

# 低解像度に劣化した画像を用いた人物対応付けにおける ダウンサンプリング処理の効果検証

遠藤 拓弥<sup>†</sup>      西山 正志<sup>†,††a)</sup>      岩井 儀雄<sup>†,††</sup>

Investigation of downsampling technique for person re-identification using low resolution images

Takuya ENDO<sup>†</sup>, Masashi NISHIYAMA<sup>†,††a)</sup>, and Yoshio IWAI<sup>†,††</sup>

あらまし 本論文では、低解像度の人物画像に合わせてダウンサンプリングを施す前処理を用いることで、人物対応付けの精度が改善されるかどうかを検証する。解像度変動の影響を受けにくい特徴量を抽出するために、学習ベースのアプローチを適用することが一般的である。ただし、同一人物について高解像度画像と低解像度画像のペアを大量に準備しなければならない課題がある。本論文では、高解像度画像と低解像度画像のペアが収集できない状況を想定し、特徴抽出の前処理としてダウンサンプリングを用いることで、低解像度画像でも安定に人物を対応付けることを狙う。カメラレンズの焦点距離、および、人物からカメラまでの距離を変化させた場合のデータセットを新たに構築し、低解像度画像に合わせるダウンサンプリングが、高解像度画像に合わせるアップサンプリングと比べて、人物対応付けの精度改善に大きく寄与することを確認した。また、ダウンサンプリングで得られる低周波成分に、個人性を表す特徴が多く含まれることをシミュレーション実験で確認した。

キーワード 人物対応付け, 低解像度画像, ダウンサンプリング

## 1. はじめに

安心かつ安全な社会に向けて、様々な場所に設置された防犯カメラを用いて、人物を対応付けるシステムが求められている。このシステムにより、迷子の捜索や、不審者の経路予測を行うことが期待されている。人物対応付けシステムを実現する上で重要となるポイントは、映像中の人物画像から、その人物の個人性を表す特徴量を如何に抽出するかである。このために様々な既存手法 [1]~[5] が提案されている。特徴量を安定に抽出するため、高解像度で撮影された人物画像を用いることが望ましい。高解像度画像には、その人物を表す個人性が多く含まれており、対応付けを正しく行える可能性は一般的に高い。しかし、防犯カメラから獲得される人物画像の解像度は、常に一定とは限

らず、カメラレンズの性能や、カメラが設置された位置により大きく変化する。低解像度で画像が撮影された場合、その人物を表す個人性を抽出することが難しくなり、対応付けの精度が低下していく。

低解像度の人物画像が引き起こす問題に対して、解像度変動を考慮した人物対応付けの既存手法 [6]~[10] が提案されている。これらの既存手法は、同一人物の高解像度画像と低解像度画像のペアを訓練サンプルとする学習ベースのアプローチを用いている。解像度変動を学習することで、対応付け精度の維持向上を図っている。ただし、学習ベースのアプローチを適用するためには、高解像度画像と低解像度画像のペアを大量に準備する必要があり、それらの収集に手間が非常にかかる課題がある。

そこで本論文では、大量の訓練サンプルを必要とする学習ベースのアプローチではなく、人物画像そのものにリサンプリング処理を施すアプローチを採用することで、人物対応付けの精度が改善されるかどうかを検証する。高解像度画像と低解像度画像のペアからなる訓練サンプルが収集できない状況でも、リサンプリ

<sup>†</sup> 鳥取大学大学院持続性社会創生科学研究科, 鳥取市  
Graduate School of Sustainability Science, Tottori University, Tottori 680-8550, Japan

<sup>††</sup> 鳥取大学大学院工学研究科, 鳥取市  
Graduate School of Engineering, Tottori University

a) E-mail: nishiyama@tottori-u.ac.jp

ング処理は単純に適用できる利点がある。リサンプリングの具体的な処理として、高解像度画像に合わせるアップサンプリング、および、低解像度画像に合わせるダウンサンプリングが考えられる。本論文では、どちらのアプローチが精度改善に適しているかを検証する。このために、カメラレンズと設置位置を変化させた場合の評価データセットを新たに構築した。実験結果より、低解像度画像に合わせるダウンサンプリングが、高解像度画像に合わせるアップサンプリングと比べて、人物対応付けの精度改善に大きく寄与することを確認した。以下では、2. で関連研究について述べ、3. でダウンサンプリングが有効であることを実験的に示し、4. と 5. にてその理由を周波数解析を用いて調査する。最後に 6. でまとめる。

## 2. 関連研究

解像度変動を考慮した人物対応付けの既存手法について述べる前に、目的が関連している顔認識の分野における解像度のサーベイ論文 [11] を挙げる。この文献では、低解像度の課題に対応するために二つのアプローチが存在すると述べられている。近年では、これらのアプローチに基づき様々な手法が提案されている。第 1 アプローチ [12], [13] では、解像度の影響を受けにくい特徴空間を設計し、その特徴空間において認識処理を行っている。第 2 アプローチ [14], [15] では、低解像度の画像を高解像度に変換することで、高解像度の画像同士で認識処理を行っている。これらのアプローチは顔認識において大きな成果を収めているが、人物対応付けでも同様のアプローチを採用した既存手法がいくつか提案されている。

人物対応付けのための第 1 アプローチとして、解像度の影響を受けにくい特徴空間を設計するために、既存手法 [6]~[8] が提案されている。同一人物の高解像度画像と低解像度画像のペアを訓練サンプルとし、解像度変化に影響を受けにくい特徴量を抽出するための統計的な学習手法が採用されている。第 2 アプローチとして低解像度の画像を高解像度に変換するために、既存手法 [9], [10] が提案されている。第 1 アプローチと同様に、同一人物の解像度ペアを訓練サンプルとし、統計的な学習を組み込んだ超解像が採用されている。

どちらのアプローチにおいても、解像度ペアの訓練サンプルを大量に準備することができれば、人物対応付けの精度を大幅に高められると期待できる。ただし、解像度ペアの訓練サンプルを大量に収集すること

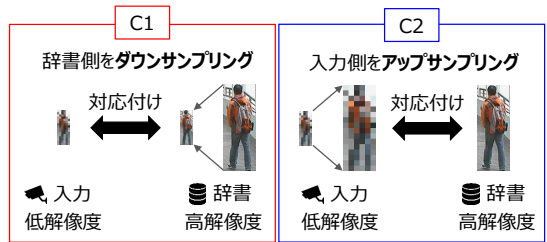


図 1 ダウンサンプリングを用いる手法 (C1) とアップサンプリングを用いる手法 (C2) の概念図。

Fig. 1 Overview of a downsampling method (C1) and a upsampling method (C2).

は実応用において非常に手間がかかる。本論文では、訓練サンプルが必要となる学習ベースのアプローチではなく、前処理であるリサンプリングを人物画像に施すアプローチを採用することで、対応付け精度の改善を狙う。

## 3. 人物対応付けにおけるリサンプリングの効果検証

### 3.1 比較するリサンプリング手法

人物対応付けシステムでは、様々な解像度の人物画像が、辞書側と入力側に与えられる可能性がある。本論文では問題設定を単純にするため、辞書側には高解像度の人物画像が与えられ、入力側には低解像度の人物画像が与えられる条件とする。辞書側と入力側との間で人物画像の対応付けはまだ完了していないため、同一人物であるかどうかは定まっていない状況である。本論文では、以下の二つのリサンプリング処理について、対応付け精度の比較を行う。

<C1> 辞書側の高解像度画像を、入力側の低い解像度に合わせてダウンサンプリングする。その後、低解像度になった辞書画像と、元の低解像度の入力画像との間で対応付けを行う。

<C2> 入力側の低解像度画像を、辞書側の高い解像度に合わせてアップサンプリングする。その後、高解像度になった入力画像と、元の高解像度の辞書画像との間で対応付けを行う。

比較手法 C1 および C2 の概要を図 1 に示す。本論文では、C1 のダウンサンプリングと C2 のアップサンプリングの具体的な処理として、ともにバイリニア補間を適用する。さらに C2 では、Super-resolution convolutional neural network (SRCNN) [16] を適用した場合も評価する。

### 3.2 人物画像の解像度を低くする要因の考察

リサンプリング処理を用いた場合の対応付け精度を検証するにあたり、辞書側と入力側に与える人物画像のデータセット構築が必要となる。データセット構築の指針を決めるために、実環境で解像度が低下する要因を考える。代表的な要因として、カメラレンズが変化する場合、および、カメラ位置が変化する場合が挙げられる。以下、それぞれについて詳細を述べる。

カメラレンズが解像度に与える影響を考える。カメラの性能は目覚ましく進歩しており、大多数のカメラセンサは大量の画素数をもつ。このため、撮影された画像自体は高解像度となることが一般的である。ただし、防犯カメラは一台のカメラで広い視野をカバーしており、一人あたりの人物画像の解像度は低くなることが多い。よって、カメラセンサの画素数が多いとしても、カメラレンズの焦点距離が異なる場合には、人物画像の解像度は低下する。具体的には、焦点距離の短い広角レンズを装着している場合に人物画像の解像度は低くなる。一方、望遠レンズを装着している場合には人物画像の解像度が高くなる。

次にカメラ位置が解像度に与える影響を考える。防犯カメラが設置される位置は、人物を見守る状況や目的に応じて様々に変化する。例えば、車両の入り口のように特定箇所を見守る場合、カメラは人物に近い位置に設置される。また、空港ロビーのような広い空間を見守る場合、カメラは見晴らしの良い離れた場所に設置される。カメラに装着されたレンズが同一だとすると、カメラから人物までの距離が近い場合は高解像度になり、距離が遠い場合は低解像度になる。

以下では、カメラレンズの焦点距離が変化した場合、また、人物からカメラまでの距離が変化した場合について、人物対応付けの評価データセットを構築していく。

### 3.3 評価データセットの構築

#### 3.3.1 レンズの焦点変化

カメラレンズの焦点距離を変化させた場合について、人物画像の撮影概要を図2に示す。入力側の人物画像を撮影する際、レンズの焦点距離を135ミリメートル、70ミリメートル、35ミリメートル、18ミリメートルに変化させた。撮影された人物画像の平均サイズは、それぞれの焦点距離で(58.3, 186.3), (31.3, 100.5), (15.4, 49.5), (7.2, 24.2)画素であった。辞書側の人物画像として、別の時刻に獲得した焦点距離135ミリメートルの撮影条件を用いた。人物からカ

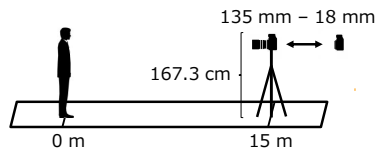


図2 カメラレンズの焦点距離を変化させた場合の撮影概要。

Fig.2 Camera setting when changing the focal length of the camera lens.

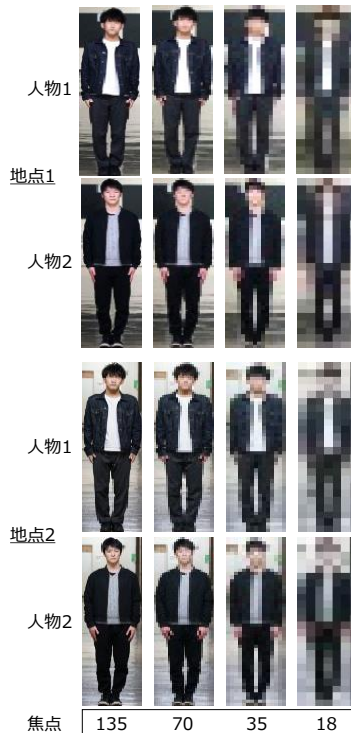


図3 カメラレンズの焦点距離を変化させた場合の人物画像の例。

Fig.3 Examples of person images when changing the focal length of the camera lens.

メラまでの距離は15メートルと固定した。カメラを床から167.3センチメートルの位置に配置した。カメラはSony α6300を、レンズはE 18-135mm F3.5-5.6 OSSを用いた。外光の影響を受けない屋内の異なる5地点で撮影を行った。撮影人数は20名(男性17名、女性3名)とし、撮影時には直立姿勢をとるよう指示した。人物の外接矩形を手で教えることで人物画像を映像から抽出した。本データセットにおける人物画像の例を図3に示す。図中から、焦点距離が短くなるにつれ人物画像の解像度が低くなるのが分かる。

### 3.3.2 カメラ位置までの距離変化

人物からカメラまでの距離を変化させた場合について、人物画像の撮影概要を図4に示す。入力側の人物画像を撮影する際、人物からカメラまでの距離を5メートル、10メートル、15メートル、20メートル、25メートルに変化させた。撮影された人物画像の平均サイズは、それぞれの距離で(59.2, 192.9), (29.2, 94.7), (19.6, 62.4), (14.3, 46.7), (11.7, 37.1)画素であった。辞書側の人物画像として、別の時刻に獲得した距離5メートルの撮影条件を用いた。人物に当たる光量を一定にするため、人物の立ち位置を固定し、カメラ位置を移動させた。カメラとして logicool HD Pro Webcam C920r を用いた。外光の影響を受けない屋内の異なる5地点で撮影を行った。撮影人数は20名(男性17名, 女性3名)とした。本データセットにおける人物画像の例を図5に示す。図中から、人物からカメラまでの距離が遠くなるにつれ人物画像の解像度が低くなること分かる。

### 3.4 人物対応付けの精度

#### 3.4.1 レンズの焦点変化

カメラレンズの焦点距離を変化させたデータセットにおいて、人物対応付けの精度を評価した。人物画像から特徴量を算出する前段階の処理として、3.1で述べたダウンサンプリング(C1)とアップサンプリング(C2)を人物画像に適用した。特徴量を抽出するために、性別や服装など人物属性の共起性を求める手法[17]を用いた。対応付け精度を高めるために、計量学習の一種である Large margin nearest neighbor (LMNN)[18]を適用した。計量学習の訓練サンプルとして、テストに用いるデータセットとは別である CUHK01 データセット[5]を用いた。辞書側の特徴量と入力側の特徴量との対応付けに、ユークリッド距離に基づく最近傍法を用いた。評価指標として、入力側に与えられた人物が、辞書側の人物の中から正しく対応付けられたことを測る第1位正解率を用いた。焦点変化データセットに含まれる5地点において、それぞれの地点で対応付けを行い、平均第1位正解率を算出した。

ダウンサンプリング(C1)とアップサンプリング(C2)を適用した場合の人物対応付けの平均第1位正解率を図6に示す。図中よりC2では、バイリニア補間の精度と比べてSRCNNの精度が良いことが分かる。辞書側の高い解像度に合わせてアップサンプリングするC2より、入力側の低い解像度に合わせてダウンサンプリングするC1が、人物対応付けの精度を改

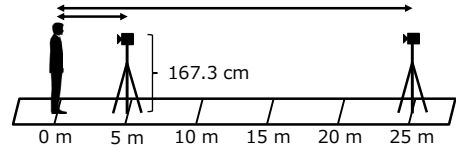


図4 人物からカメラまでの距離を変化させた場合の撮影概要。

Fig. 4 Camera setting when changing the distance from the camera to the person.

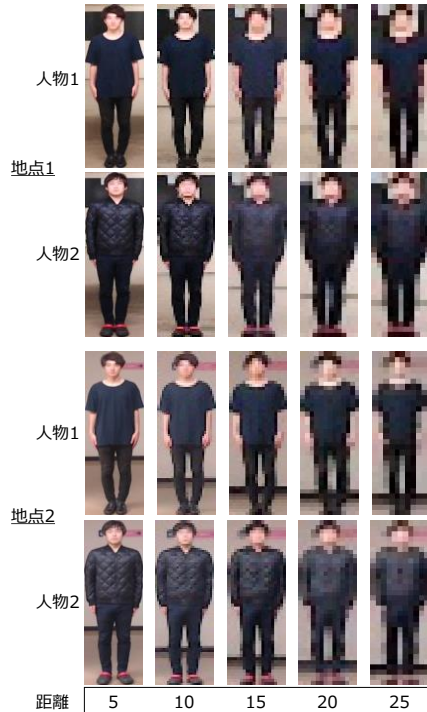


図5 人物からカメラまでの距離を変化させた場合の人物画像の例。

Fig. 5 Examples of person images when changing the distance from the camera to the person.

善させることが分かる。カメラレンズの焦点距離により解像度が低下した場合、アップサンプリングではなくダウンサンプリングが、人物対応付けの前処理として適していると考えられる。

#### 3.4.2 カメラ位置までの距離変化

人物からカメラまでの距離を変化させたデータセットにおいて、人物対応付けの精度を評価した。データセット以外の実験条件は3.4.1と同じとした。3.1で述べたC1とC2を用いた場合の人物対応付けの平均第1位正解率を図7に示す。図中よりC2では、バイリニア補間の精度と比べてSRCNNの精度が良いこと

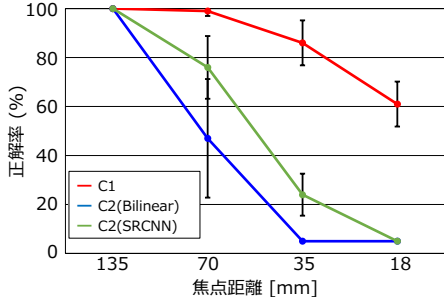


図 6 カメラレンズの焦点距離を変化させた場合の人物対応付けの精度。

Fig. 6 Accuracy of person re-identification when changing the focal length of the camera lens.

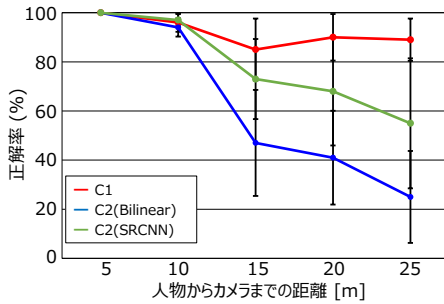


図 7 人物からカメラまでの距離を変化させた場合の人物対応付けの精度。

Fig. 7 Accuracy of person re-identification when changing the distance from the camera.

が分かる。この実験においても 3.4.1 と同様に、アップサンプリングする C2 より、ダウンサンプリングする C1 が、人物対応付けの精度を改善させることが分かる。人物からカメラまでの距離により解像度が低下した場合、アップサンプリングではなくダウンサンプリングが、人物対応付けの前処理として適していると考えられる。

#### 4. ダウンサンプリングで対応付け精度が改善される理由の調査

##### 4.1 方針

実際のカメラで撮影された低解像度の人物画像を用いた 3.4 の実験により、アップサンプリングよりダウンサンプリングの前処理が、人物対応付けの精度を改善することが明らかになった。以下では、空間方向の周波数解析を用いて、ダウンサンプリングで精度が高まる理由を実験的に調査する。本論文では、高解像度画像は低周波成分と高周波成分に分離することができ、

その低周波成分が低解像度画像に相当すると仮定を置く。人物対応付けでは、辞書側の特徴量と入力側の特徴量の距離計算を行い、距離が近ければ同一人物としている。以下では、特徴量は人物画像の画素値そのものとし議論を進める。辞書側と入力側が互いに高解像度である場合、人物画像同士の距離  $d^2$  は、先ほど置いた仮定に従い以下の式 (1) で表すことができる。

$$d^2 = d_L^2 + d_H^2 \quad (1)$$

ここで、 $d_L^2$  は低周波成分のみで対応付けた場合の距離、 $d_H^2$  は高周波成分のみで対応付けた場合の距離を表す。本実験では、それぞれの距離を用いた場合で、対応付け精度がどのように変化するかをシミュレーションで評価する。以下に比較する手法を挙げる。

(F1) 全帯域の周波数成分を用いて距離  $d^2$  を算出。

(F2) 低周波帯の成分のみを用いて距離  $d_L^2$  を算出。

(F3) 高周波帯の成分のみを用いて距離  $d_H^2$  を算出。

本実験では辞書側を高解像度画像とする。入力側の低解像度画像をアップサンプリングし、疑似的に高解像度にした画像を対象とする。F1 は 3.1 のアップサンプリングを用いる C2 に対応し、F2 はダウンサンプリングを用いる C1 に対応すると想定できる。次節では  $d_L^2$  と  $d_H^2$  の計算処理の詳細について述べる。

##### 4.2 周波数領域における距離計算

周波数領域における距離計算の詳細を以下で述べる。本節には新規性は無いが、調査方法の確認のために計算処理の詳細を説明する。辞書画像を  $f^t$ 、アップサンプリングされた低解像度の入力画像を  $g^{q\uparrow}$  とする。この時、辞書画像と入力画像のサイズは  $(N_w, N_h)$  画素で統一されているとする。それぞれの画像が、正規直交関数  $u_{ij}$  で直交関数展開されると、

$$f^t = \sum_i \sum_j^{N_h} c_{ij}^t u_{ij} \quad (2)$$

$$g^{q\uparrow} = \sum_i \sum_j^{N_h} c_{ij}^q u_{ij} \quad (3)$$

と表すことができる。ここで、 $c_{ij}^t$  は辞書画像の係数、 $c_{ij}^q$  は入力画像の係数とする。なお本論文では、正規直交関数  $u_{ij}$  における直交関数展開として、離散フーリエ変換を用いる。次に、画像間で差分をとる式 (4) の関数を考える。

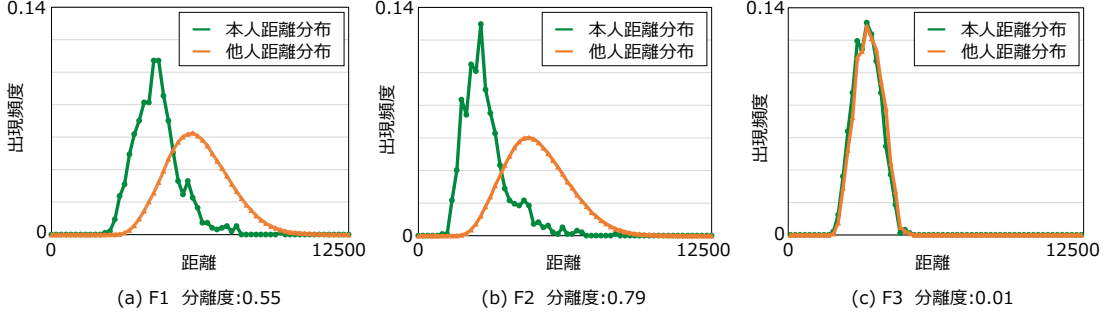


図8 手法 F1 から F3 を用いた場合の本人距離分布と他人距離分布。  
Fig. 8 Distribution of the distances between the same persons and the distribution of the distances between other persons when applying F1 to F3.

$$h(\mathbf{f}^t, \mathbf{g}^{q\uparrow}) = \mathbf{f}^t - \mathbf{g}^{q\uparrow} = \sum_i^{N_w} \sum_j^{N_h} (c_{ij}^t - c_{ij}^q) \mathbf{u}_{ij} \quad (4)$$

対応付けのための最近傍法において、辞書画像と入力画像との間でユークリッド距離を計算することは、関数  $h(\mathbf{f}^t, \mathbf{g}^{q\uparrow})$  のノルムの二乗を計算することと等価である。よって距離  $d^2$  は

$$d^2 = \|h(\mathbf{f}^t, \mathbf{g}^{q\uparrow})\|^2 \quad (5)$$

と表される。正規直交関数系が完備であれば、パーセバルの等式が成立し、

$$d^2 = \sum_i^{N_w} \sum_j^{N_h} (c_{ij}^t - c_{ij}^q)^2 \quad (6)$$

となる。ある次数  $M_w$  (ただし  $M_w < N_w$ ) と  $M_h$  (ただし  $M_h < N_h$ ) を用いて、低周波成分と高周波成分に分離すると、それぞれの距離は

$$d_L^2 = \sum_i^{M_w} \sum_j^{M_h} (c_{ij}^t - c_{ij}^q)^2 \quad (7)$$

$$d_H^2 = \sum_{M_w+1}^{N_w} \sum_{M_h+1}^{N_h} (c_{ij}^t - c_{ij}^q)^2 \quad (8)$$

と表される。上記で算出された距離を用いることで、低周波成分と高周波成分が人物対応付けの精度に与える影響を考察する。

### 4.3 実験条件

画像が低解像度に劣化するモデルを設定しシミュレーション実験を行った。入力側で低解像度の人物画

像  $\mathbf{g}^q$  が獲得されるモデルを式 (9) で表す。

$$\mathbf{g}^q = \mathbf{B}\mathbf{H}\mathbf{f}^q + \mathbf{n} \quad (9)$$

ここで、 $\mathbf{B}$  はダウンサンプリング、 $\mathbf{H}$  はぼけ、 $\mathbf{f}^q$  は元の高解像度の入力画像、 $\mathbf{n}$  はノイズを表す。 $\mathbf{B}$ 、 $\mathbf{H}$  はシフトインバリエントと仮定した。具体的には  $\mathbf{B}$  をバイリニア補間、 $\mathbf{H}$  をガウスぼけ、 $\mathbf{n}$  を白色ガウスノイズとした。なお式 (9) では、 $\mathbf{g}^q \in \mathbf{B}^{M_w \times M_h}$  から  $\mathbf{g}^q \in \mathbf{B}^M$  に変換されているとし、 $\mathbf{f}^q \in \mathbf{B}^{N_w \times N_h}$  から  $\mathbf{f}^q \in \mathbf{B}^N$  に変換されているとする。

本実験では CUHK01 データセット [5] を用いた。このデータセットには 971 人の人物画像が含まれており、異なる 2 つの時刻で撮影されたものを用いた。辞書側と入力側に、それぞれの時刻の人物画像を割り振り、辞書画像を 971 枚、入力画像を 971 枚で評価を行った。このデータセットの人物画像のサイズは (60, 160) 画素で統一されていたため、 $N_w = 60$ 、 $N_h = 160$  とした。本実験では、入力側の人物画像  $\mathbf{g}^q$  を式 (9) を用いて生成した。 $\mathbf{g}^q$  のサイズは (6, 16) 画素とした。低周波成分と高周波成分に分離するための次数を  $M_w = 6$ 、 $M_h = 16$  とした。この低解像度画像を (60, 160) 画素までバイリニア補間でアップサンプリングすることで、入力画像  $\mathbf{g}^{q\uparrow}$  を生成した。

### 4.4 実験結果

辞書画像と入力画像との間で、4.1 で述べた F1 から F3 の手法を用いて距離を算出した。それぞれの手法で得られた距離の分布を図 8 に示す。図中の緑線は辞書画像と入力画像が同一人物である時の距離分布 (本人距離分布) を示し、橙線は異なる人物である時の距離分布 (他人距離分布) を示す。本人距離分布と他人距離分布との重なりが少ないほど、人物を正しく対応

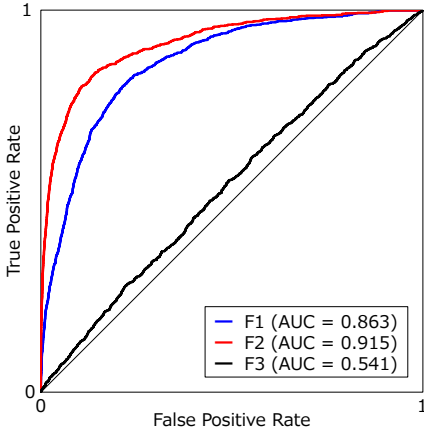


図9 手法 F1 から F3 を用いた場合の ROC カーブ。  
Fig. 9 ROC curves when applying F1 to F3.

付けられる可能性が高いと言える。距離分布の重なりを測るために分離度を用いた。図 8 より、全周波数の成分を含む F1 よりも、低周波成分のみを含む F2 が、分離度が高いことが分かった。また、高周波成分のみを含む F3 では、分離度が非常に低いことが分かった。アップサンプリングにより付加された高周波成分には、対応付けに必要な人物間での違いが、ほぼ含まれていないと言える。

手法 F1 から F3 を用いた場合の ROC 曲線を図 9 に示す。図中の青線は F1、赤線は F2、黒線は F3 を表す。ROC 曲線を描くため、scikit-learn に実装されている関数 `roc_curve` を用いた。全周波数の成分を含む F1、および、高周波成分のみを含む F3 と比較して、低周波成分のみを含む F2 が、人物対応付けの精度を高めることが分かった。実験結果より、低周波成分には個人性を表す重要な特徴が多く含まれていると言える。一方、アップサンプリングで疑似的に生成された高周波成分には、個人性がほぼ含まれておらず、むしろ対応付けに悪影響を及ぼす成分が誤って付加されたと言える。手法 F2 は 3.1 の C1 に対応することから、低解像度に合わせてダウンサンプリングすることで、人物対応付けに有効な低周波成分の特徴を残しつつ、高周波成分の特徴を除去したため、対応付け精度が高くなったと考えられる。

#### 4.5 バイリニア以外の補間を用いた場合の検証

ここでは 4.4 のバイリニア補間以外のアップサンプリング手法を用いた場合について検証した。比較手法として SRCNN [16] を用いた。補間手法以外の実験

表 1 異なる補間方法を用いた場合の第 1 位正解率と AUC の比較。

Table 1 First matching rate (%) and AUC when applying different interpolation techniques.

	第 1 位正解率			AUC		
	F1	F2	F3	F1	F2	F3
Bilinear	9.1	<b>35.0</b>	0.0	0.863	<b>0.915</b>	0.541
SRCNN	18.0	<b>44.1</b>	0.0	0.875	<b>0.917</b>	0.549

条件は 4.4 と同じとした。評価指標として、第 1 位正解率、および、ROC 曲線の Area under the curve (AUC) を用いた。AUC は 1.0 に近いほど人物対応付けの精度が高いことを表す。

異なるアップサンプリング手法を用いた場合の人物対応付けの精度を表 1 に示す。どの補間方法を用いたとしても、全周波数の成分を含む F1、および、高周波成分のみを含む F3 と比較して、低周波成分のみを含む F2 が、第 1 位正解率が高く、かつ、AUC が高いことが分かった。先の 4.4 の結果と同様に、対応付けのための弁別性が高い特徴は低周波成分に含まれていると言える。F2 は 3.1 の C1 に対応することから、アップサンプリングに対するダウンサンプリングの有効性が本実験でも確認された。この実験では解像度を (6, 16) 画素としたが、(30, 80), (24, 64), (18, 48), (12, 32) 画素とした場合でも、F1 と F3 の精度に比べて、F2 の精度が改善される傾向が同様に見られた。

既存手法 [6]~[10] など学習ベースの既存手法の精度と比較した場合、表 1 の提案手法の精度は低いと考えられる。ただし、学習ベースの既存手法は、訓練サンプルを十分に収集することが必須である。提案手法の利点として、訓練サンプルの収集が不要であることが挙げられる。一方、提案手法の欠点として、学習ベースの既存手法と比べて、大きな精度改善を見込めないことが挙げられる。今後の課題として、学習ベースの手法に提案手法を組み込むことで、さらなる精度の向上を目指す必要がある。

## 5. 公開データセットにおいて低解像度画像から人物属性の共起特徴量を抽出した場合の検証

### 5.1 リサンプリングの比較

本実験では、3.4.1 で述べた人物属性の共起特徴量をを用いることで、公開データセット CUHK01 における人物対応付けの精度を評価した。低解像度の入力画像を準備するため、解像度が (30, 80), (24, 64), (18, 48),

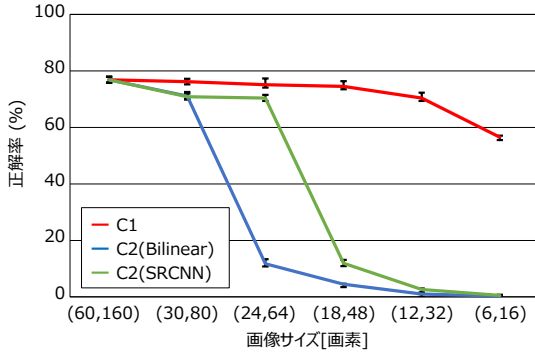


図 10 CUHK01 における人物対応付けの精度.

Fig.10 Accuracy of person re-identification on the CUHK01 dataset.

(12, 32), (6, 16) 画素のものを、式 (9) を用いて生成した。式中のパラメータは 4.3 と同じとした。また、高解像度の入力画像として、(60, 160) 画素のものをを用いた。辞書画像として、入力画像とは別の時刻で撮影されたものをを用いた。辞書画像の解像度は (60, 160) 画素とした。CUHK01 データセットに含まれる 971 名の中から、ランダムに 486 名を選択し、LMNN の計量行列を学習するための訓練サンプルとした。残りの 485 名の人物画像を用いて、対応付けの入力画像と辞書画像とを設定した。このランダム選択を 3 回実施した。各回で対応付け精度を算出し、それらの精度の平均と標準偏差を求めた。本実験では、3.1 で述べた C1 のダウンサンプリングと、C2 のアップサンプリングとの精度を比較した。なお C2 のアップサンプリングの手法として、バイリニア補間と SRCNN を用いた。

公開データセット CUHK01 において、低解像度の入力画像に C1 または C2 を適用し、リサンプリングされた画像から共起特徴量を抽出することで、人物対応付けを行った場合の精度を図 10 に示す。C2 のアップサンプリングで用いた SRCNN は、バイリニア補間より、解像度が (24, 64), (18, 48) 画素の時に精度を高めた。C1 のダウンサンプリングは、C2 のアップサンプリングより、精度を大きく改善することを確認した。

## 5.2 低周波成分と高周波成分の比較

公開データセット CUHK01 において、低周波成分と高周波成分の画像から、人物属性の共起特徴量を抽出した場合について検証した。高解像度の入力画像  $f^q$  から、低解像度の入力画像  $g^q$  を式 (9) で生成した。 $f^q$  は (60, 160) 画素、 $g^q$  は (6, 16) 画素とした。この  $g^q$  を (60, 160) 画素までアップサンプリングし、入力

表 2 各周波数帯の画像における RGB 値の平均と標準偏差.

Table 2 Average and standard deviation of RGB values for images in each frequency band.

		R	G	B
元画像	$f^q$	100.8±60.4	102.1±60.2	107.0±60.1
	$f_l^q$	100.8±49.7	102.1±49.3	107.0±49.2
	$f_h^q$	0.0±34.3	0.0±34.5	0.0±34.5
Bilinear	$g^{q\uparrow}$	99.1±40.4	100.4±40.8	105.4±40.4
	$g_l^{q\uparrow}$	99.1±40.7	100.4±40.2	105.4±39.8
	$g_h^{q\uparrow}$	0.0±7.4	0.0±7.2	0.0±7.0
SRCNN	$g^{q\uparrow}$	98.4±47.5	99.7±47.1	104.7±47.0
	$g_l^{q\uparrow}$	98.4±46.5	99.7±46.7	104.7±46.1
	$g_h^{q\uparrow}$	0.0±9.3	0.0±9.1	0.0±9.1

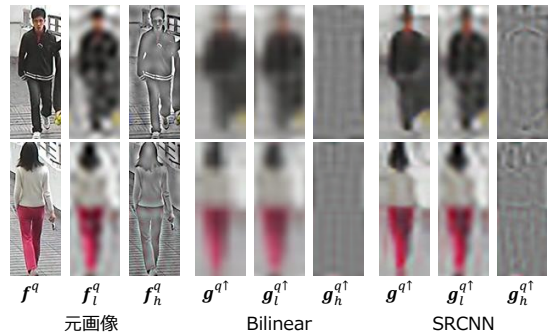


図 11 各周波数帯の画像の可視化.

Fig.11 Visualization of images in each frequency band.

画像  $g^{q\uparrow}$  を生成した。アップサンプリングにはバイリニア補間または SRCNN を用いた。人物属性の共起特徴量を用いる場合、特徴量同士の距離を、周波数領域で直接計算することは難しい。本実験では、 $g^{q\uparrow}$  を周波数領域で高周波成分と低周波成分に分離し、逆離散フーリエ変換を用いて、それぞれの成分を空間領域へ変換した。高周波成分の画像を  $g_h^{q\uparrow}$  で表し、低周波成分の画像を  $g_l^{q\uparrow}$  で表す。同様に、高解像度の入力画像  $f^q$  の高周波成分の画像を  $f_h^q$  で表し、低周波成分の画像を  $f_l^q$  で表す。

各周波数帯の画像において、RGB 値の平均と標準偏差を求めた。その結果を表 2 に示す。 $f_h^q$  では、 $g_h^{q\uparrow}$  と比べて標準偏差が大きい。SRCNN の  $g_h^{q\uparrow}$  では、バイリニア補間の  $g_h^{q\uparrow}$  と比べて標準偏差が僅かに大きい。高周波成分の平均値は 0 であるが、標準偏差はこれらの値を持つため、高周波成分にはアップサンプリングの効果で何らかの値が入っていると見える。次に、高周波成分の画像と低周波成分の画像を可視化した結果を図 11 に示す。なお、高周波成分の画像は負値も



表 3 各周波数帯の画像を用いた場合の第 1 位正解率と AUC の比較.

Table 3 First matching rate (%) and AUC when using images in each frequency band.

	第 1 位正解率			AUC		
	F1	F2	F3	F1	F2	F3
Bilinear	0.3	<b>65.9</b>	0.2	0.527	<b>0.978</b>	0.505
SRCNN	0.6	<b>71.1</b>	0.2	0.545	<b>0.980</b>	0.502

取るため、高周波成分の可視化のみ、画素値の範囲を  $[-128, 128]$  から  $[0, 255]$  へ射影した。バイリニア補間の  $g_h^{qt}$  を可視化した場合、画像中で僅かな変化が見られた。SRCNN の  $g_h^{qt}$  では人物の輪郭周辺に変化が生じていた。次に低周波成分に注目すると、SRCNN の  $g_l^{qt}$  では、バイリニア補間の  $g_l^{qt}$  と比べて、人物の体型がより視認できる。よって、SRCNN は、高周波成分よりも低周波成分に、効果が表れたと考えられる。

各周波数帯の画像から、人物属性の共起特徴量を抽出した場合について、対応付け精度を表 3 に示す。高周波成分の画像を用いた F3 では、バイリニア補間と SRCNN の両方で精度が低い。全周波数帯の画像を用いた F1 でも同様であった。この理由として、辞書側の高周波成分は存在するものの、入力側の高周波成分が正しく復元されていないため、誤対応が発生したと考えられる。低周波成分の画像を用いた F2 は、F1 と F3 と比べて精度が高い。F2 はダウンサンプリングに相当することから、本実験でもその有効性が確認された。

## 6. まとめ

本論文では、低解像度の人物画像に合わせてダウンサンプリングを施すアプローチを採用することで、人物対応付けの精度が改善されるかどうかを検証した。カメラレンズの焦点距離、および、人物からカメラまでの位置を変化させた場合の評価データセットを構築した。実験結果より、低解像度画像に合わせるダウンサンプリングが、高解像度画像に合わせるアップサンプリングと比べて、人物対応付けの精度改善に大きく寄与することを確認した。アップサンプリングされた画像の高周波成分と低周波成分について、人物対応付けの精度を比較し、ダウンサンプリングで得られる低周波成分に個人性を表す特徴が多く含まれる可能性が高いことをシミュレーション実験で確認した。

今後の課題として、対応付けの精度をより高めることができるリサンプリング手法の検討が挙げられる。

本論文では、性別、体型、服装、髪型などの属性に起因する人物の見え方を個人性と考へ、対応付けの実験を行ったが、様々な要因で見え方は変化するため、多様なデータセットでさらなる評価が必要である。本論文で得た知見に対して、実際の環境における画像生成過程や光学系との関連性を考察することも必要である。

## 謝 辞

本研究の一部は、JSPS 科研費 JP20K11864 の助成を受けたものである。

## 文 献

- [1] O. Hamdoun, F. Moutarde, B. Stanculescu, and B. Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *Proceedings of the Second ACM/IEEE International Conference on Distributed Smart Cameras*, pp. 1–6, 2008.
- [2] T. Wang, S. Gong, X. Zhu, and S. Wang. Person re-identification by video ranking. In *Proceedings of the European Conference on Computer Vision*, pp. 688–703, 2014.
- [3] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio. Full-body person recognition system. *Pattern recognition*, Vol. 36, No. 9, pp. 1997–2006, 2003.
- [4] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. In *Proceedings of the Computer Vision and Pattern Recognition*, pp. 4184–4193, 2015.
- [5] W. Li, R. Zhao, and X. Wang. Human re-identification with transferred metric learning. In *Proceedings of the Asian Conference on Computer Vision*, pp. 31–44, 2012.
- [6] X. Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, and B. Xu. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 695–704, 2015.
- [7] X. Li, W. S. Zheng, X. Wang, T. Xiang, and S. Gong. Multi-scale learning for low-resolution person re-identification. In *Proceedings of the Proceedings of the IEEE International Conference on Computer Vision*, pp. 3765–3773, 2015.
- [8] Z. Wang, R. Hu, Y. Yu, J. Jiang, C. Liang, and J. Wang. Scale-adaptive low-resolution person re-identification via learning a discriminating surface. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, Vol. 2, p. 6, 2016.
- [9] Zheng W., Mang Y., Fan Y., Xiang B., and Shin'ichi S. Cascaded sr-gan for scale-adaptive low resolution person re-identification. In *Proceedings of the International Joint Conference on Artificial Intelligence*,

- pp. 3891–3897, 2018.
- [10] J. Jiao, W. S. Zheng, A. Wu, X. Zhu, and S. Gong. Deep low-resolution person re-identification. In *Proceedings of the Thirty-second Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*, 2018.
- [11] Z. Wang, Z. Miao, Q. M. J. Wu, Y. Wan, and Z. Tang. Low-resolution face recognition: a review. *The Visual Computer*, Vol. 30, No. 4, pp. 359–386, 2014.
- [12] Y. Shi and A. K. Jain. Probabilistic face embeddings. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6902–6911, 2019.
- [13] S. P. Mudunuri and S. Biswas. Low resolution face recognition across variations in pose and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 5, pp. 1034–1040, 2015.
- [14] W. W. W. Zou and P. C. Yuen. Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, Vol. 21, No. 1, pp. 327–340, 2011.
- [15] J. S. Park and S. W. Lee. An example-based face hallucination method for single-frame, low-resolution facial images. *IEEE Transactions on Image Processing*, Vol. 17, No. 10, pp. 1806–1816, 2008.
- [16] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 2, pp. 295–307, 2015.
- [17] M. Nishiyama, S. Nakano, T. Yotsumoto, H. Yoshimura, Y. Iwai, and K. Sugahara. Person re-identification using co-occurrence attributes of physical and adhered human characteristics. In *Proceedings of the International Conference on Pattern Recognition*, pp. 2085–2090, 2016.
- [18] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, Vol. 10, No. Feb, pp. 207–244, 2009.

(平成 xx 年 xx 月 xx 日受付)

### 西山 正志 (正員：シニア会員)

2000 年 岡山大学工学部情報工学科卒業。2002 年 同大学院博士前期課程了。同年株式会社東芝入社。同社研究開発センターを経て、現在鳥取大学大学院工学研究科准教授。2011 年 東京大学大学院学際情報学府にて博士(学際情報学)を取得。カメラを用いた人物認識を始めとするパターン認識およびインタラクションの研究に従事。山下記念研究賞や画像センシングシンポジウム優秀学術賞など受賞。電子情報通信学会、情報処理学会各会員。

### 岩井 儀雄 (正員)

1992 年(平成 4 年)大阪大学基礎工学部情報工学科卒業。1997 年(平成 9 年)大阪大学大学院基礎工学研究科博士課程後期修了。同年同大学院助手。2003 年(平成 15 年)同大学院助教授。2004 年(平成 16 年)5 月～2005 年(平成 17 年)3 月英国ケンブリッジ大学客員研究員。2007 年(平成 19 年)同大学院准教授。2011 年(平成 23 年)鳥取大学大学院工学研究科教授。コンピュータビジョン、パターン認識の研究に従事。博士(工学)

### 遠藤 拓弥

2019 年 鳥取大学工学部電気情報系科卒業。2019 年 鳥取大学大学院持続性社会創生科学研究科工学専攻博士前期課に在学中。