作業している」