

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 24 日現在

機関番号：17102

研究種目：若手研究(B)

研究期間：2015～2016

課題番号：15K16028

研究課題名(和文)多様な人物属性の事前学習に基づく深層特徴を用いた人物画像の照合と検索

研究課題名(英文) Person re-identification and search using deep features based on pre-training of various human attributes

研究代表者

松川 徹 (Matsukawa, Tetsu)

九州大学・システム情報科学研究所・助教

研究者番号：80747212

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：カメラ間人物照合への応用を目指し、多様な人物属性(服装、所持物体など)を反映した深層特徴を開発した。畳込ニューラルネットワーク(CNN)の学習を人物属性ラベルの付与された歩行者画像データベースを用いて行い、中間層の特徴を取得し、カメラ間人物照合へ転用した。人物の判別性のより高い特徴をCNNに生成するため、人物属性の組み合わせを用いてCNNを学習する手法を提案した。この手法では、上半身の服装、下半身の服装、それらの色などの属性グループ間の組み合わせに着目し、CNNの学習するラベルの詳細化を行う。ベンチマークデータベースを用いた評価実験により、提案手法の有効性を定量的に確認した。

研究成果の概要(英文)：Aiming at to apply for person re-identification, we have developed deep features based on various human attributes (clothing, carrying object etc.). We conducted a learning of Convolutional Neural Network (CNN) using a pedestrian dataset which is annotated by human attributes. We extracted intermediate features from the learnt CNN and transferred them to person re-identification. To obtain more discriminative features in CNN, we developed a method to learn CNN using combination of human attributes. This method produces fine-grained labels for the CNN learning by focusing on the attribute combinations among different attribute groups, eg., upper body clothing, lower body clothing, and their colors. The quantitative evaluation on the benchmark datasets confirmed the effectiveness of the proposed method.

研究分野：パターン認識

キーワード：カメラ間人物照合 人物属性 深層特徴 CNN

1. 研究開始当初の背景

ある監視カメラで撮影された特定の人物画像を全身画像同士の照合により異なる監視カメラで撮影された多数の人物画像から見つける技術は、例えば駅構内の広域なエリアで人物追跡など、捜査現場への応用の期待が高まっている。しかし、異なるカメラ間では、人物の向き姿勢、照明、カメラ角度や背景に大きな変動が生じ、人物照合は大変困難な問題となる。

人物画像に対して服装や所持物体の種別に基づき、様々な意味概念のラベル(人物属性)を付与できる。人物属性は、人物の姿勢や照明条件などの変動によらない記述であり、人物属性を基に照合を行うことでより高い精度での照合が行えると期待される。また、人物属性を学習することにより、目撃証言を基にした人物画像検索を行うことも可能となる。

しかしながら、従来の研究では、人物属性の種類が 15 種類と少ないため、報告されている人物照合精度はあまり高くない。これはたとえば、鞆という人物属性があったとしても、ビジネス用のブリーフケースやハンドバック等、様々な鞆が存在するため、特定の人物を同定するための記述力が不足するためである。そのため、従来の研究よりも多様な人物属性を用いることが実用レベルでのカメラ間人物照合では必要不可欠となる。

2. 研究の目的

(1) 人物の持つあらゆる人物属性を認識することに有効な特徴を、あらかじめ設計することは難しい。学習データから高精度を対象の認識が可能な特徴を学習できる畳み込みニューラルネットワーク(Convolutional Neural Network:CNN)(引用文献)を用いて、人物属性を認識する技術を開発する。

(2) 人物の照合の精度向上のため、(1)のCNN で得られる特徴ベクトルを、人物照合のための判別度を加味して圧縮する。カメラ間人物照合の標準的なベンチマークである Viewpoint Invariant Pedestrian Recognition (VIPeR) データベース(引用文献)において、2014 年時点の最高認識率 35% を 15% 向上した 1 位認識率 50% を実現する。

3. 研究の方法

(1) CNN を用いた人物属性認識を行った。人物属性認識を行うデータベースとして、PEdesTrian Attribute (PETA) データベース(引用文献)を用いた。このデータベースにおいて複数の CNN の構成の認識精度を比較した結果、ImageNet で事前学習された AlexNet (引用文献)をこのデータベースで fine-tuning を実施した場合の精度が最も高かった。この AlexNet による人物属性認識の結果は、目撃証言などに基づく人物画像検索を行うためには、十分に精度が高いもので

はなかった。しかし、同一人物の画像をカメラ間で照合するための特徴を CNN から取得するためには、必ずしも個々の人物属性認識の精度が高精度である必要でない。そこで、人物属性認識を行う CNN としては AlexNet を用いることとした。

(2) カメラ間の人物画像を照合する特徴量として、人物属性識別用に学習された CNN から取得し、計量学習を適用した。PETA などのベンチマークデータベースで定義されている人物属性は、個別の人物を判断するためには粗すぎる。そこで、個別の人物属性よりも人物固有の情報を記述可能な人物属性の組み合わせに着目した(図 1)。これは、単にセーターを着用しているという属性と比較して、赤いセーターと青いジーンズを同時に着用しているという属性は個人を特定するための詳細な属性を記述していることに基づく。このような人物属性の組み合わせを識別するように CNN を学習することで、人物固有の特徴が得られると期待し、人物属性の組み合わせを識別する CNN を学習した。



図 1 : 人物属性の組み合わせの例

4. 研究成果

本研究の成果は、カメラ間人物照合のための(1)人物属性の組み合わせに基づく CNN 特徴(学会発表), 及び(2)階層的なガウシアン記述子(学会発表 -)の 2 点である。

(1) 人物属性の組み合わせに基づく CNN 特徴
提案 CNN 特徴の概要を図 2 に示す。CNN の構成としては AlexNet を用いる。AlexNet は 5 つの畳み込み層と 3 つの全結合層からなり、プーリング層が 1、2、5 層に取り付けられている。CNN は 1.2M 枚の画像から 1000 クラスの物体を認識する ImageNet の識別問題を識別するように初期化されている。提案手法は、補助データセットでの fine-tuning の実施(Phase 1)、ターゲットデータセットでの特徴抽出(Phase 2)の 2 段階からなる。

Phase 1. PETA データセットで、AlexNet の fine-tuning を実施する。このデータセットは、複数の人物属性グループにおける複数の人物属性ラベルが付与されている。したがって、CNN は以下の 2 つの補助タスクを連

带的に学習する。(a) 各人物属性グループの複数の人物属性の識別タスク。このタスクのため、出力層数が各属性ラベルと同等となる全結合層を取り付ける。(b) 組み合わせ人物属性の識別タスク。PETA データベースの各サンプルにつき、頻出する人物属性の組み合わせのインデクスである、組み合わせラベルを定義し、この組み合わせラベルを識別する。このタスクのため、出力数が組み合わせ属性ラベルと同等となる全結合層を取り付ける。誤差逆伝播法を用いて、(a)、(b) のタスクの損失関数を同時に最小化するように CNN を最適化する。

Phase 2. ターゲットとなる人物照合データセットで CNN 特徴を用いる。各データセットの画像から、4096 次元の特徴ベクトルを 1 番目の全結合層より抽出する。この特徴ベクトルに対して、対象データセットにおいて計量学習を適用することで、そのデータセットにおける人物の判別性を高める。

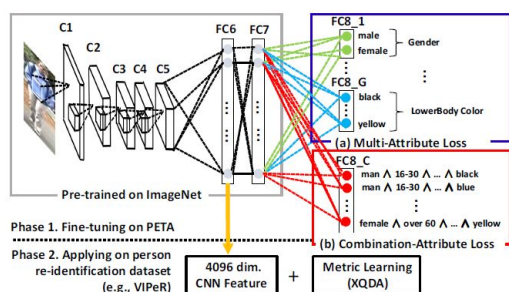


図 2 : 提案 CNN 特徴

Fine-tuning の設定

PETA データベースは、歩行者属性認識用の最大の公開データベースであり、61 個の人物属性の 19,000 枚の画像からなる。このデータベースの画像は、10 個の人物照合データベースから抽出されている。本研究の目的は、転移可能な CNN 特徴の設計にあるため、人物照合に用いるデータベースを除き、fine-tuning を実施する。

アノテーションが行われている全人物属性より、手動で 7 つ相互に排他的な人物属性グループを選び出した。それぞれ、性別、年齢、持ち物、上半身服装、上半身色、下半身服装、下半身色、である。各人物属性群に属する人物属性を表 1 に示す。

表 1 : 相互に排他的な人物属性群

Group (g)	Attributes
Gender	male, female
Age	less 15, 15-30, 31-45, 46-60, over 60
Luggage	backpack, other, folder, luggage case nothing, plastic bags, suitcase
UpperBody Clothing	sweter, tshit, suit, jacket, no sleeve, other
UpperBody Color	black, blue, brown, green, grey orange, pink, purple, red, white, yellow
LowerBody Clothing	suit, shorts, shirt skirt, long skirt trousers, hot pants, jeans, capri
LowerBody Color	black, blue, brown, grey, pink, red, white, yellow

評価結果

Fine-tuning が施された CNN 特徴量を用いて、カメラ間人物照合の評価を行った。評価には、ここで報告する VIPeR データベースの他に 3 つのデータベースを用いた。VIPeR データベースは、屋外で観測された 632 人のカメラペアの画像からなる。各画像ペアの半数の人数を学習画像として用い、それ以外の人物の画像をテスト画像に用いる。全手法で、特徴ベクトルに L2 ノルム正規化と Cross-view Quadratic Discriminant Analysis (XQDA) 計量学習(引用文献)を適用する。

CNN 特徴を用いた手法の結果を表 2 に示す。提案特徴 (Comb.+Multi) の 1 位認識率は 42.5% であった。(Multi-) は、提案手法において組み合わせ人物属性の損失を用いず、各人物属性群の複数の人物属性の識別タスクの損失のみを用いた場合、(Person-) は、損失関数として人物のラベルを用いた場合をそれぞれ示す。提案特徴はこれらベースライン手法比較して高い精度であると確認できる。これらの結果は、組み合わせ属性を識別する損失を CNN 特徴の学習に用いることの有効性を示す。

表において FFN は、従来手法である Feature Fusion Net を示す(引用文献)。(ImageNet) は、人物属性データベースにおける fine-tuning 実施前の AlexNet (引用文献)を用いた場合である、これらの結果と比較することにより、人物属性データベースにおけ fine-tuning の実施を行っている手法で大幅に認識率が向上していることが分かる。

表 3 に CNN 特徴と従来特徴を併用した場合の結果を示す。Local Maximal Occurrence (LOMO) 特徴量(引用文献)と提案特徴量を併用して用いることにより、目標値以上の 52.1% の 1 位認識率を達成している。従来は、FFN と LOMO 特徴を連結した特徴(引用文献)や、LBP, SIFT, ImageNet での事前学習された CNN 特徴のアンサンブル(引用文献)が高い性能を示していたが、提案特徴と LOMO 特徴量の併用はそれらを上回る性能であることが確認できる。

表 2 : CNN 特徴量の Rank 認識率

	Rank1	Rank10	Rank20
提案法 (Comb.+Multi)	42.5%	83.0%	92.0%
(Multi)	39.6%	81.5%	90.6%
(Person)	37.9%	78.5%	88.4%
従来法 (FFN)	31.8%	73.7%	85.3%
(ImageNet)	19.7%	58.1%	72.9%

表 3 : CNN 特徴と従来特徴との併用による Rank 認識率

	Rank1	Rank10	Rank20
提案法+LOMO	52.1%	89.2%	95.0%
FFN+LOMO+KMFA	51.1%	91.4%	96.9%
MetricEnsemble	45.9%	88.9%	95.8%

(2) 階層的なガウシアン記述子

人物画像の照合のためには、色やテクスチャなどの特徴量を記述子が特に重要な役割を果たす。そこで本研究の派生研究として、カメラ間人物照合に有効な階層的な記述子として、画素特徴ベクトルの2段階のガウス分布に基づく特徴記述子(Gaussian Of Gaussian: GOG 記述子)を開発した(成果-)。この特徴単体でVIPeR データベースにおける1位認識率約50%を達成した(表4)。この成果はトップ国際会議CVPRに採択され、カメラ間人物照合分野において大きな注目を集めた。

表4 : GOG 記述のランク認識率.

	Rank1	Rank10	Rank20
GOG 記述子	49.7%	88.7%	94.5%

[引用文献]

A.Krizhevsky, I.Sutskever, and G.E.Hinton, ImageNet Classification with Convolutional Neural Networks, In Proc. NIPS, 2012.

D.Gray, and H.Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in ECCV 2008.

Y.Deng, P.Luo, C.C.Loy, and X.Tang, Pedestrian Attribute Recognition at far distance, in Proc. ACM MM, 2014.

S.Liao, Y.Hu, X.Zhu, and S.Z.Li, Person Re-Identification by Local Maximal Occurrence Representation and Metric Learning, In Proc. CVPR2016.

S.Wu, Y-C. Chen and W.S.-Zheng, An enhanced deep feature representation for person re-identification, in Proc. WACV2016.

S.Paisitkriangkrai, C.Shen, A. van den Hengel, Learning to rank in person re-identification with metric ensemble, In Proc. CVPR2015.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 0 件)

[学会発表](計 4 件)

Tetsu Matsukawa, Einoshin Suzuki, "Person Re-Identification Using CNN Features Learned From Combination of Attributes", in Proceedings of 23rd International Conference on Pattern Recognition (ICPR2016), pp.2429-2434, Cancun (Mexico), 2016年12月7日.

松川徹, 岡部孝弘, 鈴木英之進, 佐藤洋

一, "階層的なガウシアン記述子を用いたカメラ間人物照合", 第6回バイオメトリクスと認識・認証シンポジウム(SBRA2016), S3-4, 芝浦工業大学(東京都), 2016年11月17日(招待発表).

Tetsu Matsukawa, Takahiro Okabe, Einoshin Suzuki, Yoichi Sato, "Hierarchical Gaussian Descriptor for Person Re-Identification (CVPR2016)", 第19回画像の認識・理解シンポジウム(MIRU2016), IS2-10, アクティシティ浜松(静岡県), 2016年8月4日(招待講演).

Tetsu Matsukawa, Takahiro Okabe, Einoshin Suzuki, Yoichi Sato, "Hierarchical Gaussian Descriptor for Person Re-Identification", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016), pp.1363-1372, Las Vegas (USA), 2016年6月27日.

[図書](計 0 件)

[産業財産権]

出願状況(計 0 件)

取得状況(計 0 件)

[その他]

ホームページ等

カメラ間人物照合の特徴量抽出コード, および抽出された特徴量を公開している.

<http://www.i.kyushu-u.ac.jp/~matsukawa/ReID.html>

6. 研究組織

(1) 研究代表者

松川 徹 (MATSUKAWA TETSU)

九州大学・システム情報科学研究院・助教
研究者番号: 80747212

(2) 研究分担者 なし

(3) 連携研究者 なし

(4) 研究協力者 なし