

移動ロボットによる人物識別のためのインタラクティブな教師付き機械学習とその演習授業への適用

村田祐樹† 渥美雅保†

創価大学理工学部情報システム工学科†

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で重要な役割を担ってきた。特に近年では、屋外空間において人物の外見的特徴を用いて個人を特定する研究が進んでいるが、一方でこうした技術が教育の分野に応用されている事例はまだ多くない。我々は、大学の演習授業を支援するシステムとして SOTARO(SOciable Teaching Assistance RObot)と名付けたオープンチャットとロボットを連係させたシステムを開発している[1]。本システムでは、その目的の1つとして、ロボットによる人物識別に学生の課題進捗状況等のデータを組み合わせることで個々の状況に適した授業支援を目指している。

本論では、この人物識別に関して、その特徴であるリモートブレインでの人間参加型機械学習と演習授業支援の運用で収集したデータを用いた実験による有用性の評価について述べる。

2. SOTARO 授業支援システムの全体概要

図1に本システムの構成を示す。本システムは2台のロボットとそれぞれのリモートブレインサーバ、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。ロボットは TurtleBot をベースに構築され、RGB-D センサとして高さおよそ 1.1m の地点に Kinect を搭載している。リモートブレインサーバでは Kinect で取得される RGB-D 映像と音声情報の処理を行うが、本論で述べる人物識別器は、そのうち RGB 画像データを用いた認識処理を行う。なお、ロボットの制御移動は現状 GUI による遠隔操作で行っている。

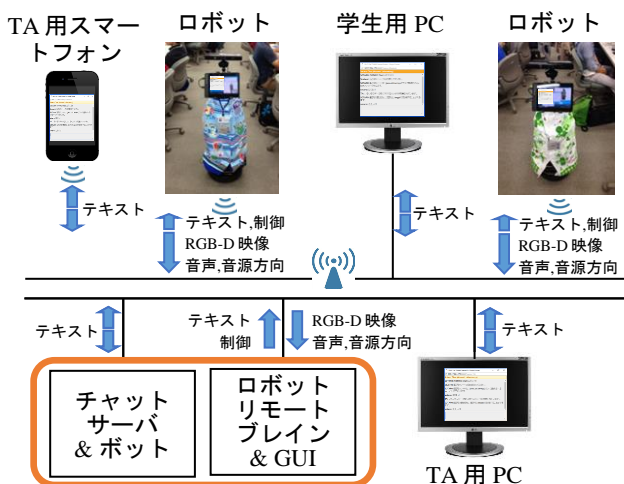


図1. SOTARO 授業支援システムの構成

Interactive Supervised Machine Learning for Person Identification by Mobile Robot and its Application in Classroom
 †Yuuki Murata, Masayasu Atsumi
 Department of Information Systems Science, Faculty of Science and Engineering, Soka University

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、SOTARO に搭載された Kinect の RGB カメラから画像を取得し、Navneet らの手法[2]を用いて作成した HOG 特徴と SVM による検出器で人物の上半身矩形を取り出す。続いて CNN(Convolution Neural Network)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[3]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。

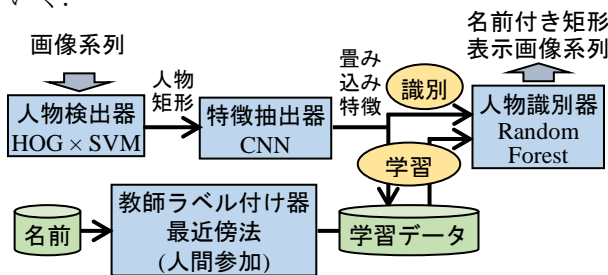


図2. 人物識別システムの構成

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

Jeff らの手法[4]に基づき、物体分類問題用に作成された CNN(Convolution Neural Network)のモデルの全結合層から畳み込み特徴を抽出する。今回は、CNN のモデルに ILSVRC 2012(ImageNet Large Scale Visual Recognition Challenge 2012)の分類タスクで優勝した AlexNet を改良した CaffeNet を使用する。検出器によって取得した人物矩形を 224×224pixel にリサイズした画像を入力し、全結合層 fc6 の 4096 次元を特徴として取り出す。

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、取得した畳み込み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を生成することとした。これにより、比較の見え方が近い同一人物の矩形に自動的に同じラベルが割り当てられる。ラベル付けされたデータはGUIを介して手動で統合・削除が可能となっており、例えば同一人物に異なるラベルが付いた場合には、ラベルの統合処理を行う。

3.5. 人物識別器の学習

ラベル付けされた特徴データをもとに Random Forest で学習を行う。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 5 分割交差検証を行い、評価する。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、アキュラシーが 0.84 となり、比較的高い精度で識別することができた。今回使用した Random Forest のパラメータを表 1 に示す。

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

4.3. 授業運用を想定した識別評価

図 3 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 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 2 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 4 に示す。

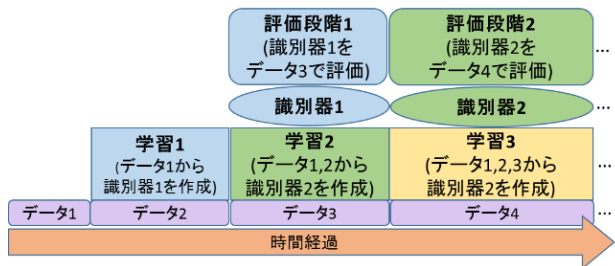


図 3 識別性能の評価方法



図 4 人物再同定の実行画面

表 2 識別性能とクラス数の推移

段階	1	2	3	4	5	6	7
登録クラス数	12	24	28	33	37	38	40
識別正解率	0.23	0.12	0.15	0.42	0.54	0.44	0.63

5. むすび

本論では演習授業支援システムの一構成要素となる人間参加型の機械学習による人物識別手法に関して述べた。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れることが確認できた。今後の課題として、より早い段階で高性能の識別器を作成することがあげられる。

参考文献

[1] 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携による演習授業支援システム—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, 2017.
 [2] Navneet, D. and Triggs, B. "Histograms of oriented gradients for human detection." CVPR'05. Vol. 1. IEEE, 2005.
 [3] Leo, B. "Random forests." Machine learning 45.1, pp.5-32, 2001.
 [4] Jeff, D., et al. "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition." ICML. 2014.