

Random Forestと畳み込み特徴に基づく インタラクティブ機械学習による人物識別 —移動ロボットの教室内巡回中の人物再同定への適用—

Person Identification by Interactive Machine Learning
based on Random Forest and Convolutional Feature
-Application to Person Re-identification by Mobile Robots Circulating around the Classroom-

村田 祐樹
Yuuki Murata

渥美 雅保
Masayasu Atsumi

創価大学理工学部情報システム工学科
Dept. of Information Systems Sci., Faculty of Sci., and Eng., Soka University

We propose a person identification method of mobile robots which support users through circulating indoor space such as classrooms. This method extracts and accumulates convolutional features from detected person's rectangles, interactively labels these features and updates person classifiers of the Random Forest. We confirm usefulness of this method from experiments of identifying persons detected by mobile robots circulating around the classroom.

1. はじめに

人物識別は犯罪の防止や捜査など主にセキュリティの分野を中心に重要な役割を担ってきた。特に近年では、屋外空間で人物の外見的特徴から個人を特定する研究が進んでいるが、一方でその技術が教育支援の目的で屋内空間に応用されている事例はまだ多くない。本研究では教室等の屋内空間を巡回する移動ロボットが出会った人物を識別・再同定するためのインタラクティブな機械学習手法を提案する。本手法では人物識別のために CNN(Convolutional Neural Network)を特徴抽出器として利用する。収集・蓄積した特徴量は GUI を用いて人間が介入する形でラベル付け処理をし、Random Forest による学習を繰り返し行い、識別器を生成する。

我々が開発を進めている SOTARO(SOciable Teaching Assistance RObot) [渥美 17a, 渥美 17b]と名付けた大学の授業支援ロボットに組み込むことにより、人物識別・再同定の実験を通じて、本手法の有用性を確かめる。

2. SOTARO 授業支援システム

本システムでは PC 教室内で行われるソフトウェア演習授業において、TA(ティーチングアシスタント)と TA 移動ロボット、TA チャットボットが連携して受講学生の学習支援を行う。本システムは図 1 に示すように、2 台のロボットとそれぞれのリモートプレインサーバ、チャットサーバとボット、及びウェブブラウザ上のチャットクライアントから構成される。移動ロボットは TurtleBot をベースに構築され、RGB-D センサとして Kinect を搭載している。Kinect は椅子に座っている受講学生に合わせて床面からおおよそ 1.1m の高さに設置している。リモートプレインサーバでは取得される RGB-D 映像と音声情報の処理、インタラクションのための発話テキストの送信、および GUI を用いたロボットの制御を行う。本論で述べる人物識別システムでは RGB 画像データを用いた認識処理を行う。

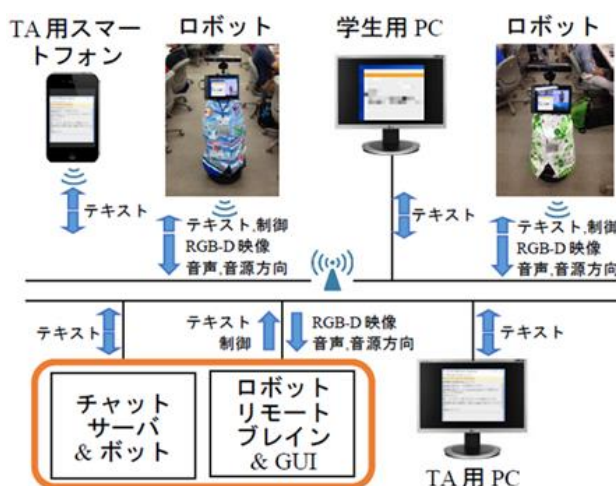


図 1 SOTARO 授業支援システム概要

3. 人物識別システム

3.1 概要

図 2 に提案する人物識別システムの構成を示す。まず、移動ロボットに搭載された Kinect の RGB カメラから画像を取得し、HOG 特徴と SVM による検出器 [Dalal 05] で人物の上半身矩形を取り出す。続いて CNN に人物矩形をリサイズした画像を入力し、全結合層の畳み込み特徴を抽出する。取得した特徴量は最近傍法のアルゴリズムを用いてクラスタリングし、GUI を使用した人間参加型のラベル付け処理を行う。特徴量とラベルは学習データとして蓄積していき、Random Forest [Breiman 01] を繰り返し学習することで、順次作成される人物識別器の性能を向上させていく。

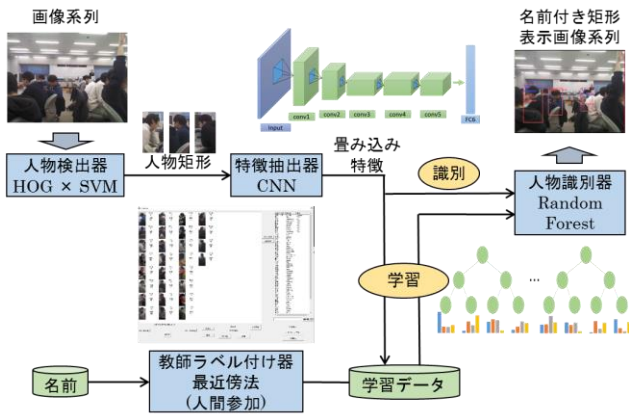


図 2 識別システム全体概要

3.2 人物検出

画像から人物領域を適切に取り出すため、事前に HOG 特徴と SVM による人物画像特徴の学習を行い、検出器を作成する。演習授業内では検出対象となる人物は椅子に座っていることが多いため、人物の頭頂部から腰にかけての上半身矩形を検出する。また Sliding Window 法によって、入力された画像全体から様々な位置・大きさの矩形領域を探索する。

3.3 特徴抽出

従来は物体の識別のために特徴量を設計する必要があったが、近年 CNN が注目されるようになり、大量の画像データセットで学習させた CNN を特徴抽出器として利用する手法が提案されている [Donahue 14]。本手法では検出器で取得した人物矩形をリサイズした画像を ImageNet で学習された CNN に入力し、全結合層から特徴を取り出す。

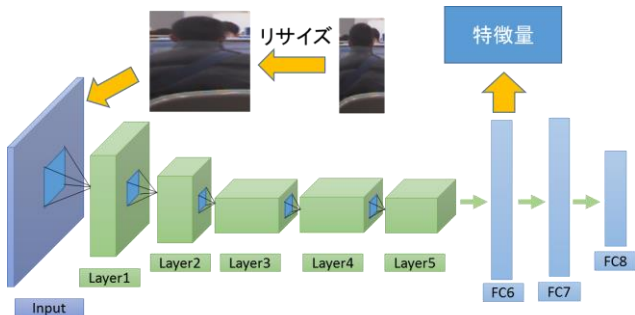


図 3 学習済み CNN を利用した画像特徴の抽出

3.4 人間参加型のラベル付け

最近傍法のアルゴリズムを用いて、取得した特徴のラベル付けを実行する。最近傍法とは、多数のデータで構成される集合に新たにデータが 1 つ与えられたとき、集合から最も近いデータを探してそのクラスターを割り当てる方法である。提案手法では閾値を設定し、与えられたデータと最近傍のデータの距離が一定以上となった場合には新しいクラスターを生成することとした。これにより、比較的見え方が近い同一人物の矩形が自動的に同じクラスターとなる。同一のクラスターに 20 枚以上の人物矩形が登録されると学習データとして登録され、図 4 に示す GUI 上に代表の画像が 1 枚表示される。この GUI はクラスター統合・調節部と表示

ラベル選択部から構成されている。画面左のクラスター統合・調節部では、自動的に登録された代表画像をもとに TA が手でクラスターの統合・調節を行うことができる。例えば同一人物が異なるクラスターに割り当てられている場合には、クラスターの統合処理を行う。また、学習データの対象となった各クラスターには自動的に数字のラベルが登録されるが、名簿等から受講学生の名前が事前に入手できる場合には、画面右の表示ラベル選択部で数字の代わりに学生の名前をラベルとして登録することも可能である。ロボットがキャプチャした映像のモニタ画面には検出した人物を囲った赤い矩形とともにここで登録したラベルが表示される。



図 4 人間参加型ラベル付け GUI

3.5 人物識別器の作成

ラベル付けされた特徴データをもとに Random Forest の学習を行う。Random Forest はアンサンブル学習の一種で、ブートストラップサンプリングによって作成された相関の低い多数の決定木を弱識別器とし、それらの合議によって識別を行う手法である。人物識別器は 4.3 で後述するように授業内で収集する画像データを用いて一定時間間隔の繰り返し学習によって生成される。

4. 実験

4.1 実験概要

ソフトウェア演習の授業支援運用で収集した画像データを用いて、まず、画像から検出した人物矩形のデータセットを作成して識別性能評価を行う。続いて実際の演習授業を想定し、順次収集される画像データを用いて識別器作成を繰り返し行った際の各段階における識別性能の評価を行う。

4.2 データセットに対する識別評価

作成したデータセットは演習授業内で取得した人物画像のうち 30 人の上半身矩形 6000 枚と誤検出を分類するための人物以外の画像 200 枚の合計 6200 枚で構成されている。このうち学習用サンプルに 4960 枚、テスト用サンプルに 1240 枚を使用して特徴抽出・Random Forest の学習を行い、5 分割交差検定で評価した。本実験で使用した CNN は、ILSVRC 2012 (ImageNet Large Scale Visual Recognition Challenge 2012) で好成績を収め

た AlexNet をはじめ、それを改良した CaffeNet, VGG(Visual Geometry Group)によるモデル 4 種類(CNN-M [Chatfield 14], CNN-S[Chatfield 14], Net-D[Simonyan 15], Net-E[Simonyan 15])の合計 6 種類で、それぞれ全結合層 FC6 から特徴を取得している。また、Random Forest のパラメータを表 1 に示す。人数分のクラスと誤検出クラスを合わせた 31 クラスの識別を行った結果は表 2 に示す通りとなり、どのモデルも比較的高い精度で識別することができた。

表 1 人物識別実験に使用したパラメータ

決定木の数	100
決定木の最大深さ	15
特徴選択数	100
閾値選択数	5
サブサンプリング率	0.5

表 2 学習済み CNN 別の識別性能評価

CNN モデル	Accuracy
CNN-M	0.85
CaffeNet	0.84
CNN-S	0.84
AlexNet	0.83
Net-D	0.79
Net-E	0.78

4.3 授業運用を想定した識別評価

図 5 に本実験における評価方法を示す。演習授業内でキャプチャした約 80 分間の映像を使用し、最初の約 4 分間の映像からは人物特徴の抽出とラベル付けのみを行い初回の学習を行う。最初の識別器生成後は識別と同時に特徴データ収集とラベル付け、及び追加されたデータを用いた学習を行い、新たな識別器が完成するまでの間を 1 段階として区切った評価を繰り返し行うことで識別性能の推移をみる。なお、人物の検出には 3.2 で述べた手法で人物上半身矩形画像 4690 枚と背景画像 8098 枚の合計 12788 枚から学習した人物検出器を使用している。また、識別のために使用する CNN には 4.2 の結果と特徴抽出にかかる時間を考慮して CaffeNet を選択し、Random Forest のパラメータは表 1 と同様のものを使用した。本実験では 7 つの識別器を生成することができた。各段階における登録クラス数と人物識別性能の推移を表 3 に示す。序盤は学習データが少ないことと、学習データに含まれない人物が多数存在することから識別性能が良くないが、時間経過によって多数の人物の様々な見え方のデータが集まることで、次第に性能が向上していくことを確認できた。本手法により人物再同定されている画面を図 6 に示す。

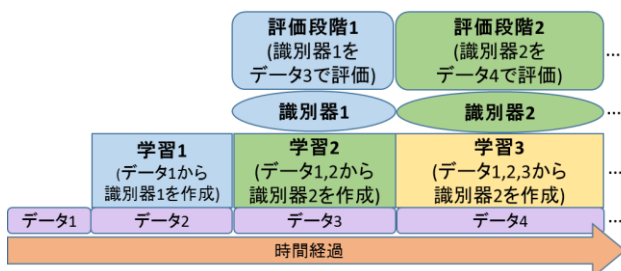


図 5 識別性能の評価方法

表 3 識別性能とクラス数の推移

時区間	1	2	3	4	5	6	7
クラス数	12	24	28	33	37	38	40
正解率	0.23	0.12	0.15	0.42	0.54	0.44	0.63



図 6 人物再同定の実行画面

5. むすび

本論では、屋内空間を移動するロボットのためのインタラクティブな機械学習による人物識別手法を提案した。そして、大学の演習授業を支援するロボットが運用で収集したデータを用いた実験によりその有効性を評価した。実験の結果、演習授業時間中に識別器を更新していくことで識別精度の向上を図ることが確認できた。今後の課題として、本格的な運用に向けてより早い段階で高性能の識別器を作成できるようにすること、識別結果に受講学生の学習状態、課題進捗状況等のデータを組み合わせることで授業支援の幅を広げていくことがあげられる。

参考文献

- [渥美 17a] 渥美雅保, 村田祐樹, 安川葵: SOTARO: オープンチャットとロボットの連係による演習授業支援システムー Human-in-the-loop 型機械学習によるアプローチ, 情報処理学会第 79 回全国大会, 6E-07 (2017)
- [渥美 17b] 渥美雅保, 村田祐樹, 安川葵: オープンチャットとロボットの連係によるティーチングアシスタントとの協働システム, 2017 年度人工知能学会全国大会, (第 31 回) 論文集 4G1-OS-14a-5 (2017)
- [Dalal 05] Dalal. N., and Triggs. B.: Histograms of oriented gradients for human detection, CVPR (2005)
- [Breiman 01] Breiman L.: Random forests, Machine learning 45(1):5-32 (2001)
- [Donahue 14] Donahue. J, et al.: DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, ICML (2014)
- [Chatfield 14] Chatfield. K., et al.: Return of the devil in the details: Delving deep into convolutional nets, BMVC (2014)
- [Simonyan 15] Simonyan, K., and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, ICLR (2015)