

# STHOG 特徴に基づく歩行者照合の CNN による高精度化

柏本 雄士朗<sup>1,a)</sup> 村松 大吾<sup>1,b)</sup> 武村 紀子<sup>1,c)</sup> 八木 康史<sup>1,d)</sup>

**概要:** 防犯カメラ映像などを用いた歩行者照合は、犯罪捜査や人流解析への応用が期待されている。歩行者照合を実現する主要なアプローチの一つに、歩容特徴を用いた手法があるが、この手法は安定した一歩行周期の歩行画像列が得られることが前提となっており、一歩行周期分の画像列が取得できない場合は適用ができない。しかし実環境を考慮した場合、カメラの画角や遮蔽物、対象人物の振り向きや躓きなどの動作により、一歩行周期分の安定した歩行画像列が得られない場合が多い。そのため、本研究では一歩行周期の歩行画像列が取得できない場合でも利用可能な STHOG 特徴などの特徴量に注目し、CNN によってそれらを用いた照合の高精度化を行うことで三フレームの歩行画像列からの歩行者照合手法を提案する。本稿では、提案手法の初期評価実験結果を報告する。

## Improvement of STHOG feature-based gait person re-identification with CNN

### 1. はじめに

近年、商業施設や駅など街中には多くの防犯カメラが設置されており、現在では犯罪の抑止などに用いられている。それらの防犯カメラを用いて画像による歩行者の認証や照合を行う研究は犯罪捜査や人流解析などへの応用が期待されている。そのような研究の中でも、視界を共有しない防犯カメラ間で人物の照合を行う歩行者照合は、顔や歩容などのバイオメトリクス情報や、服装などの色やテクスチャの情報、周囲の人物との位置関係や集団単位での情報、また、それらを補助するためにカメラの位置関係と移動に要する時間による制約情報など様々な情報を用いる手法が提案されている。これらの中で顔による手法は、高解像度かつ対象人物の顔が映っている画像が不可欠であるため、防犯カメラの映像を用いることを考えた場合には後ろ向きの人物などには適用することができず、適用可能シーンが限定されてしまう。顔が見えなくても個人を特定できる手法として、歩容認証 [1] も適用可能と考えられる。しかし、従来使われている多くの歩容特徴に基づく手法が一歩行周期

分の歩行画像列が安定して得られることを前提としているため、遮蔽物や対象人物の行動の影響により一歩行周期の歩行画像列が得られないシーンでは適用が難しい。現在最も多く報告されている手法は人物の画像の色やテクスチャの情報を用いるものである (例えば [1] など)。最近では深層学習の普及と共に、人物の画像を入力とした畳み込みニューラルネットワーク等を用いることで特徴を抽出し、それによる歩行者照合を行うことで従来に比べて高い精度を実現している研究も多くみられる [2], [3], [4], [5], [6]。また、そのような研究の中には画像を上中下に 3 分割して与える手法 [7] や、画像のどの部位に注目を行えば良いかの学習も並行させる手法 [8], [9], [10] など、上半身と下半身で分かれることの多い服装の情報や画像内の人物領域の情報を積極的に利用する取り組みも行われている。他にも 1 枚の画像のみでなく、画像列を入力としたネットワークによる研究 [11], [12], [13], [14] や、画像による情報にカメラ間の移動に要する時間的な制約を加えて精度を向上させた研究 [15] などが報告されている。しかし、色情報に基づく手法では、学習データと評価データ間や評価データのカメラ間で照明状況などの撮影環境が異なる場合に対しては精度が悪化する。また、スーツやコートなど類似した服装の人物が多いシーンにおいては、照合精度に限界がある。その為、本研究では、照明環境の変化にも強い歩行者の動作特徴も考慮した照合を行うことを目的として、主に動

<sup>1</sup> 大阪大学  
Osaka University

a) kashimoto@am.sanken.osaka-u.ac.jp

b) muramatsu@am.sanken.osaka-u.ac.jp

c) takemura@ids.osaka-u.ac.jp

d) yagi@am.sanken.osaka-u.ac.jp

作(歩容)とテクスチャの情報を取り扱っていると考えられるSTHOG (Spatio Temporal Histograms of Oriented Gradients) 特徴 [16] に注目する。STHOG 特徴は従来、歩行者照合や歩容認証などに用いられており [17], [18] 他の一歩行周期分の画像系列を必要とする歩容特徴と違い三フレームの画像列から抽出できる。本研究ではSTHOG 特徴をCNN (畳み込みニューラルネットワーク) の入力として用いることで、従来のSTHOG 特徴を用いた手法の精度改善を実現する。

STHOG 特徴は、局所領域の輝度の勾配方向をヒストグラム化した特徴であるHOG (Histograms of Oriented Gradients) 特徴 [19] を時間方向に拡張し、3枚の時系列画像から計算された空間ヒストグラムと時間ヒストグラムからなる特徴量であり、従来歩行者に対する人物検出や人物照合に用いられてきた。本研究では、STHOG 特徴を用いた歩行者照合を高精度化させるため、CNNを導入する。具体的には、大きく分けて次の3ステップにより照合を行う。

- (1) 学習用データセットに含まれる歩行者の歩行画像列から抽出されたSTHOG 特徴を入力として、CNNによる人物分類器を学習する。
- (2) 学習された人物分類器のネットワークにおける中間層出力を人物特徴と考え、特徴抽出器として用いる。
- (3) 得られた人物特徴どうしを比較することで相違度を求め、その相違度により人物の対応付けを行う。

本研究においては、上記のフレームワークを用いて10,307人分の歩行者の歩行画像列からなるデータセットであるOUMVLP データセット [20] を用いて実験を行う。また、同時に当手法の照明などの撮影環境の変化への頑健性を確認するため、元となる画像列に対してガンマ補正をかけることで疑似的に明るさを変化させたデータセットも用意し評価に用いた。

## 2. 提案手法

本節では、提案手法であるCNNを用いたSTHOG 特徴による歩行者照合について述べる。

### 2.1 STHOG 特徴

STHOG 特徴とは、画像内の  $(i, j)$  の位置にあるピクセルにおける、それぞれ左右、上下方向に隣接したピクセルどうしの輝度値の差である  $I_x(i, j)$ ,  $I_y(i, j)$  を用いて、エッジの勾配方向  $\phi(i, j) = \tan^{-1}(I_y(i, j)/I_x(i, j))$  をピン、勾配強度  $I(i, j) = \sqrt{I_x(i, j)^2 + I_y(i, j)^2}$  を度数として作成されたヒストグラム特徴であるHOG 特徴を時間方向に拡張し、三フレームの時系列画像に対して適用したものである。

まず、対象人物の連続する三フレームの時系列画像に対して共通して用いるウィンドウを設定する。ウィンドウ内の画像に対し一定のサイズとなるセルを設定し、セル内の

各ピクセルに対して時系列画像から得られるエッジの空間勾配方向  $\phi$  と時間勾配方向  $\theta$ , 勾配強度  $I$  を計算し、それらを用いることで、セルにおける空間勾配方向と時間勾配方向のそれぞれをピンとし、勾配強度を度数としたヒストグラムを作成する。セルをウィンドウ内でスライドさせつつ各セルにおいて同様にヒストグラムを作成し、一定のセルの集まりとなるブロックにおいて正規化されたヒストグラムをベクトルとして並べたものがSTHOG 特徴における特徴ベクトルである。空間勾配方向により作成されたヒストグラムはHOG 特徴と同様のものとなり、時間勾配方向により作成されたヒストグラムは対象となる時系列画像内の物体の動きを表す特徴となる。

位置  $(i, j)$  のピクセルにおける勾配強度  $I$  と、前述の空間勾配方向  $\phi$ , 時間勾配方向  $\theta$  は、左右、上下、時間的前後の輝度値の差  $I_x(i, j)$ ,  $I_y(i, j)$ ,  $I_t(i, j)$  を用いて以下のように設定する。

$$I(i, j) = \sqrt{I_x(i, j)^2 + I_y(i, j)^2} \quad (1)$$

$$\phi(i, j) = \tan^{-1}\left(\frac{I_y(i, j)}{I_x(i, j)}\right) \quad (2)$$

$$\theta(i, j) = \tan^{-1}\left(\frac{I_t(i, j)}{\sqrt{I_x(i, j)^2 + I_y(i, j)^2}}\right) \quad (3)$$

### 2.2 CNN による歩行者照合

図1は提案手法であるSTHOG 特徴を入力としたCNNによる歩行者照合の概要を示したものである。提案手法では、まず人物IDをラベルとして有する歩行画像列を学習データセットとして用いて人物分類器の学習を行う(図1(1))。次に、学習された人物分類器に評価データセットの各データを入力し、中間層出力を得る(図1(2))。これにより得られた中間層出力は、入力されたデータからその人物固有の情報が抽出された特徴量とみなすことができる。つまり、STHOG 特徴からより洗練された特徴を抽出したものとみることができる。照合の際には、中間層出力を照合用特徴として用い、比較する画像列間からの特徴の差のL2ノルムを計算することで歩行者照合を行う(図1(3))。

## 3. 実験

本研究では、学習データセットと評価データセットで撮影環境が同一のデータと、ガンマ補正により学習データセットと評価データセットで明るさが疑似的に異なるデータの2種類データによる実験を行った。また、比較手法として類似のCNNを用いたシルエット画像列やSTHOG 特徴の抽出を行う前のグレースケール画像列による歩行者照合の評価、及びCNNを使用せずシルエット特徴どうしやSTHOG 特徴どうしをL2ノルムによって比較することにより相違度を計算した場合の評価も行った。実験の評価

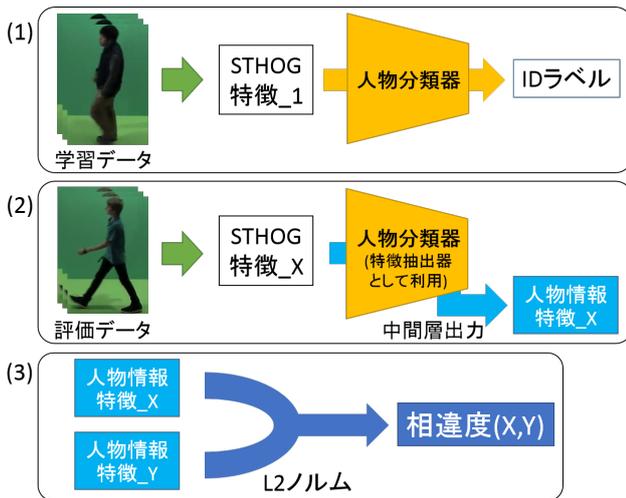


図 1 提案手法の概要



図 2 OUMVLP の歩行者画像の例

指標としては、ある入力データに対し、相違度を基に登録データの順位付けを行った場合に、本人の登録データが1位及び5位までに含まれる割合である1位認証率 (Rank1) と5位認証率 (Rank5) と、入力データと登録データの相違度と本人受入のしきい値と比較することで、他人が本人として判断される割合である他人受入誤り率 (False acceptance rate, FAR) と、本人が他人として判断される割合である本人拒否誤り率 (False rejection rate, FRR) を考えた場合にしきい値を変化させ、FAR と FRR が等しくなった時のそれらの値である等価誤り率 (Equal error rate, EER) を用いた。

### 3.1 データセット

実験には OUMVLP(OU-ISIR Gait Database, Multi-View Large Population Dataset) の撮影角度 90°のデータを使用した。このデータは、グリーンバックを背景にして歩く 10,307 人の歩行者を進行方向に対し、左回りで 90°の方向にあるカメラにより 25fps で撮影したものであり、各歩行者は同一の状況で歩行を行い、それぞれ一歩行周期分の 2 つの歩行画像列データを持つ。今回の実験では、それらの歩行画像列から歩行者の高さを基準として 11:16 のアスペクト比で歩行者領域を切り出したのち、幅 88(pixel)、高さ 128(pixel) にリサイズして用いた。

用いたデータの例を図 2 に示す。また、これらのデータを用いて STHOG 特徴を抽出する際の設定として、セルのサ

イズは 9(pixel)×9(pixel)、ストライドは 1(pixel)×1(pixel) とし、空間勾配方向、時間勾配方向共にピンの数は 9 とした。また、セルの正規化のためのブロックのサイズは 2(cells)×2(cells) とし、正規化の方法は以下の式に従った。

$$h_{norm}(n) = \frac{h(n)}{H} \quad (4)$$

$$H = \sqrt{\left( \sum_{k=1}^{2 \times 2 \times 9} h(k)^2 \right) + \epsilon} \quad (5)$$

ただし、 $h(n)$  は正規化前のブロックに含まれる  $n$  番目のヒストグラムの度数、 $h_{norm}(n)$  は正規化後のブロックに含まれる  $n$  番目のヒストグラムの度数、 $\epsilon$  はゼロ除算を防ぐための微小な定数とする。

OUMVLP の全 10,307 人分のデータのうち、データが欠落している者を除き ID 番号が奇数番のデータを学習用データに、偶数番のデータを評価用データとして用いた。また、評価データの 2 つある歩行画像列のうち片方は登録データとして、もう片方は入力データとして扱う。ただし、登録データが欠落している ID のデータは入力データからも取り除いている。学習データには 4,971 人分の合計 226,891 個の三フレームからなる画像列データが含まれており、評価データのうち登録データは 4,977 人分の 113,605 個の画像列データが、入力データには 4,862 人分の 111,451 個の画像列データが含まれている。

### 3.2 CNN の設定

STHOG 特徴を入力とする CNN による人物分類器の構造とパラメータに関しては図 3 のように設定した。また、比較手法として用いたシルエット画像列とグレースケール画像列を入力とする CNN による人物分類器は図 4 のように設定した。なお、それぞれの図中の Conv2D は畳み込み層を Max pool 2D はプーリング層を Full connection は全結合層を表している。各 CNN を用いて学習を行う際、それぞれの学習データに対しては全学習データの平均値を減算し、その後減算後の全学習データの最大値によって除算することでデータを正規化した。学習後の CNN を用いた評価データの特徴抽出を行う際、学習データに対して行った処理と同様の正規化処理を学習データと同じ数値を用いて行った。また、学習の際のバッチサイズは STHOG 特徴や画像列データ 100 個、エポック数は 1000 で統一した。

### 3.3 学習データと評価データの撮影環境が同一の場合の実験

#### 3.3.1 実験設定

OUMVLP データセットをそのまま用い、学習データと評価データの撮影環境が同一の状況での実験を行った。その際、歩行者照合の結果は歩行の位相に影響を受けることが考えられるため以下の 2 通りの実験を実施した。

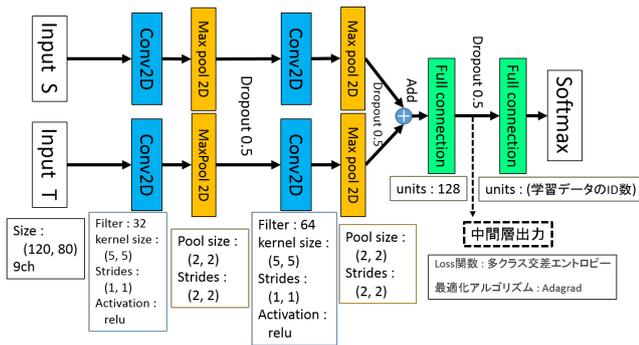


図 3 STHOG 特徴を入力とした CNN による人物分類器の構造

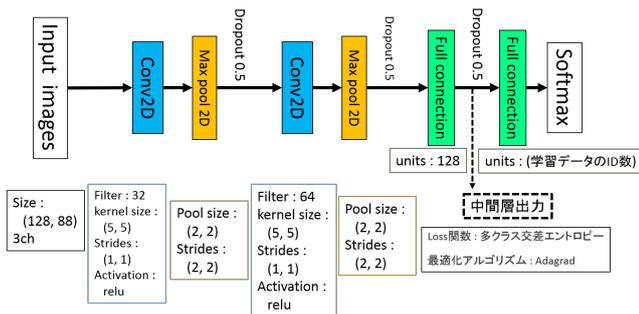


図 4 シルエット画像列或いはグレースケール画像列を入力とした CNN による人物分類器の構造

(1) 登録データに一歩行周期分のデータが利用可能な場合の評価。評価用データが多すぎるため、評価時の入力データは、一歩行周期の画像系列の中から連続する 3 枚の画像セットを一組ランダムに選択した。

(2) 登録データに連続した三フレーム画像のみ利用可能な場合の評価。評価時の入力データは、一歩行周期の画像系列の中から連続する 3 枚の画像セットを一組ランダムに選択した。

(1) の実験では各 ID の一歩行周期分の登録データと 1 つの入力データの相違度のうち、最小値を ID 間の相違度として設定した。これらの実験においてシルエット画像列とグレースケール画像列を入力として与えた場合を比較実験として行い、(1) の実験ではそれらに加えて、CNN を用いずに L2 ノルムによって相違度を計算した場合の実験も行った。

### 3.3.2 実験結果

登録データとして一歩行周期分のデータが利用可能な場合の実験の結果を表 1 に、登録データとして連続した三フレーム画像のみ利用可能な場合の実験の結果を表 2 に示す。

単純に L2 ノルムでの比較を行う場合に比べ、CNN によって人物情報を抽出することで精度に大幅な改善がみられる。また、シルエット画像列とグレースケール画像列、STHOG 特徴を比較すると、グレースケール画像が最もよい精度を示した。また、実験 (2) では歩行の位相の影響によりどの手法においても精度が低下している。

表 1 一歩行周期分の画像を登録画像として利用できる場合の精度 (評価 (1))

Method	Rank1 (%)	Rank5 (%)	EER (%)
Silhouette + L2	51.5	63.1	17.13
STHOG + L2	66.6	75.8	9.11
Silhouette + CNN	86.5	92.9	3.25
Grayscale + CNN	95.5	98.6	0.88
STHOG + CNN	91.4	96.9	1.01

表 2 連続する 3 枚の画像列のみが登録されている場合の精度 (評価 (2))

Method	Rank1 (%)	Rank5 (%)	EER (%)
Silhouette + CNN	68.9	83.3	5.62
Grayscale + CNN	90.2	96.2	1.54
STHOG + CNN	82.6	92.3	1.52

## 3.4 学習データと評価データの撮影環境が異なる場合の実験

### 3.4.1 実験設定

学習データは OUMVLP データセットをそのまま使い、評価データはガンマ補正によって疑似的に明るさを変化させ、STHOG 特徴とグレースケール画像を入力とした実験を行った。ただし、グレースケール画像列に対しては通常のものに加え、3.2 章で述べた正規化を行わず、代わりに予めすべての学習データと評価データの画像列に対して、それぞれの画像列における輝度値の平均と標準偏差を用いて Z 標準化したものの両方の評価を行った。

実験は 3.3.1 章の実験 (1) に準拠し、一歩行周期分の登録データに対して 1 つの入力データを用いた。ガンマ補正による明るさの変化については、以下の 2 通りの実験を行う。

- 入力データ・登録データ両方に同じガンマ補正を行った実験
- 入力データのみに対してガンマ補正を行った実験

また実験におけるガンマの値は 0.9, 0.7, 0.5 を用いた。それぞれのガンマの値での画像の例を図 5 に示す。

### 3.4.2 実験結果

表 3 は入力データ・登録データ両方に同じガンマ補正を行った実験の結果である。グレースケール画像列を用いた場合の結果が共にガンマ値が 0.7 以下で精度を落としているのに対し、STHOG 特徴を用いた場合ではガンマ値が 0.7 では精度の低下は見られず、ガンマ値 0.5 での精度低下も抑えられている。

表 4 は入力データのみに対してガンマ補正を行った実験の結果である。Z 標準化を行っていないグレースケール画像列を用いた場合の結果は、ガンマ値が低くなるにつれて、入力・登録間での明るさの差が広がるとともに精度の低下が深刻なものとなっている。Z 標準化を行ったグレースケール画像列を用いた場合の結果は、Z 標準化を行っていない結果に比べて精度の低下を抑えられてはいるものの、ガンマ値 0.7 以下では STHOG 特徴を用いた場合の結

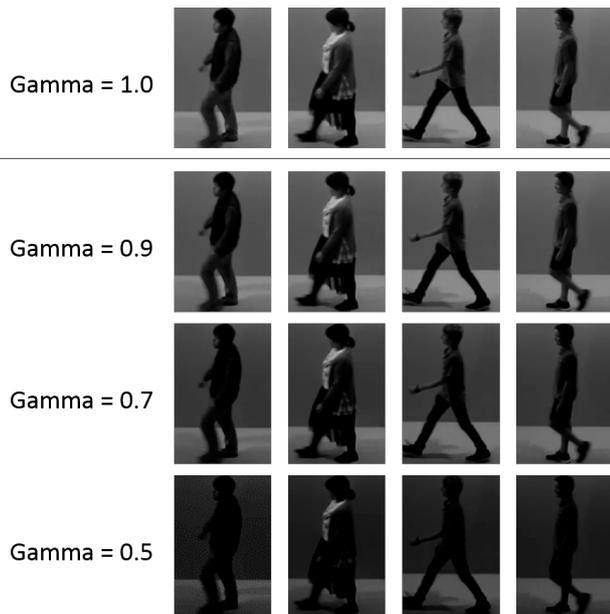


図 5 ガンマ補正を行った画像の例

表 3 入力データ・登録データ両方に同じガンマ補正を行った実験結果 (数字は%)

gamma	Method	Rank1	Rank5	EER
1.0	Grayscale + CNN	95.5	98.6	0.88
	Grayscale (Z) + CNN	92.8	97.2	2.53
	STHOG + CNN	91.4	96.9	1.01
0.9	Grayscale + CNN	94.4	98.0	1.03
	Grayscale (Z) + CNN	91.2	96.7	2.53
	STHOG + CNN	91.2	97.0	0.98
0.7	Grayscale + CNN	88.6	94.8	2.27
	Grayscale (Z) + CNN	84.6	93.0	2.89
	STHOG + CNN	90.9	96.9	0.99
0.5	Grayscale + CNN	79.9	90.2	3.11
	Grayscale (Z) + CNN	70.6	80.5	6.36
	STHOG + CNN	88.8	95.9	1.10

表 4 入力データのみに対してガンマ補正を行った実験結果 (数字は%)

gamma	Method	Rank1	Rank5	EER
1.0	Grayscale + CNN	95.5	98.6	0.88
	Grayscale (Z) + CNN	92.8	97.2	2.53
	STHOG + CNN	91.4	96.9	1.01
0.9	Grayscale + CNN	21.0	40.1	5.27
	Grayscale (Z) + CNN	91.7	96.9	2.56
	STHOG + CNN	91.2	96.9	0.96
0.7	Grayscale + CNN	2.4	6.3	26.24
	Grayscale (Z) + CNN	66.5	85.6	3.19
	STHOG + CNN	87.4	95.8	1.16
0.5	Grayscale + CNN	0.2	0.7	39.26
	Grayscale (Z) + CNN	16.2	31.7	12.0
	STHOG + CNN	67.3	86.1	2.51

果に大きく差を付けられている。

### 3.4.3 考察

グレースケール画像列を用いる手法に対して、STHOG 特徴を用いる手法は、学習データと評価データ間、及び評価データ内のカメラごとのデータに明るさの違いが生じた場合でも精度の低下を大きく抑えられている。この結果から、グレースケール画像列を用いた場合、CNN は主に画像内の輝度値の情報を学習しているのに対して、グレースケール画像から歩行動作や輝度勾配の情報を抽出した特徴量である STHOG 特徴は、明るさの変化に対して頑健であると考えられる。

## 4. 結論

本研究では CNN を用いて STHOG 特徴による歩行者照合の精度を向上させることで、照明環境の変化に強い歩行者照合を行うことを提案した。また、OUMVLP を用いた実験により、CNN を用いることで STHOG 特徴による歩行者照合の精度が向上すること、及び照明環境の変化により色情報が変わる状況での歩行者照合における STHOG 特徴の有用性を確認した。

## 参考文献

- [1] Jinguang Han and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, No. 2, pp. 316–322, 2006.
- [2] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deep-reid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152–159, 2014.
- [3] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1335–1344, 2016.
- [4] Ejaz Ahmed, Michael Jones, and Tim K Marks. An improved deep learning architecture for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3908–3916, 2015.
- [5] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1249–1258, 2016.
- [6] Xiaobin Chang, Timothy M Hospedales, and Tao Xiang. Multi-level factorisation net for person re-identification. In *CVPR*, Vol. 1, p. 2, 2018.
- [7] Evgeniya Ustinova, Yaroslav Ganin, and Victor Lempit-sky. Multi-region bilinear convolutional neural networks for person re-identification. In *Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on*, pp. 1–6. IEEE, 2017.
- [8] Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, and Wanli Ouyang. Attention-aware compositional network for person re-identification. *arXiv preprint*

- arXiv:1805.03344*, 2018.
- [9] Liming Zhao, Xi Li, Yueting Zhuang, and Jingdong Wang. Deeply-learned part-aligned representations for person re-identification. In *ICCV*, pp. 3239–3248, 2017.
  - [10] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1077–1085, 2017.
  - [11] Lin Wu, Chunhua Shen, and Anton van den Hengel. Deep recurrent convolutional networks for video-based person re-identification: An end-to-end approach. *arXiv preprint arXiv:1606.01609*, 2016.
  - [12] 加藤直樹, 箱崎浩平, 田麻雅基, 古山純子, 里雄二, 青木義満. 畳み込みニューラルネットワークによる距離学習を用いた動画像人物再同定. 精密工学会誌, Vol. 83, No. 12, pp. 1117–1124, 2017.
  - [13] Niall McLaughlin, Jesus Martinez del Rincon, and Paul Miller. Recurrent convolutional network for video-based person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1325–1334, 2016.
  - [14] Jiahuan Zhou, Bing Su, and Ying Wu. Easy identification from better constraints: Multi-shot person re-identification from reference constraints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5373–5381, 2018.
  - [15] Jianming Lv, Weihang Chen, Qing Li, and Can Yang. Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7948–7956, 2018.
  - [16] Chunsheng Hua, Yasushi Makihara, and Yasushi Yagi. Pedestrian detection by using a spatio-temporal histogram of oriented gradients. *IEICE TRANSACTIONS on Information and Systems*, Vol. 96, No. 6, pp. 1376–1386, 2013.
  - [17] Ryo Kawai, Yasushi Makihara, Chunsheng Hua, Haruyuki Iwama, and Yasushi Yagi. Person re-identification using view-dependent score-level fusion of gait and color features. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 2694–2697. IEEE, 2012.
  - [18] Hazem El-Alfy, Daigo Muramatsu, Yuuichi Teranishi, Nozomu Nishinaga, Yasushi Makihara, and Yasushi Yagi. A visual surveillance system for person re-identification. In *Thirteenth International Conference on Quality Control by Artificial Vision 2017*, Vol. 10338, p. 103380D. International Society for Optics and Photonics, 2017.
  - [19] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, pp. 886–893. IEEE, 2005.
  - [20] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSSJ Transactions on Computer Vision and Applications*, Vol. 10, No. 1, p. 4, 2018.