人物再同定のための重み付き特徴統合の検討*

阿部 香織** 宮下 侑大*** 片岡 裕雄† 中村 明生††

Weighted Feature Integration for Person Re-identification

Kaori ABE, Yudai MIYASHITA, Hirokatsu KATAOKA and Akio NAKAMURA

Person re-identification is the task of finding and matching the same individual in different camera views. For robust person re-identification, we propose a weighted feature integration method that adapts to illumination changes and appearance differences caused by different camera views. First, we extract four kinds of local features (color histograms, frequency features, gray-level co-occurrence matrices, and histogram of oriented gradient features) from the image of a person as cloth appearance information. Second, in the pre-training phase, we calculate the difference in value of each local feature for a pair of images taken by different cameras. Local features are then weighted and integrated based on these differences. We tested three weighting functions: reciprocal, probability density function, and average Bhattacharyya distance. In the experiments, we utilized four public datasets, iLIDS-VID, GRID, PRID, and VIPeR, and verified the effectiveness of the proposed method. The results demonstrate a general improvement in person re-identification performance when the feature integration is weighted by the average Bhattacharyya distance.

Key words: person re-identification, appearance-based features, feature integration, weighting

1. 緒 言

監視カメラの増加に伴う映像の自動解析技術の需要から、図1のような複数のカメラを連携し同一人物を照合する問題は人物再同定と呼ばれ、重要課題の1つである¹⁾²⁾。人物再同定は、商業施設における防犯やマーケティングなど複数分野での応用が可能であり、コンピュータビジョン分野で広く研究されている。

人物再同定では外観情報を利用し照合を行うことが多く、主に顔、歩容、着用衣服の局所特徴を利用した手法が提案されている。それぞれの従来手法の詳細は第2章で示すが、緒言においても簡単に述べる。

Fig. 1 Example of Person Re-Identification

顔は個人の同定に有効であるが、歩容や衣服に比べて画像内に占める割合が小さいため、カメラと人物の距離が離れた場合、精度低下が否めない。また、歩容も同定に有効な特徴を抽出可能であるが、外観情報を利用する場合は視点変化の影響を受けやすい。ところで、衣服領域は画像内において比較的大きな面積を占め、特徴抽出に有利である。顔、歩容、衣服とそれぞれ長所短所があるが、本稿においては衣服からの局所特徴を用いる手法に取り組む。

さて、衣服の外観からは色、模様、形状といった局所特徴を抽出可能であり、それらを用いた人物再同定手法も提案されている。しかし、これらの特徴はカメラ間の視点変化に伴う照明変動や見えの違い(衣服の模様、形状情報の変動)といった、撮影条件の変動の影響を受けやすい。これに対応するために、局所特徴を使い分ける、あるいは連結による特徴統合を行う、などの手法が提案されている。前者に関し、適切に使い分けるためには撮影環境を予め知悉する必要がある。また、後者に関しては、本稿の先行研究として大槻らが複数局所特徴の連結統合を行っている³⁾。ただ、深度情報による方向推定をして横向きの際にのみ固定値の重み付けを行っているだけであり、それ以外では単純連結による特徴統合である。すなわち、撮影条件の変動に対して、本来利用すべきではない脆弱な局所特徴に対して不適切な重み付けがなされている可能性がある。

これに対し、本稿では衣服の外観情報を利用することを前提として、照明変動や衣服の模様・形状などの見えの違いといった、視点変化による撮影条件の変動が存在する人物再同定に対し、汎用的な特徴統合のフレームワークを提案する。複数の局所特徴を用いた照合に際して、局所特徴の差分による撮影条件の変動の表現、およびそれに基づく重み付け手法の選定を行い、人物再同定を行う。

以降、第2章では人物再同定問題における従来研究について概説する。第3章では提案するフレームワークについて詳述する。第4章では提案手法の有用性を示すため、提案手法による人物再同定に関する評価実験、およびその結果について考究を

* 原稿受付 平成 28 年 5 月 11 日

掲載決定 平成 28 年 8 月 17 日

** 学生会員 東京電機大学大学院 (東京都足立区千住旭町 5)

*** 非 会 員 東京電機大学大学院 (現, (株)シー・エス・イー; 東京都渋谷区渋谷 3-3-1)

† 非 会 員 産業技術総合研究所 (茨城県つくば市梅園 1-1-1)

†† 正 会 員 東京電機大学大学院 (東京都足立区千住旭町 5)

Fig. 2 Overview of our proposed method

加える。以上を踏まえたうえ、第5章で本稿を総括する。

2. 従来研究

前述の通り、人物再同定では外観情報を利用し照合を行うことが多く、主に顔、歩容、着用衣服の特徴を利用した手法が提案されている。

顔に基づく人物再同定の従来研究について述べる。Viola & Jones は Haar-like 特徴と AdaBoost を組み合わせることで顔検出を実用レベルとし、顔認証研究を活性化した⁴⁾。Li らは顔全体を複数局所領域へ分割することを繰り返すことで得られる Probabilistic Elastic Part モデルという階層的なグラフモデルを提案し、表情変化に頑健な顔認証を提案した⁵⁾。Hassner らは目や鼻などの器官位置、および顔の輪郭である face-alignment を利用し顔の見えの違いに頑健な手法を提案した⁶⁾。以上より、顔は表情変化や頭部の姿勢変化による見えの違いに頑健であるため、人物再同定に有効である。しかし、歩容や衣服に比べ、画像内に占める割合が小さいため、カメラと人物の距離が離れた場合、精度低下が否めない。

歩容に基づく人物再同定の従来研究について述べる。楨原らは歩行中のシルエット画像から取得した歩行の周波数成分を歩容特徴として利用し人物再同定を実現した⁷⁾。Wang らは HOG3D を用いて歩幅や歩行周期を抽出し、外観情報のみによる歩容認証を提案した⁸⁾。今成らは個人の癖といった微小動作が判別可能である Dense Trajectories を利用し、密な時系列特徴の取得、量子化および記述による人物再同定に取り組んだ⁹⁾。

さらに、川合らは歩容の方向変換モデルを提案し、カメラの視

点変化への対応に取り組んでいる¹⁰⁾。歩容も人物再同定に有効な特徴を抽出可能であるが、外観情報を利用する場合は視点変化の影響を受けやすい。

衣服に基づく人物再同定の従来研究について述べる。Cheng らは衣服の色情報をヒストグラムとして抽出し照合する Incremental Major Color Spectrum Histogram Representation という衣服認証を提案した¹¹⁾。Hamdoun らは衣服の模様情報を Speed-Upped Robust Feature 特徴より抽出し照合する衣服認証を提案した¹²⁾。従来手法では衣服の外観から色、模様、形状といった局所特徴を抽出して利用している。しかし、これらの特徴はカメラ間の視点変化に伴う照明変動や見えの違い(衣服の模様、形状情報の変動)といった、撮影条件の変動の影響を受けやすい。これに対応するために、撮影条件ごとに局所特徴を使い分けることが考えられるが、適切に使い分けるためには撮影環境を予め知悉する必要があり、汎用性が低くなる。それに対し、複数の局所特徴を統合して利用することも考えられる。前述の通り、大抵らは複数局所特徴の連結統合を行っている³⁾。深度情報による身体方向推定をして、横向きの際にのみ色情報の重みを 1.5 倍、模様情報の重みを 0.5 倍している。ただし、それ以外の向きでは単純連結しているのみであり、撮影条件の変動に対応した適切な重み付けに関する議論はなされていない。

以上をまとめると、顔、歩容、衣服を利用した人物再同定にはそれぞれ長所短所がある。本稿においては衣服からの局所特徴を用いる手法に取り組む。その際、複数の局所特徴を利用するが、撮影条件の変動に対応する適切な重み付けの検討が必要である。

3. 提案手法

第3章では、図2に示す、フレームワークについて詳述する。全体の流れを示す。提案手法は1. 特徴抽出、2. 事前学習、3. 識別の3段階の処理で構成される。それぞれ、3.1節、3.2節、3.3節で詳細を述べる。

3.1 特徴抽出

まず、画像内の人物領域内に設定した上半身を含む矩形領域を特徴抽出のための衣服領域と定義し、衣服の外観情報を抽出する。なお、本稿では上半身を「腰部以上、肩峰以下」と定め、切り出しを行った。下半身は上半身と比較して背景が複雑に混在するために使用しない。

また、前提として、本稿では上半身は既に抽出されているものとする。先行研究である大規らの研究においては、人体の寸法に基づいた上半身の推定・抽出手法が提案されており、これを利用することを考えれば無理のない前提と考える。本研究では、特徴統合手法について焦点を絞って議論を進める。

人物再同定に衣服を利用する場合、主に衣服の色、模様、形状情報が使用されることから、局所特徴としては色に対応して色ヒストグラム特徴、模様に対応して周波数特徴および同時生起行列特徴、形状に対応して輝度勾配ヒストグラム特徴を用いる³⁾。

3.1.1 色ヒストグラム特徴

色ヒストグラム(Histogram (Hist))特徴は、図3に示すとおり、矩形領域内においてHSV表色系に変換後、照明変化に比較的頑健なHue値、Saturation値のヒストグラムを連結したものである。Value値は明るさに対して脆弱であるため使用しない。

3.1.2 周波数特徴

周波数(Frequency (Freq))特徴は、図4に示すとおり、矩形領域内を2次元高速フーリエ変換して得られたパワースペクトル画像の画素値を並べたものとする。

Fig. 3 Extraction of the Color Histograms Features

Fig. 4 Extraction of the Frequency features

Fig. 5 Extraction of the Gray Level Co-occurrence Matrix Features

3.1.3 同時生起行列特徴

同時生起行列(Gray Level Co-occurrence Matrix (GLCM))特徴の抽出手法を図5に示す。同時生起行列 $P_{\delta}(i, j)$ とは、ある画素 i から d 画素離れた位置にある画素 j に関して画像内における画素対 (i, j) の画素値の発生頻度を確率に変換したものである。

まず、矩形領域内をグレースケール化し32階調に変換する。画素間距離 d を1、画素対の角度 $\theta = 0, 45, 90, 135[\text{deg}]$ のときの、 L 行 L 列の正方行列となる同時生起行列 $P_{\delta}(i, j)$ を計4個求める。次に同時生起行列ごとに、以下(1)~(7)式が示す7個の統計値を求める。

エネルギー (Energy)

$$ENE = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \{P_{\delta}(i, j)\}^2 \quad (1)$$

逆差分モーメント (Inverse the Difference Moment)

$$IDM = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{1}{1 + (i - j)^2} P_{\delta}(i, j) \quad (2)$$

コントラスト (Contrast)

$$CNT = \sum_{k=0}^{L-1} k^2 P_{x-y}(k) \quad (3)$$

相関 (Correlation)

$$CRR = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} ij P_{\delta}(i, j)}{\sigma_x \sigma_y} \quad (4)$$

分散 (Variance)

$$VAR = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu_x)^2 P_{\delta}(i, j) \quad (5)$$

エントロピー (Entropy)

$$ENT = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P_{\delta}(i, j) \log\{P_{\delta}(i, j)\} \quad (6)$$

サムエントロピー (Sum Entropy)

$$SEP = - \sum_{k=0}^{2L-2} P_{x+y}(k) \log\{P_{x+y}(k)\} \quad (7)$$

ただし、それぞれのパラメータについては以下の式(8)~(15)のように定義する。

Fig. 6 Extraction of the Histograms of Oriented Gradient features

x 方向の画素値の同時生起行列

$$P_x(i) = \sum_{j=0}^{L-1} P_\delta(i, j) \quad i = 0, 1, \dots, L-1 \quad (8)$$

y 方向の画素値の同時生起行列

$$P_y(j) = \sum_{i=0}^{L-1} P_\delta(i, j) \quad j = 0, 1, \dots, L-1 \quad (9)$$

領域全体の同時生起行列

$$P_{x+y}(k) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P_\delta(i, j) \quad k = 0, 1, \dots, 2L-2 \quad (10)$$

ただし, $|i+j|=k$ を満たす.

$$P_{x-y}(k) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P_\delta(i, j) \quad k = 0, 1, \dots, 2L-2 \quad (11)$$

ただし, $|i-j|=k$ を満たす.

$$\mu_x = \sum_{j=0}^{L-1} j P_x(j) \quad (12)$$

$$\mu_y = \sum_{j=0}^{L-1} j P_y(j) \quad (13)$$

$$\sigma_x^2 = \sum_{j=0}^{L-1} (j - \mu_x)^2 P_x(j) \quad (14)$$

$$\sigma_y^2 = \sum_{j=0}^{L-1} (j - \mu_y)^2 P_y(j) \quad (15)$$

最後に(1)-(7)式が示す 7 個の統計値を L2 ノルムにより正規化し, その結果を同時生起行列特徴とする.

3.1.4 輝度勾配ヒストグラム特徴

輝度勾配ヒストグラム(Histograms of Oriented Gradient (HOG))特徴の抽出方法を図 6 に示す. 矩形領域内をグレースケール画像へ変換し, 8×8 画素の局所領域(セル)ごとに, 輝度勾配の方向と強度を取得して輝度勾配ヒストグラムを生成する. 16×16 画素の局所領域(ブロック)ごとに L2 ノルムにより正規化し, ベクトル化したものとする.

3.2 事前学習

撮影条件の変動から受ける影響を低減するために, 事前学習として局所特徴に付与する統計的な重みを求める. まず, 本稿において, 人物再同定ではカメラごとの撮影条件は大きく変わらないことを前提とする. 例えばカメラ A とカメラ B があつた場合, カメラ A とカメラ B の撮影条件は異なるが, カメラ A で撮影した複数人物の撮影条件はほぼ一様である, カメラ B の場合も同様, であることを想定する.

まずカメラ間の局所特徴ごとの変化量 d^f を求める. 異なるカメラで撮影した同一人物の画像より局所特徴を抽出して距離を

Fig. 7 Step of Pre-training

求め, 変化量 d_i^f とする. ここで, 距離としては人物再同定で利用されることが多い 3 種類, L1 ノルム, L2 ノルムまたはバタチャリヤ距離を考える. 学習人数分の距離の総和が撮影条件における局所特徴の変化量 d^f となる.

この変化量 d^f に反比例するような重み導出関数 $w^f(d^f)$ を適用し, 撮影条件に対し適応する重み w^f を求める. これにより, 撮影条件の変動に脆弱な局所特徴の重みを下げることができる.

重み導出関数としては, Method 1: 変化量 d^f (L1 ノルム)の逆数, Method 2: 変化量 d^f (L2 ノルム)を母集団とした標準正規分布の確率密度関数, Method 3: 変化量 d^f (バタチャリヤ距離)の平均を考案し, 比較する.

ここで, Method 1 は, 最も単純であり, 変動が大きい局所特徴の重みを下げることができる. しかし, これが極端に作用し, 特定の特徴のみが選択される可能性もある. そのような場合を考慮し, 単純連結ではないが, 特定の特徴のみを重視あるいは軽視することなく, 同程度の重み付けをするような手法として Method 2 を考える. さらに, Method 1, 2 では取得特徴の次元数の偏りを考慮していないため, 導出した重みが次元数に依存する可能性がある. そこで取得特徴の次元数の偏りも考慮できるような重み付け手法として Method 3 を考える.

以上をまとめて図 7 に図示する.

3.2.1 Method 1: 変化量 d^f (L1 ノルム)の逆数

まず, 異なるカメラで撮影した同一人物 i における特徴の変化量を求める. 図 7 のように, Cam A が取得した画像を検索人物(Query)画像 Q_i とし, 衣服領域から取得した局所特徴を f^{Q_i} とする. さらに, Cam B が取得した画像を候補人物(Gallery)画像 G_i とし, 衣服領域からした局所特徴を f^{G_i} とする. 対応する Gallery 画像 G_i から抽出した特徴の差分を算出し, L1 ノルムを計算して式(16)の変化量 d_i^f を求める.

$$d_i^f = \|f^{Q_i} - f^{G_i}\|_1 \quad (16)$$

$$f = \{\text{Hist, Freq, GLCM, HOG}\}$$

3.1 節で説明する通り, この変化量 d^f に基づいて重み w^f を導出し, 次に述べる正規化を行う. はじめに, 学習人数 N の分のだけ変化量 d_i^f を算出する. 次に, これらをまとめた $4N$ 次元の列ベクトル $\mathbf{F}$ を L2 ノルムによる正規化を行

い, $\hat{\mathbf{F}} = [\hat{d}_1^{\text{Hist}} \dots \hat{d}_N^{\text{Hist}} \dots \hat{d}_N^{\text{HOG}}]^T$ を求める.

提案手法では、局所特徴の変化量に対して反比例するような重み関数を使用する。そのため個々の変化量 $\hat{d}_i^f$ が極端に小さい場合においては、その平均値を重み関数に適用すると重みが発散する恐れがある。そのため、正規化後の変化量 $\hat{d}_i^f$ の総和 d^f を重み関数に適用することで発散を低減する。Method 1 では、局所特徴ごとの変化量の総和 d^f の逆数を重み $w^f(d^f)$ とする(式(17))。

$$w^f(d^f) = \frac{1}{d^f}$$

$$d^f = \sum_{i=1}^N \hat{d}_i^f \quad (17)$$

3.2.2 Method 2: 変化量 d^f (L2 ノルム)を標準正規分布の確率密度関数

Method 2 では正規分布モデルに当てはめ重み関数を設計する。まず、Method 1 と同様の手順で差分を算出し L2 ノルムを計算して式(18)の変化量 d_i^f を求める。

$$d_i^f = \|f^{Q_i} - f^{G_i}\|_2 \quad (18)$$

次に Method 1 と同様の手順で局所特徴ごとの変化量の総和 d^f を求める。Method 2 では局所特徴ごとの変化量の総和 d^f を母集団とする。まずこの母集団を L2 ノルムにより正規化する。その後、母集団を統一するため正規分布を標準化する。局所特徴ごとに標準化変数 $x^f = (\hat{d}^f - \mu)/\sigma$ を求め、標準正規分布の確率密度関数に代入して式(19)の変化量 d_i^f を求める。

$$w^f(x^f) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^{f^2}}{2}\right) \quad (19)$$

3.2.3 Method 3: 変化量 d^f (バタチャリヤ距離)の平均

まず、異なるカメラで撮影した同一人物 i における局所特徴ごとのバタチャリヤ距離を計算して式(20)の変化量 d_i^f を求める。

$$d_i^f = B(Q_i, G_i) = \sqrt{\sum_{l=1}^k f^{Q_i} f^{G_i}} \quad (20)$$

次に、変化量 d_i^f を各特徴に対して事前学習で用いるデータの数 N だけ求め、局所特徴ごとの変化量の総和 d^f を算出する。最後に、局所特徴ごとの変化量の総和 d^f を学習人数で割り、式(21)の変化量 d_i^f を求める。

$$w^f(d^f) = \frac{1}{N} \cdot d^f \quad (21)$$

3.3 識別

図 2 の Cam A が取得した画像の集合を Query 画像群、Cam B が取得した取得した画像の集合を Gallery 画像群とする。

識別では、同一人物 i ごとに対して Query 画像 Q_i と候補人物 Gallery 画像 G_j の対応付けを行う。

図 8 に識別の流れを示す。

いま図 8 の赤枠の人物画像を Query 画像 Q_i として、これに対応する Gallery 画像 G_i を Gallery 画像群から検索し、対応付けを行う。

まず、Query 画像と各 Gallery 間において局所特徴ごとに計算して、式(22)の類似度 $E(Q_i, G_j)$ 、または式(23)の相違度 $B(Q_i, G_j)$ ($i, j \in J$: 識別人数)を求める。提案手法では撮影条件の変動に対する汎用性を高めるため、学習を必要としない距離

Fig. 8 Step of Identification

指標であるユークリッド距離、バタチャリヤ距離を用いてそれぞれ類似度 $E(Q_i, G_j)$ または相違度 $B(Q_i, G_j)$ として用いる。

$$E(Q_i, G_j) = \sqrt{\sum_{l=1}^k (f^{Q_i} - f^{G_j})^2} \quad (22)$$

$$B(Q_i, G_j) = \sqrt{\sum_{l=1}^k f^{Q_i} f^{G_j}} \quad (23)$$

つぎに、算出した局所特徴ごとの類似度 $E(Q_i, G_j)$ (または相違度 $B(Q_i, G_j)$)と、事前学習で算出した局所特徴ごとの重み w^f との積和を計算し、式(24)の D_i を求める。

$$D_i = \begin{cases} \sum_f E(Q_i, G_j)^f \cdot w^f & (\text{if } w^f = \text{Method 1, 2}) \\ \sum_f B(Q_i, G_j)^f \cdot w^f & (\text{if } w^f = \text{Method 3}) \end{cases} \quad (24)$$

$E(Q_i, G_j)^f$: 特徴 f におけるユークリッド距離
 $B(Q_i, G_j)^f$: 特徴 f におけるバタチャリヤ距離
 w^f : 特徴 f への重み

この積和結果 D_i が Query 画像 Q_i と各 Gallery 画像 G_i との重みづけされた類似度 (また相違度)である。この積和結果 D_i を順位付けし、最も順位が高い場合に相当する Gallery 画像 G_i の人物を Query と同一人物とみなす。

4. 評価実験

提案手法の有効性を評価するために、次に示す 3 種類の比較手法を従来手法として比較実験を行った。

- (1)複数の局所特徴に対し単純連結による特徴統合手法を利用する人物再同定手法
- (2)畳み込みニューラルネットワーク (Convolutional Neural Network)のうち AlexNet¹⁶⁾の第 6 層より抽出した CNN 特徴のみによる人物再同定手法
- (3)局所特徴ごとの総度数を揃え、単純連結による特徴統合手法を利用する人物再同定手法

従来手法について説明する。(1)は最も単純な特徴統合手法である単純連結を用いている。提案手法と比較し、特徴統合における重みづけの有効性について評価する。(2)は深層学習フレームワークである Caffe にて公開されている学習済みの AlexNet より CNN 特徴を抽出している¹⁷⁾。第 6 層に

Table 1 Details of the dataset

Dataset	iLIDS-VID ⁸⁾	GRID ¹³⁾	PRID ¹⁴⁾	ViPeR ¹⁵⁾
Scene	Indoor (Airport)	Indoor (underground station)	Outdoor (Road)	Outdoor (Road)
The number of person	300	248	198	630
The number of data	600	496	396	1260
Illumination variation	○	×	○	×
Appearance variation	×	○	×	○
Data Sample

において CNN 特徴の分類精度が向上することが判明しているため用いた¹⁸⁾。一般物体認識において、認識率が優れている CNN 特徴と比較し提案手法の有効性を評価する。(3)は単純連結による特徴統合を用いた手法であるが、局所特徴ごとの総度数を 1 に揃えている。よって、同じく局所特徴ごとの総度数を揃えている Method 3 と比較し、重みづけの有効性を評価する。

実験対象として表 1 に示す 4 種類の公開データセットを用いた。

提案手法の有効性を評価するため、人物再同定において一般的な指標である累積照合特性(Cumulative Matching Characteristic (CMC))曲線を用いる。CMC 曲線とは Query を Gallery の集合内で上位 n 位以内までに照合できた累積確率を表す曲線である。CMC 曲線は縦軸に再同定精度[%], 横軸に Rank Score をとる。つまり上位 Rank Score における再同定精度が高いほど手法が有効であることを示す。

評価実験の実験結果について説明する。評価実験では、データセットの 1/3 を重み導出のための事前学習として利用し、残り 2/3 を利用して人物再同定を行う。各データセットにおける計 3 回の人物再同定から算出される各 Rank Score での平均の再同定精度を求める。この平均の再同定精度に基づく CMC 曲線を実験結果とする。

4.1 各データセットの詳細

4 種類のデータセットはそれぞれ屋内、または屋外環境におけるデータセットである。さらに各撮影環境において、データセットごとに変化する撮影条件が異なる。評価実験では、提案手法の汎用性についても検証する必要がある。これらを選択した。実験結果の考察を行う前に、データセットごとの性質と重み関数の性質から、予めどのような評価結果になるのかについて予め仮説を立てる。

4.1.1 iLIDS-VID Dataset

iLIDS-VID は空港内の異なる地点に設置された 2 台のカメラが取得した男女合計 300 名のデータより構築される。iLIDS-VID は、カメラの視点が人物に対して横向きに統一されている。よってカメラ間において見えの違いは生じない。しかし iLIDS-VID は屋内環境であるが照明条件が後述する GRID Dataset と比較して大きい。よって本実験では、iLIDS-VID を屋内環境かつ、カメラ間で照明条件が変動する場合の人物再同定と想定する。iLIDS-VID での実験結果から、提案手法の屋内環境における照明変動への適応性を検証する。

4.1.2 GRID Dataset

GRID は地下鉄駅構内の異なる地点に設置された 2 台のカメラが取得した男女合計 250 名のデータより構築される。GRID は、カメラの視点が人物の前向きまたは後向きといっ

ように一定でない。よってカメラ間で見えの違いが生じる。しかし GRID はカメラ間での照明変動が iLIDS-VID と比較して小さい。よって本実験では GRID を屋内環境かつ、カメラ間で見えの違いがある場合の人物再同定を想定する。これにより提案手法の屋内環境における見えの違いへの適応性を検証する。

4.1.3 PRID Dataset

PRID は異なる地点の横断歩道上に設置された 2 台のカメラが取得した男女合計 200 名のデータより構築される。PRID はカメラ間で視点が人物の横向きに統一されている。よってカメラ間において見えの違いは生じない。しかし、PRID は日陰の領域を含むため、カメラ間で照明条件の変化が後述する ViPeR に比べ大きい。よって本実験では、PRID2011 を屋外環境かつ照明条件が変化する場合での人物再同定と想定する。これにより提案手法の屋外環境における照明条件の変化への適応性を検証する。

4.1.4 ViPeR Dataset

ViPeR は、屋外の異なる 2 地点に設置された 2 台のカメラが取得した男女合計 632 名のデータより構築される。ViPeR は、カメラは設置地点が異なるため視点も異なり、カメラ間における見えの違いが PRID と比べ大きい。よってカメラ間で見えの違いが生じる。しかし、ViPeR は屋外環境であるが、カメラ間での照明条件の変動が PRID と比較して小さい。よって本実験では、ViPeR を屋外環境かつ、カメラ間で見えの違いがある場合の人物再同定を想定する。これにより提案手法の屋外環境における見えの違いへの適応性を検証する。

4.2 評価実験の結果に対する仮説

提案手法を評価するために、予め提案手法の変動する撮影条件ごとに対する有効な重み関数と、重み関数同士の汎用性について仮説を立てる。

4.2.1 仮説 1:変動する撮影条件ごとに有効な重み関数に関する仮説

まず、カメラ間で照明条件が変動する撮影環境において有効である可能性がある重み関数について説明する。評価実験では使用するカメラ間で照明変動がある人物再同定を想定したデータセットは、iLIDS-VID と PRID である。

基礎実験では変動する撮影条件に関わらず、色特徴の重要度が高くなる傾向が判明している。よって、照明条件が変動する場合には、有効度が高い色特徴の変化量が増幅してしまう。そのため、全局所特徴に対し、同程度の重みを付加するような Method 2, 3 の重み関数が有効である。これにより、特徴統合をした際に、照明変動より影響を受けた色特徴への再同定精度の依存が低減できると考える。よって、特徴統合が再同定手法の精度向上が最も大きくなると考える。

Fig. 9 CMC curve for each dataset

次に、カメラ間で見えの違いが生じる撮影環境において有効である可能性がある重み関数について説明する。評価実験で使用する、カメラ間で照明変動がある人物再同定を想定したデータセットは、GRID と VIPeR である。

基礎実験では、カメラ間で見えの違いが生じる人物再同定では、模様・形状特徴の重要度が高くなる傾向が判明している。よって、見えが変動する場合において有効度が高い模様・形状特徴の変化量が増幅する。しかし、上述したように見えの違いが生じる環境では、人物再同定において有効度が高い色特徴の有効度が高い。よって、撮影変動に対して有効度の高い特徴のみ大きく重み付けする Method 1 の再同定精度の向上が最も大きくなると考える。

以上に述べた各データセットの性質から提案手法に関して以下のような仮説を立てる。

<仮説 1> カメラ間で照明条件が変化する場合では Method 2,3 に基づく再同定が有効であり、見えの違いが生じる場合では Method 1 に基づく再同定が有効

4.2.2 仮説 2: 変動する撮影条件に対する重み関数の汎用性に関する仮説

さらに提案手法同士の比較として、局所特徴の次元数の違いが与える重み付けへの影響について考察する。Method 1, 2 では局所特徴ごとの変化量 df に対して次元数の違いが考慮されていない。よって重み w^f が特徴の次元数に依存する恐れがある。そこで Method 3 のように次元数の偏りを考慮したうえで特徴統合を行う必要があると考える。次元数の統一手法として、次元数が大きい局所特徴に対し主成分分析 (Principal Component Analysis(PCA))を用いて、次元数の小さい特徴の次元数まで次元圧縮を行う手法が主に使用される。しかし本稿で扱う局所特徴は特徴間で次元数の偏りが大きいため、PCA による次元圧縮では次元を過削減する恐れがある。よって提案手法では特徴ごとの総度数を揃えることで対応する。以上より、次のような仮説を立てる。

<仮説 2> Method 3 の重み付けに基づく再同定手法がすべてのデータセットにおいて最も精度向上率が高い

4.3 実験結果および考察

実験結果として図 9 に各データセットにおける CMC 曲線を示す。図 9 において、Not Weighted は複数の局所特徴に対し単純連結による特徴統合手法を利用した人物再同定手法、Alex は AlexNet の第 6 層より抽出した CNN 特徴のみによる人物再同定手法、L1 は局所特徴ごとの総度数を揃え、単純連結による特徴統合手法を利用した人物再同定手法を示す。さらに、人物再同定で上位 Rank Score の再同定精度の比較によっても手法を評価するため、表 2 に Rank Score = 1, 5, 10, 20 のときの再同定精度を示す。図 10 に比較手法のうち、最も単純な特徴統合である、重みづけを用いない単純連結による特徴統合を基準としたときの各重み付け手法による再同定精度の向上率を示す。なお、ここで述べた向上率とは、Rank Score ごとの提案手法と比較手法の再同定精度の差の平均として表したものである。

4.2 節では評価実験の実験結果全体について評価し、重み付け手法ごとの各局所特徴への重み(図 11)を用いて 4.1 節で立てた仮説の検証を通し考察とする。

図 9、表 2 より全ての実験環境において、3 種類の重み関

Fig.10 Improvement rate by proposed methods for each dataset

Table 2 Compared previous methods and propose methods: “R” is “Rank Score”. “M1” is “Method 1”, “M2” is “Method 2”, “M3” is “Method 3”.

(a) iLIDS-VID					(b) GRID				
	Identification rate %					Identification rate %			
	R=1	R=5	R=10	R=20		R=1	R=5	R=10	R=20
Not weighted	1.5	4.3	7.5	13.7	Not weighted	2.4	8.2	14.1	20.3
weighted by M1	4.2	12.8	21.7	29.0	weighted by M1	13.9	28.5	43.8	56.8
weighted by M2	2.2	7.0	11.8	20.7	weighted by M2	3.6	11.8	18.5	27.9
weighted by M3	6.0	15.3	22.3	31.7	weighted by M3	11.4	29.7	37.6	49.6
Alex	2.6	8.1	13.7	21.1	Alex	8.4	18.9	25.5	41.4
L1	6.2	15.8	23.0	31.3	L1	11.6	24.5	31.3	42.4

(c) PRID					(d) VIPeR				
	Identification rate %					Identification rate %			
	R=1	R=5	R=10	R=20		R=1	R=5	R=10	R=20
Not weighted	3.5	10.9	15.9	26.3	Not weighted	0.7	3.7	5.7	9.5
weighted by M1	5.6	17.9	25.3	36.4	weighted by M1	5.1	11.8	16.9	23.6
weighted by M2	5.1	15.9	25.5	38.6	weighted by M2	1.3	3.7	7.5	14.2
weighted by M3	3.8	14.9	23.0	32.6	weighted by M3	6.1	15.3	20.8	28.8
Alex	3.0	8.1	16.7	31.6	Alex	3.1	5.7	8.4	11.7
L1	3.8	10.1	15.7	24.2	L1	6.7	15.0	20.2	27.2

Fig.11 The weight of each local features of each dataset: “Hist” is the color histograms features. “GLCM” is the Gray-Level Co-occurrence Matrix features. “Freq” is the frequency features. “HOG” is the Histograms of Oriented Gradients

数を用いたいずれかを用いた提案手法の再同定精度が、従来手法を上回った。これにより、人物再同定において重みづけ特徴統合手法である提案手法の有効性が示唆された。

評価実験では一般物体認識において有効とされるCNN特徴を用いた再同定精度との比較も行った。図9より全実験環境において再同定精度の向上を確認した。深層学習手法は画素値の畳み込み演算であるため、出力される特徴は画像の色成分に依存している可能性があると考えられる。よって図9(a), (c)のようにカメラ間で照明条件に変動がある場合に再同定精度が低下したと考える。CNN特徴による再同定精度との比較により、従来では画像潜在情報も抽出可能であるCNN特徴による認識手法が、既存のhand-craftな局所特徴による認識手法より優れているものとされている。しかし、提案手法のように複数の局所特徴を用い、かつ適切な重み付けと特徴統合によって認識精度を上回ることが可能であることが示唆された。

次に4.1節で立てた仮説について検証する。

<仮説検証 1> 照明条件が変化する場合ではMethod 2, 3に基づく再同定が有効であり、見えの違いが生じる場合ではMethod 1に基づく再同定が有効

照明変動に対する提案手法の効果について考察する。図10よりカメラ間で照明条件が変化するデータセットであるiLIDS-VIDでは、仮説と反しMethod 1による重み付けに基づく手法の精度向上率が最大であった。

PRIDにおいてMethod 1では図11(a)より同時生起特徴に対して最も重みがかけており、Method 3では図11(c)より周波数特徴に対して最も重みがかけている。PRIDは逆光により模様取得できない場合があった(図12)。

Fig.12 Example of appearance difference by Backlight in PRID

よって色特徴の重要度が模様情報の重要度を上回ったため、Method 1の重み付けに基づく再同定手法が有効となるような評価結果になったと推測する。

見えの違いに対する提案手法の効果について考察する。図10よりカメラ間で見えの違いが生じるデータセットであるVIPeRでは、仮説と反しMethod 3による重み付けに基づく手法の精度向上率が最大であった。VIPeRは他のデータセットと比較し、空港、地下鉄構内、道路上といったようにカメラの設置位置が固定されていない。よってデータセット間で人物ごとに照明条件も変動しているとみなされたため、照明条件の変動において有効なMethod 3の重み付けに基づく再同定手法が有効となるような評価結果になったことが示唆された。

<仮説検証 2> Method 3の重み付けに基づく再同定手法がすべてのデータセットにおいて最も精度向上率が高い

表3に提案手法ごとの再同定精度の向上率の平均、および分散を示す。

表3よりMethod 3が3種類の重み付け手法のうち再同定精度の向上率の平均がMethod 1より低い、分散がMethod 1のおよそ半分である。よって本実験で評価した重み導出関数の中で撮影条件に対し再同定精度を向上し、最も安定した重み付け手法であることがわかる。Method 3では変化量

Table 3 The mean and variance of improvement rate

	Method 1	Method 2	Method 3
Mean	12.7	6.2	12.3
Variance	44.8	0.9	24.3

d^f としてバタチャリヤ距離を用いた。そのため、局所特徴ごとの総度数を1に揃えた。ここで総度数を揃えることで変化量 d^f の値に範囲が発生する。よって、特徴間の変化量のスケールを次元数に関係なく評価できたと考える。

評価実験の結果から Method 3 に基づく重み付け手法による特徴統合が、撮影条件の変動に対して汎用性の高い人物再同定の実現のために有効であることが示唆された。

5. 結 言

衣服を利用することを前提とする人物再同定の問題点である照明変動や衣服の模様・形状などの見えの違いといった、撮影条件の変動に対応する汎用的な特徴統合手法を提案した。

複数の局所特徴を用いた照合の際、撮影条件の変動を局所特徴(色ヒストグラム特徴、周波数特徴、同時生起行列特徴、HOG 特徴)の差分によって表現し、これに基づく重み付け手法の選定した。重み付けには、Method 1: 変化量(差分のL1 ノルム)の総和の逆数、Method 2: 変化量(差分のL2 ノルム)を母集団とした確率密度関数、Method 3: 変化量(バタチャリヤ距離)の平均の3種類の関数を採用し、比較した。本研究の貢献は以下の3点である。

- (1) 撮影環境が異なる4種類の公開データセットにおいて提案手法による人物再同定を行った。複数の局所特徴を単純連結する特徴統合手法に比べ、再同定精度が平均 8.8 [%](iLIDS-VID), 16 [%](GRID), 6.5 [%](PRID), 9.9 [%](VIPeR) 向上した。よって、提案手法が複数環境に対して汎用性を持つことを確認した。
- (2) 実験で利用したデータセットの特性に基づき、撮影環境毎に有効な重み関数について検討した。カメラ間で照明条件が変動する場合、複数局所特徴に対し同程度の重みを与える Method 2, Method 3 が有効であり、見えが変動する場合、変動する撮影条件に有効な局所特徴のみに重みを与える Method 1 が有効であった。
- (3) 採用した重み関数の性質に基づき、撮影条件の変動に対応する汎用性について検討した。Method ごとの再同定精度の向上率の平均と分散を比較した。Method 3 が精度向上率に対する分散が小さかったため、撮影条件の変動に対して最も汎用性が高い重み関数であることが示唆

された。

仮説検証より変動する撮影条件ごとに有効な重み関数の性質が示唆された。そこで今後は、変動する撮影条件を予め推定し、適切な重み関数による重みづけ特徴統合を行う予定である。

謝 辞

本研究の一部は、JSPS 科研費 15K01482 の援助を受けた。

参 考 文 献

- 1) Apurva Bedagkar-Gala et al.: A Survey of Approaches and Trends in Person Re-identification, *Journal of Image and Vision Computing*, **32**, 4 (2011), p.270.
- 2) R. Vezzani et al.: People Reidentification in Surveillance and Forensics: A Survey, *Journal of ACM Computing Surveys (CSUR)*, **46**, 2 (2013), p.1.
- 3) 大槻 雅文: 衣服特徴を用いた人物検索システム, 電気学会論文誌 D, **134**, 3 (2014), p.350.
- 4) Paul Viola et al.: Robust Real-time Face Detection, *International Journal of Computer Vision*, (2004), **57**, 2, p.137.
- 5) Haoxiang Li et al.: Hierarchical-PEP Model for Real-world Face Recognition, *IEEE Conference on Computer Vision and Pattern Recognition*, (2015), p.4055.
- 6) Tal Hassner et al.: Effective Face Frontalization in Unconstrained Images, *IEEE Conference on Computer Vision and Pattern Recognition*, (2015), p.4295.
- 7) 横原 靖: 周波数領域における方向変換モデルを用いた歩容認証, 情報処理学会論文誌, **48**, (2007), p.117.
- 8) Taiqing Wang et al.: Person Re-Identification by Video Ranking, *European Conference on Computer Vision*, (2014), p.1.
- 9) 今成 隆了: Dense Trajectories を用いた歩行人物再同定に関する検討, 日本ロボット学会学術講演会, IH3-04, (2015), p.1.
- 10) 川合 諒: 方向変換モデルを用いた複数カメラ間での人物照合, 電子情報通信学会 信学技報, **185**, 11 (2011), p.1.
- 11) E.D.Cheng et al.: Matching of Objects Moving Across Disjoint Cameras, *International Conference on Image Processing*, (2006), p.1769.
- 12) Omar.Hamdoun et al.: Person Re-identification in Multi-camera System by Signature Based on Interest Point Descriptors Collected on Short Video Sequences, *Distributed Smart Cameras*, (2008), p.1.
- 13) Chunxiao Liu et al.: On-the-fly Feature Importance Mining for Person Re-Identification, *Journal of Pattern Recognition*, (2014), **47**, p.1602.
- 14) Martin Hirzer et al.: Person Re-Identification by Descriptive and Discriminative Classification, *In Proc. Scandinavian Conference on Image Analysis*, (2011), **6688**, p.91.
- 15) Doug Gray et al.: Evaluating Appearance Models for Recognition, Reacquisition, and Tracking, *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, (2007), p.1.
- 16) Alex Krizhevsky et al.: ImageNet Classification with Deep Convolutional Neural Networks, *Conference on Neural Information Processing Systems*, (2012), p.1.
- 17) Caffe Model Zoo, "http://caffe.berkeleyvision.org/model_zoo.html", アクセス日時: 2015 年 12 月 15 日 18:00.
- 18) Jeff Donahue et al.: DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, *International Conference on Machine Learning*, (2014), p.647.