

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

檜山敦子[†] 小林一郎[‡]

[†]お茶の水女子大学 人間文化研究科 数理・情報科学専攻

[‡]お茶の水女子大学 理学部 情報科学科

1 はじめに

近年、セキュリティ対策として監視カメラを多用する機会が増えている。しかし、人がカメラを常時監視することには大きな負担を要する。そのため、画像処理技術を用いて、画像の中から監視対象物体を特定する研究が多く報告されている [2,4,5,6,7,8]。本研究では、特定された監視対象の動きを言語で説明する手法を提案する。具体的には、監視対象空間を生活空間とし、その空間が持つ機能と監視対象である人の動作とのインタラクションをとらえることにより、人物の行動を推定し、言語で説明を行う手法を提案する。

2 提案手法

提案する手法の処理のながれを以下に示す。

step1. 画像の特徴抽出

取得した動画より画像処理技術を用いて、対象物体の動きを表す特徴データを抽出する。

step2. 特徴データの分析

step1 で取得された特徴データの分析を行う。

step3. 言語に基づく画像理解

step1 において分析された特徴データとその空間が持つ機能や時間情報とのインタラクションをとらえることにより、人物行動を推定し、動画画像中の状況を言語で説明する。

システムの概要を図1に示す。

2.1 画像の特徴抽出

本研究では Intel 社が提供する画像処理ライブラリである OpenCV[1] を利用した。利用する画像処理技術として、背景差分処理、色情報の抽出、膨張収縮処理、輪郭線追跡等を行っている。

本研究では、室内における人間の行動を対象としている。そこで人を認識するために、肌色領域を検出しその代表点の座標を特徴データとしてとらえた。入力画像の背景差分をとり変化のある画素を抽出し、RGB から HSV に色空間を変換し、H と S を用いて肌色を抽出し、膨張収縮処理を施した後、輪郭線追跡で人間

Natural Language Explanation of Human Behavior at Life Space based on Image Understanding

Atsuko HIYAMA[†], Ichiro KOBAYASHI[‡]

[†]Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Ootsuka Bunkyo-ku Tokyo 112-8610

[‡]Dept. of Information Sciences, Faculty of Science, Ochanomizu University, 2-1-1 Ootsuka Bunkyo-ku Tokyo 112-8610 {atsuko, koba}@koba.is.ocha.ac.jp

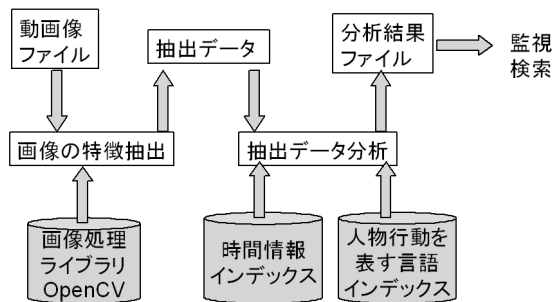


図 1: システムの概要

の顔や手の領域を求めた。抽出する特徴データは領域の重心の座標値とした。

2.2 特徴データの分析

本研究では、空間のモデルを予め設定しているため、その空間が持つ機能に応じた行動を人間はするものと考えられる。つまり、空間内の家具等の配置が決まっていれば、それを使う人間の行動はある程度予測でき、画像中での肌色領域の変化の仕方にも、一定の変化の仕方が見られる。よって、抽出された特徴データを画像中の領域別に振分けラベル付けし、誤認識や未認識の場合の許容回数を設定したうえで、それぞれの領域ごとのラベルの連続の仕方でも人間の行動を推論する。

2.3 言語に基づく画像理解

人間の行動変化から得られた特徴データを分析することで推論された人間の行動と、状況情報（時間的制約、空間的制約）をも加えて言語を選択する。

本研究では、研究対象を室内の人間の行動としたが、人間の行動は周囲の環境によってある程度推測可能と考えられる。例えば、人間が椅子に座るという行動を考えたとき、椅子があるから座るのであって、何も無い床で腰をかがめることはあまりなく、椅子の上に立つということも椅子が本来持つ機能とは異なる。同じ椅子でも食卓にあるのか、PCなどが置かれた作業テーブルがあるのか等でも、人間が何を行う目的で座るのが異なる。そのように、室内の空間が持っている機能は人間の行動を予測する上で重要な要因である。

人間の行動には日常的に行われる行動が多くあり、例えば、食卓の椅子に座るのでも、時間帯でその目的が朝食、昼食、夕食など様々である。よって、行動が観測された時刻や曜日などの時間情報も重要な要因である。

以上を踏まえ、本研究では人間行動の推定にあたり、

人間の行動の変化をとらえるだけでなく、その空間が持つ機能や時間情報にも着目し、言語で説明する手法を提案する。

3 実験

画像取得のためのカメラとして、LogicoolのQcam Orbit MP を利用し、1秒間に30フレームでとられた動画画像から1秒間に3フレームごとに処理を行った。画像解像度は320×240とした。

画像認識データと空間機能の制約に基づく言語による解釈の様子を図2に示す。図3は、動画画像中の一画面とそれに対応して生成された言語を示す。言語に対して、図4に示すように登録された時間情報も加味することで、より詳細な状況を言語で説明することも可能となる。

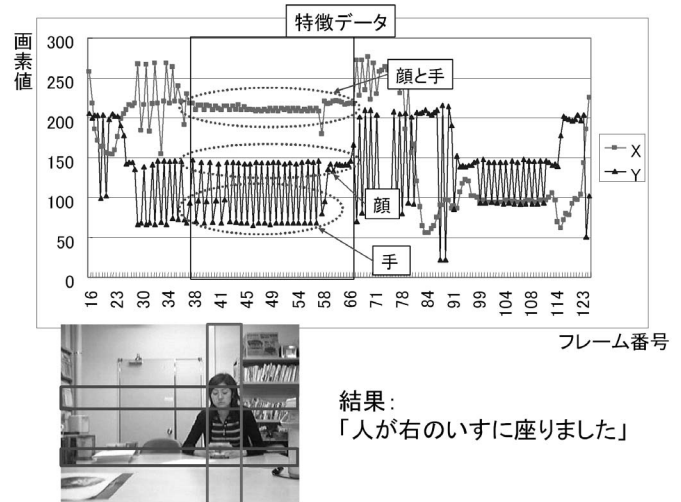
4 まとめ

本研究では、まず生活空間内における人間の行動を、取得した動画画像から画像処理技術を用いて肌色領域を検出し、特徴データを抽出した。その後、抽出された特徴データを画像中の領域別に振り分けラベル付けし、ラベルの連続の仕方から人間の行動を推論し、時間情報も加えて言語で表現させた。

特に本研究では、人間の動きのみに着目しているのではなく、生活空間がもつ機能に着目することで、精度の高い画像処理を必要とすることなく画像処理による認識を強化させ、また状況情報にも着目することで、より詳細な言語での説明が可能となった。これらに着目することで画像理解を補強できると考えている。

参考文献

- [1] <http://www.intel.com/technology/computing/opencv/>
- [2] 動画画像における人物行動の自然言語による自然言語による説明の生成, 小島篤博 他, 電気情報通信学会論文誌, D, vol.J81-D- , No.8, pp.1867-1875, Aug. 1998.
- [3] 本村陽一, 西田佳史, 日常環境における支援技術のための行動理解, 人工知能学会誌, No.20, No.5, pp587-592, 2005.
- [4] 小野口一則, 和田将司, 視差情報を加えた Temporal Templates によるジェスチャー認識, 信学技報 NLC2006-14, 2006.
- [5] 星野考総 他, ペット型ロボットをもちいた独居者の在宅データ取得と隠れセミマルコフモデル (HMM) による体調不良推定に関する研究, 第22回ファジィシステムシンポジウム 7A3-4, 2006.
- [6] M.Higuchi, S.Aoki, A.Kojima, and K.Fukunaga: "Scene Recognition Based on Relationship between Human Actions and Objects", In *Proc. of the 17th International Conference on Pattern Recognition*, Vol.3, pp.73-78, Aug. 2004.
- [7] T.Kitahashi, A.Kojima, M.Higuchi, and K.Fukunaga: "A Cooperative Recognition of Human Movements and Related Objects Guided by Functional Model of Objects", In *Proc. of the 2005 International Conference on Cyberworlds*, pp.187-193, Nov. 2005.
- [8] M.Aaitou, A.Kojima, T.Kitahashi, and K.Fukunaga: "A Dynamic Recognition of Human Actions and Related Objects", In *Proc. of the First International Conference on Innovative Computing, Information and Control*, Vol.1, pp.306-309, Aug. 2006.



結果:
「人が右のいすに座りました」

図2: 画像認識の基づく言語での説明

	人が右のいすに座りました
	人が立ちました
	人が左のいすに座りました

図3: 画像中の一画面と言語表現

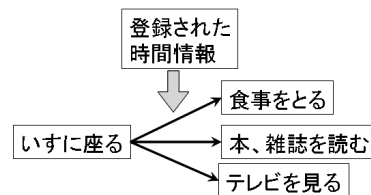


図4: その状況を示す適切な言語の選択