

人込みの中での人物と物体のインタラクション開始・終了検出

三橋 優人^{†1} 阿部 亨^{†2} 菅沼 拓夫^{†2}

東北大学大学院情報科学研究科^{†1} 東北大学サイバーサイエンスセンター^{†2}

1. はじめに

映像から人物や物体を検出する手法が種々提案されているが、多数の人物が存在する環境で人物と物体の様々なインタラクションを検出することは依然として困難な課題である。これに対し本研究は、手荷物の置き引きや不審物の置き去りの監視、店頭での購買活動の状況把握等への応用を想定し、人込みの中での人物と物体のインタラクションの開始・終了を映像から検出する手法の実現を目指している。この手法では、映像から抽出した個々の人物領域内で人物の手に対応する箇所を特定することにより、物体を手にとる・手から離す等、人物と物体のインタラクションの開始・終了の安定した検出を図っており、このアプローチは、人物と物体のインタラクションにおいて人物の手が主要な役割を果たすという知見に基づいている。本稿では、本提案手法の概要と基本設計について述べる。

2. 関連研究

人物と物体のインタラクションを検出する従来手法の多くは、人物と物体を独立に検出し、その後で両者を相互に関連付けインタラクションの判定を行っている。このため、例えば[1]のように、検出可能な対象物体や向きが制限されている場合は、検出できるインタラクション自体も制限されることになる。また、手荷物の置き引きや不審物の置き去りなど特定の状況を監視映像から検出することを目的とした[2]の手法のような場合でも、人込みなど多数の人物が検出される環境では、検出された個々の人物と物体との関係を特定することが困難となり、人物と物体のインタラクションを正確に検出することが難しくなる。

3. 提案手法

本稿で提案する手法では、時系列画像から

個々の人物の領域を抽出した後、各抽出領域から人物の腕に対応する箇所を特定し、その周辺で物体の存在を判定することで人物と物体のインタラクション検出を行う。このように、提案手法では、人物検出と物体検出を関連付けて実行することにより、人込みの中での人物と物体の多様なインタラクションの正確な検出を図っている。提案手法の処理の流れを図1に示し、以下に各処理の詳細を述べる。

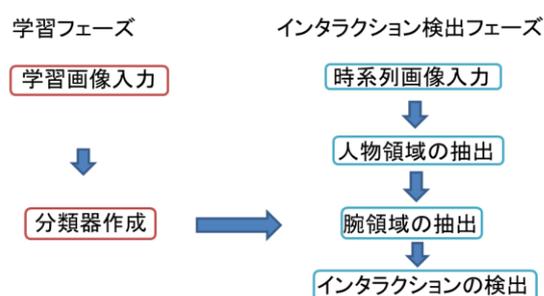


図1：提案手法の処理の流れ

3.1 人物領域の抽出

まず、入力された時系列画像から、HOG 特徴量[3]を用いて個々の人物領域を矩形として抽出する。HOG 特徴量は輝度勾配方向を複数のヒストグラムで表現したものであり、提案手法では、事前に人の形状を学習した検出器を用いて抽出を行う。人物領域の抽出例を図2に示す。

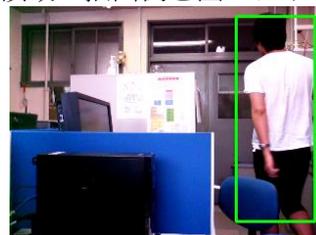


図2：人物領域の抽出例（矩形部分）

3.2 腕領域の抽出

次に、抽出した人物領域を小領域に分割し、人物の腕に対応する小領域を特定する。

人物と物体のインタラクションの開始・終了前後では、見かけの速度ベクトルであるオプティカルフローが人物領域の部位毎に異なると考えられる。そこで、人物領域を小領域に分割するために、提案手法では、時系列画像からオプティカルフローを求め、その結果に対し k 平均

Detecting human-object interactions in crowded environments

Yuto Mitsuhashi^{†1}, Toru Abe^{†2}, Takuo Suganuma^{†2}

^{†1} Graduate School of Information Sciences, Tohoku University

^{†2} Cyberscience Center, Tohoku University

法[4]によるクラスタリングを行う。

人物の腕に対応する箇所を特定する際、腕の向き等の影響を抑制するために、提案手法では、対象の回転や照明の変化に頑健である SURF 特徴量[5]を用いた Bag of Features (BOF) [6]として小領域の特徴を表現する。具体的には、オプティカルフローに基づき分割した小領域毎に、SURF 特徴量をヒストグラム化した BOF を作成する。この BOF を分類器に入力することで、腕に対応する小領域を決定する。分類器は、事前に用意した正解データ（腕の画像）と不正解データ（腕以外の部位の画像）から求めた BOF を用い、Support Vector Machine (SVM) により学習することで作成する。

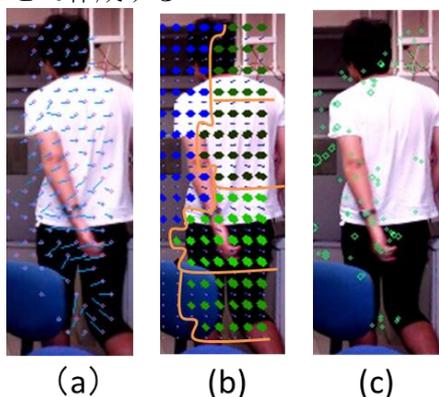


図 3 : (a) オプティカルフロー, (b) 分割された小領域, (c) SURF 特徴量

3.3 インタラクションの検出

人物の腕として特定した各小領域の周辺に対し、オプティカルフロー等を用いた領域分割、背景差分による領域検出を適用することで物体の存在の有無を判定し、人物と物体のインタラクションの検出を行う。

4. 予備実験

提案手法を検証するために、SURF 特徴量を用いた BOF により人物の腕の領域が特定できるか、人物の部位ごとの画像を対象とした予備実験を行った。

4.1 実験環境

実験には、USB カメラ (UCAM-DLW500TABK) で撮影した 320×240 画素の画像 1218 枚を用いた。内訳は、正解データ（腕の部位）721 枚、不正解データ（腕以外の部位）497 枚である。実験に使用した画像の例を図 4 に示す。この画像をランダムに 10 分割し、交差検定法（1 つをテストデータに、残りの 9 つのデータを分類器の学習に用いる操作を 10 回反復）により腕領域の分類精

度を評価した。

SURF 特徴量を用いた BOF を求める際は、BOF の次元数を 15 とし、最大値を 1 に正規化した。なお、実験にあたって OpenCV 2.3[7]、LIBSVM 3.17[8]を使用した。



図 4 : 実験に使用した画像の例

4.2 実験結果

交差検定法による分類精度は 79.06 % (963/1218) となった。この結果から、SURF 特徴量を用いた BOF により人物の腕の領域を特定することは十分可能であると考えられる。

5. おわりに

本稿では、人物領域の腕の箇所に着目することで、人物と物体のインタラクションを人込みの映像から検出する手法を提案した。今後は、人物領域の分割方法や小領域を記述する特徴量についての具体的な検討を行い、提案手法の実装、評価実験を進める予定である。

参考文献

- [1] A. Prest, et al., "Explicit modeling of human-object interactions in realistic videos," IEEE Trans. Pattern Anal. Mach. Intell., Vol.35, No.4, pp.835-848 (2013).
- [2] Y.L. Tian, et al., "Robust detection of abandoned and removed objects in complex surveillance videos," IEEE Trans. Syst., Man, Cybern. C, Vol.41, No.5, pp.1094-6977 (2011).
- [3] N. Dalal, et al., "Histograms of Oriented Gradients for Human Detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol.1, pp.886 - 893 (2005).
- [4] P.S. Bradley, et al., "Refining Initial Points for K-Means Clustering," 15th International Conference on Machine Learning, pp.91-99 (1998).
- [5] H. Bay, et al., "SURF:Speeded Up Robust Features," ECCV 2006, Lecture Notes in Computer Science, Vol.3951, pp.404-417 (2006).
- [6] F. Li, et al., "A Bayesian Hierarchical Model for Learning Natural Scene Categories," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol.2, pp.524-531 (2005).
- [7] <http://sourceforge.net/projects/opencvlibrary/>.
- [8] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.