

屋内空間を巡回するロボットのための協調型転移深層学習に基づく 人物再同定とその TA 業務支援への適用

Person Re-identification based on Interactive Transfer Deep Learning for Mobile Robots and Its Application to TA Task Support

村田 祐樹^{*1}
Yuki Murata

渥美 雅保^{*1}
Masayasu Atsumi

^{*1} 創価大学大学院工学研究科情報システム工学専攻
Information Systems Science, Graduate School of Engineering, Soka University

This paper proposes interactive person re-identification method for mobile robots that periodically provide services to specific groups. This method consists of a CNN-based person feature extractor that is trained based on Triplet Loss, and a CNN-based person re-identifier that is trained based on transfer learning. Person re-ID is executed through a cooperative human-in-the-loop learning approach. As an example of a service, we apply this method to a Teaching Assistant (TA) support. This application aims to support students' study based on their identification by the proposed method and student card reading in which appearances and names are linked. Performance of the proposed method is evaluated by experiments using a large open dataset and a self-made dataset periodically collected for the same group by a mobile robot. In addition, the feasibility of the TA support is verified by experiments in which robots are operated in actual classes.

1. はじめに

人物識別は、犯罪の捜査やその防止、マーケティング情報の収集など、社会の中で重要な役割を果たしている。その中でも人物再同定は、視野を共有しない複数の固定カメラに映る人物画像から、同一人物を識別する問題と定義される。本研究では、特定の集団に定期的にサービスを提供するロボット向けの人物再同定を提案する。公開されている大規模データセットと、移動ロボットで収集した自作のデータセットを使用した実験により、本手法の有用性を評価する。

また、ロボットが各個人に合わせた最適なサービスを提供するためには、人物再同定に加えて名前等の個人識別子を用いた情報検索を組み合わせることが重要である。本研究では、提供するサービスの一例として演習授業の TA (Teaching Assistant) 支援を取り上げ、人間とロボットが互いの特性を活かして受講学生の学習支援を提供するシステムを構築する。ロボットには名前と外見を結びつけるための学生証読み取り機能を搭載し、名前から得られる情報に基づいて各個人に合わせた学習支援の提供を目指す。実際の演習授業でロボットを動作させ、TA 支援の可能性を検証する。

2. 関連研究

2.1 人物再同定

近年の深層学習に基づく人物再同定には、画像のペア [Radenovic 16] に基づく siamese ネットワークや 3 つ組 [Schroff 15] に基づく Triplet ネットワークモデルが用いられることが多い。これらは複数の画像を入力として受け取り、損失関数に基づいて各人物固有の特徴を出力するように学習される。テスト時には出力された特徴同士を、コサイン類似度などの距離関数を使用して人物の識別に役立てる。これらモデルは、入力サンプルの

組が同一人物によるものか否かを考慮するだけで学習ができるという利点がある。一方、全てのサンプルに事前のラベル付けが必要な点は問題である。また、類似度比較に基づいて人物を識別するには、データベース中の全てのサンプルと新たな入力サンプルを比較する必要があるため、データベースのサイズによっては計算量が膨大になる恐れがある。

これらに対処するため本研究では、人物固有の特徴を出力する特徴選択 CNN に、人物識別を実時間で行うための人物再同定 CNN を組み合わせたモデルを提案する。そして、人物データを収集しながら転移学習により短時間で繰り返し学習を実行する。

2.2 学習支援ロボット

コミュニケーションロボットを教育の分野に取り入れる試みに関心が高まっている。高石ら [高石 07] は、教師と学生のコミュニケーション促進を目的としたロボット型の IP ホンを TA として導入し、システムの試作を行った。彼らが開発したシステムでは、教師やその補助者 (人間の TA) が専用のアプリケーションを介して、聞く・考える・質問する等のロボットの操作を任意のタイミングで行った。

一方 Shiomi ら [Shiomi 15] は、小学校の科学の教室内に、TA としてではなく、“科学の知識が豊富な転校生”という立場で振る舞うロボットを導入し、実地調査を行った。ロボットの行動は半自律的な遠隔操作を介して、科学に関する質問に答えられるように設計され、子供からの質問を促した。彼らは一か月間に及ぶ実験から、特に意欲的に質問をする子供らの科学的な好奇心を高められると結論付けた。

本研究は、大学のプログラミング演習授業におけるロボット TA による学習支援サービスの提供を目的としている。遠隔操作によって教室を巡回するロボットを TA として取り入れ、学生証を読み取る機能を搭載することで名前から得られる様々な情報に基づいて学生とコミュニケーションを行う。実際にロボットを授業中に動作させる中で、得られる効果や今後の可能性を検証した。

連絡先: 村田 祐樹, 創価大学大学院工学研究科,
〒192-8577 東京都八王子市丹木町 1-236, Tel: 042-691-2211,
Email: y.murata0105@gmail.com

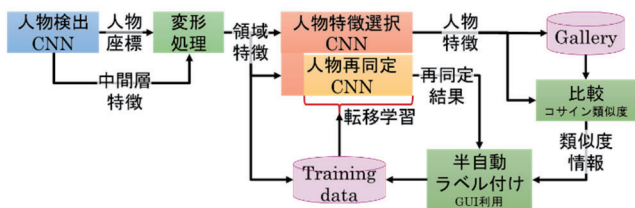


図1 人物再同定手法

3. 移動ロボット向け人物再同定

3.1 概要

提案する手法は、図1に示すように3種類の畳み込みニューラルネットワーク(CNN)—人物検出 CNN、人物特徴選択 CNN、人物再同定 CNN—と半自動ラベル付け処理で構成される。まず、人物検出 CNN に画像を入力し、人物が存在する領域の座標と隠れ層の特徴マップを出力する。これらの出力から人物領域に限定された特徴マップが人物ごとに抽出され、後半2つのCNNへ入力するための変形処理が施される。人物特徴選択 CNN は人物ごとに不変の特徴を選択して出力するため、人物再同定 CNN は人物の識別を行うために利用する。出力される人物特徴は Gallery として保存され、新たに取得されるサンプルとの比較に利用される。自動的なラベル付けは、比較結果として出力される類似度情報と再同定 CNN が出力する分類確信度に基づいて行われる。また、類似度と分類確信度が共に一定以下の場合は、GUI を用いた協調ラベル付け処理に切り替える。このように本システムでは、学習用データを随時作成し、人物再同定 CNN の転移学習を実施する。

3.2 転移深層学習に基づく人物再同定

高精度かつ実時間の人物再同定を実現するため、本研究では人物検出 CNN として YOLOv2[Redmon 17]を使用する。YOLOv2 は 23 の畳み込み層と 5 つのプーリング層で構成されるモデルで、人物を含む複数の物体をリアルタイムに検出することができる。ネットワークには高さと幅が 416 ピクセルの 3 チャネルカラー画像を入力し、物体の座標とクラス、それらの確信度を含む特徴を出力する。また、本手法では検出と同時に第 2 畳み込み層の特徴マップから人物領域の特徴を取り出す。

人物特徴選択 CNN(図2(a))は、人物検出 CNN で取得した人物領域の特徴マップを入力して人物固有の特徴を選択して出力するネットワークである。学習時には損失関数として、式(1)で表される Triplet Loss を使用する。

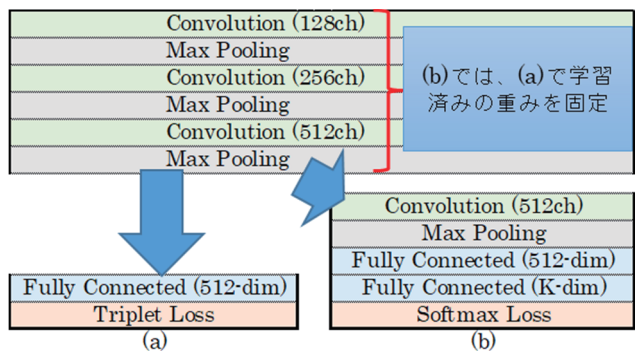


図2 特徴選択 CNN(a)と再同定 CNN(b)の構成

$$L(a, p, n) = \frac{1}{N} \left(\sum_{i=1}^N \max\{d(a_i, p_i) - d(a_i, n_i) + \text{margin}, 0\} \right) \quad (1)$$

where $d(x_i, y_i) = \|x_i - y_i\|_2^2$

ここで、 N はバッチサイズ、margin はマージンパラメータを表す。Triplet (a, p, n) は基準となるサンプル a (anchor)、それと同クラスのサンプル p (positive)、別クラスのサンプル n (negative) の 3 つの組で構成される。それぞれが重みを共有する同一の特徴選択 CNN へ入力され、人物固有の特徴が選択・出力される。これらの特徴同士を、距離関数を用いて比較したときに、同一クラス間では距離が小さく、別クラス間で距離が大きくなるように学習が行われる。

人物再同定 CNN(図2(b))は、人物特徴選択 CNN と同様に人物領域の特徴マップを受け取り、人物 ID を識別結果として出力する。本研究で扱う人物再同定は、 K 人の特徴データが存在すると仮定したとき、ネットワークの出力層が K 個のユニットを持つ、 K クラス分類問題として扱うことができる。出力層の各ユニット k ($k = 1, \dots, K$) の総入力を u_k とするとき、学習に用いるソフトマックス関数は以下の式(2)のように表すことができる。

$$y_k = \frac{\exp(u_k)}{\sum_{j=1}^K \exp(u_j)} \quad (2)$$

また、学習時の損失関数として式(3)の交差エントロピーを用いる。

$$E(\mathbf{w}) = - \sum_{n=1}^N \sum_{k=1}^K t_{nk} \log y_k(x_n; \mathbf{w}) \quad (3)$$

ここで \mathbf{w} は学習するパラメータ、 t は正解ラベル、 x はネットワークに入力する特徴を表す。

本研究で提案する人物再同定モデルは、入力に近い層については人物特徴選択 CNN と共通の形になっており、出力側の層は人物識別結果を出力するために、畳み込み層、プーリング層、全結合層を追加する形で構成されている。訓練には、あるタスクに関して学習させたモデルを別のタスクを解くために利用する手法である転移学習を用いる。提案モデルにおいては、前半部分は人物特徴選択 CNN を事前に学習させた時点の重みを利用し、後半に追加された層のみを学習する。これによって、人物のアピアランスの変化への適応とモデル学習の収束の促進を図ることができる。

3.3 GUIを用いた半自動ラベル付け

提案手法では、以下の 3 つのステップに沿って、自動的なラベル付けが可能か、GUI を用いた人の手によるラベル付けが必要かを判断する。

(1) 人物再同定器を用いた自動ラベル付け

既に学習された人物再同定器が存在し、新たに取得されたサンプルが、あらかじめ設定された閾値以上の確信度で分類された場合、分類されたクラスをそのままラベルとして使用する。

(2) 人物特徴の類似度比較に基づく自動ラベル付け

新しく取得されたサンプルの人物特徴を、既にラベルが付けられた全サンプルの人物特徴と比較し、類似度が高い順にソートする。最上位サンプルの類似度が事前に設定された閾値以上となった場合、そのラベルが新しいサンプルに割り当てられる。

(3) GUIを用いた手動ラベル付け

上記(1)、(2)で自動的にラベル付けがされなかった場合、そのラベルは人間のユーザによって GUI を使って決定する。使用

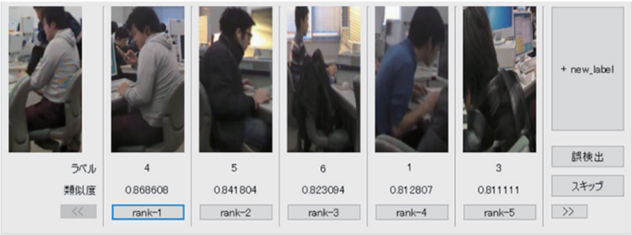


図 3 ラベル選択 GUI



図 4 ラベル調整 GUI

する GUI を図 3 に示す。入力サンプルは左端に表示され、類似度が高い順に 5 つのラベル付きのサンプルが右側に表示される。ユーザは、双方の画像を比較して最も類似する(同一人物である)と考えられる画像のラベルを選択することでラベル付けを行う。新しい画像に一致すると考えられる画像がない場合、ボタン操作によって新たなラベルを割り当てることができる。

図 4 のラベル調整 GUI には、クラス名、各ラベルに含まれるサンプル数やその代表となる画像が表示される。ユーザは既に与えられているクラス名を、人名などの任意の名前に手動で変更することができる。

4. 実験

4.1 使用するデータセット

実験に使用するデータセットの構成を表 1 に示す。人物再同定向けデータセットの多くは、人物領域が既に切り取られた状態の画像で構成されている。一方、本研究の提案手法は人物検出 CNN から人物領域の特徴を抽出するため、人物領域の座標情報を含んだ画像全体のデータが必要になる。Person Re-identification in the Wild (PRW) [Zheng 17] データセットは、公開されている大規模データセットの中でこの条件を満たすデータセットの 1 つである。

SOTARO-1 および SOTARO-2 は、当研究室で開発を進めているロボット SOTARO で収集したデータを再構成して作成したデータセットである。SOTARO の詳細については 5.2 節で改めて後述する。2 種類のデータセットは別日に収集されており、服装等見え方の異なる同一人物のサンプルが含まれている。

表 1 データセット構成

	PRW	SOTARO -1	SOTARO -2
撮影方法	固定カメラ 6 台	移動ロボット搭載のカメラ	
画像枚数	約 34,000 枚	約 16,000 枚	
含まれる 人数	学習用 483 人 評価用 450 人	41 人	48 人
用途	特徴選択 CNN の学習	特徴選択 CNN の学習・ 再同定 CNN の転移学習	

表 2 転移学習の性能評価

CNN の学習に使用する データセット		評価			
特徴選択	再同定	Top-1	Top-5	Top-10	Top-20
なし	SOTARO -1	66.4	83.9	89.4	94.5
PRW		78.9	90.6	94.9	98.3
SOTARO-2		80.3	91.0	95.3	99.3
なし	SOTARO -2	60.6	76.4	86.1	92.7
PRW		74.4	87.2	91.3	95.5
SOTARO-1		79.4	90.4	94.4	97.3

4.2 人物再同定 CNN の転移学習の評価

前述の 2 つのデータセットを用いて、特徴選択 CNN の学習と再同定 CNN の転移学習を実施した。評価には、分類確信度が高い上位 K 位に正解ラベルが含まれるか否かを測定する Top-K 精度を使用した。結果は表 2 に示す通りである。表からは、特徴表現の事前学習が再同定精度の向上に有効であることが分かる。また、特徴選択 CNN と再同定 CNN の学習に、類似の環境下で収集したデータセットを用いることで、更なる精度の向上に繋がることが分かった。

4.3 データセット分割による協調システムの評価

協調型ラベル付け処理が正確に行われたと仮定し、サンプル数の増加に対する人物再同定精度の推移を評価する。本実験では SOTARO データセットを 6 分割し、5 段階の再同定精度を比較する。表 2 は、特徴選択 CNN の学習に SOTARO-2 を、再同定 CNN の転移学習に SOTARO-1 を使用した場合の結果を、表 3 は逆の条件で実験を行った場合の結果を示している。学習用サンプルの増加に伴って再同定精度が改善されることから、多くのサンプルを早い段階で収集することが重要であるといえる。

表 3 データセット分割実験 1

	Stage-1	Stage-2	Stage-3	Stage-4	Stage-5
Sample 数	2,663	5,327	7,991	10,654	13,318
Class 数	35	40	40	40	40
Accuracy	23.4	40.2	60.5	64.2	81.4

表 4 データセット分割実験 2

	Stage-1	Stage-2	Stage-3	Stage-4	Stage-5
Sample 数	2,668	5,336	8,004	10,672	13,340
Class 数	28	42	45	48	48
Accuracy	32.5	52.1	46.1	64.7	80.7

5. TA 業務への適用

5.1 ロボットによる TA 支援

我々が TA を担当しているプログラミング演習授業ではこれまで、複数人の TA が、受講学生側からの意思表示（質問をするために手を挙げる・TA を呼ぶなど）に基づいて学習の支援を行ってきた。しかしこの方法では、質問する学生は限定され、TA とのコミュニケーションに抵抗のある学生の学習を支援することができない。事前に行ったアンケート調査からは、およそ 4 割の学生が TA とのコミュニケーションに何らかの問題を抱えていることが分かっている。こうした背景から我々は、授業にロボット TA を導入することを考えた。

5.2 学生証読み取りに基づく名前識別

ロボット (SOTARO[渥美 17]) を用いた TA 業務支援において、各個人に合わせたサービスを提供するためには、人物再同定に加えて名前等の個人識別子を把握することが重要である。本研究では図 5 に示すように、SIFT[Lowe 04]特徴とテンプレートマッチングによる学生証読み取り機能をロボットに搭載することで、受講学生の名前を外見画像に加えて取得し、名前から得られる情報を元にコミュニケーションを行う。

5.3 TA 業務支援の想定実験

図 6 に、実際にロボット TA による業務支援の実験を行った際の実験の流れを示す。まず、学生にプログラミングに関する基本的な知識を問う穴埋め問題を提示し、その後、ロボットがそれに関連するクイズを学生に出題する。クイズは問題・ヒント・答え・補足の 4 つで構成され、遠隔操作によって各種スクリプトの読み上げを行う。クイズは学生が希望した場合繰り返し出題し、最後にアンケートを実施する。

アンケートでは、実験協力者のうち 9 名から有効な回答が得られた。図 7 に結果の一部を示す。ロボットによるクイズの実施により、授業に関する理解を深める一定の効果が見られたといえる。今後もロボットがより多くの学生に受け入れられるよう、コミュニケーション対象の選択方法や学生への話しかけ方の検討が必要である。

6. むすび

本論では、特定の集団に定期的にサービスを提供する移動ロボット向けの人物再同定手法を提案した。公開されている大規模データセットと、移動ロボットで収集した自作のデータセットを用いた実験により、提案手法の有用性を評価した。

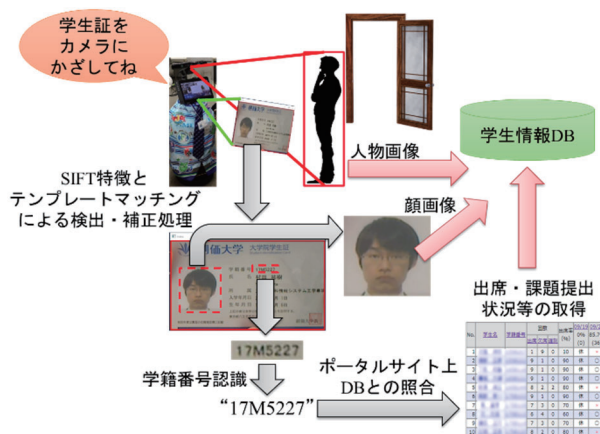


図 5 学生証読み取りと名前を介した情報取得

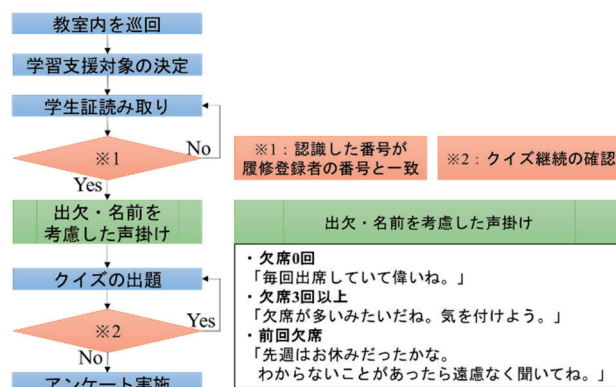


図 6 TA 業務支援実験の流れ

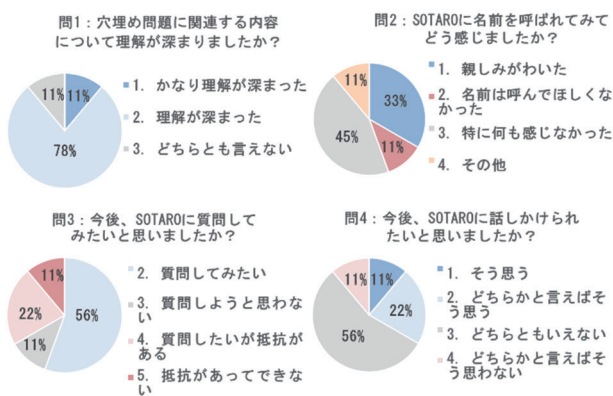


図 7 実験後のアンケート結果

また、提供するサービスの一例として演習授業の TA 業務支援を取り上げ、学生証読み取り機能を搭載したロボットを授業内で動作させる実験によって、その可能性を検証した。

参考文献

- [Radenovic 16] F. Radenovic, G. Tolias, and O. Chum: CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples, ECCV, 2016.
- [Schroff 15] F. Schroff, D. Kalenichenko, and J. Philbin: Facenet: A unified embedding for face recognition and clustering, CVPR, 2015.
- [高石 07] 高石悦史, 安村通晃: ロボット TA をもちいた対面型授業支援システムの試作, WISS2007 論文集, 2007.
- [Shiomi 15] M. Shiomi, T. Kanda, I. Howley, K. Hayashi, and N. Hagita: Can a social robot stimulate science curiosity in classrooms?, International Journal of Social Robotics, 2015.
- [Redmon 17] J. Redmon and A. Farhadi: YOLO9000: Better, Faster, Stronger, CVPR, 2017.
- [Zheng 17] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian: Person re-identification in the wild, CVPR, 2017.
- [渥美 17] 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.
- [Lowe 04] D. G. Lowe: Distinctive image features from scale-invariant feature points, International Journal of Computer Vision, 2004.