

Jensen Shannon カーネルとカーネル最大マージン成分分析による カメラの違いの影響を受けにくいカメラ間人物照合

井尻 善久^{†,††a)} 勞 世竝^{††b)} 村瀬 洋^{††c)}

Robust Human Re-Identification Based on Kernel Large Margin Component Analysis and Jensen Shannon Kernel

Yoshihisa IJIRI^{†,††a)}, Shihong LAO^{††b)}, and Hiroshi MURASE^{††c)}

あらまし 人物の追跡は監視カメラの重要なアプリケーションの一つである。一台の監視カメラの視野より広い領域において人物の追跡を行うためには、複数の監視カメラを利用し、それらカメラ間で人物対応付け（人物照合）をする必要がある。一方複数のカメラを用いるとしても、完全に死角のない領域を広範囲にわたって構築することは、コスト的あるいは景観的な問題から、実現が困難であることも多い。したがって本論文では、視野を共有しない複数のカメラ間での人物照合を考える。視野を共有しないカメラ間では、各カメラの色校正法や、監視対象の人物の体の向きや姿勢、照明条件が異なる。これに対し従来研究では、人物の服装の色が短時間には変化しないことを前提に、各人物の服装色ヒストグラムによる単純な照合が提案されてきた。しかしながら、異なる色校正を受けた色ヒストグラムを単純な距離指標で照合することは精度低下につながると考えられる。このため提案手法では、ヒストグラム間の距離指標を高次元非線形空間で学習し、カメラによる違いの影響を受けにくい照合を目指す。提案手法の有効性は、実験により実証する。

キーワード 距離指標, カーネル, 学習, 人, 照合

1. ま え が き

安心安全な社会の実現に向けて、人物監視を目的とした大量の監視カメラが導入されている。これに伴い、増大する監視業務を支援するために、様々な自動監視や監視支援システムが提案されてきた。中でも人物追跡は重要なアプリケーションの一つである。特に一台の監視カメラの視野は限られていることから、広い領域で自由に行き交う人物の動きを追跡し続けるためには、複数の監視カメラによる追跡が実用上重要となる。一方カメラの台数が増えても、完全に死角のない監視

エリアを構築するためには、相当数の監視カメラが必要とされ、コスト的にも景観的にも現実的でない場合が多い。したがって一般的には、視野を共有しない複数の監視カメラによる追跡が必要となる。このためには、カメラ視野から外れた人物の情報を記録しておき、その人物が別のカメラ視野に入ってきた際に、あらかじめ記録した人物情報をもとに同一人物として照合することが必要となる。そこで本論文においては、こうした視野を共有しないカメラ間での人物照合問題を考える。

カメラ間人物追跡が高精度にできるようになれば、例えば、マーケティングのために店舗内やショッピングモールにおける人物の来店経路を知ることができる。また、ショッピングモール等における万引き・窃盗等の際には、被疑者の写真をもとに場内監視カメラ映像と照合を行い、早期検挙につなげることができる。また犯罪捜査等で犯人の足取りを知るのにも有効である。

カメラ間人物照合を実現するためには、既存の顔認識 [1] や顔属性推定 [2] を用いることが考えられる。しかしながら、図 1 に示したように、人物は自然な動きをしているため様々な方向を向いて撮影される。また

[†] オムロン株式会社技術本部コアテクノロジーセンタ, 木津川市
Core Technology Center, Corporate R&D, OMRON Corp.,
9-1 Kizugawadai, Kizugawa-shi, 619-0283 Japan

^{††} 名古屋大学大学院情報科学研究科, 名古屋市
Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya-shi, 464-8603 Japan

^{†††} オムロンソーシャルソリューションズ株式会社, 木津川市
OMRON Social Solutions Co. Ltd., 9-1 Kizugawadai,
Kizugawa-shi, 619-0283 Japan

a) E-mail: joyport@ari.ncl.omron.co.jp

b) E-mail: lao@ari.ncl.omron.co.jp

c) E-mail: murase@is.nagoya-u.ac.jp



図1 VIPeR DBの例:(a)(b)は異なるカメラで撮られたID1019の人物, 以下同様に(c)(d)はID302, (e)(f)はID188, (g)(h)はID358

Fig. 1 Examples of VIPeR DB: (a)(b) are from ID19. Similarly, (c)(d) from 302, (e)(f) from 188, and (g)(h) from 358.

監視カメラは監視エリア全体を撮影することに主眼においており、追跡対象人物に焦点が合っているとは限らない。したがって実用的な条件では、監視対象人物の鮮明な顔画像をいつも取得できることは想定することはできず、顔認識だけでは不十分であることが多い。

したがって本論文では、顔認識以外の人物照合方法として、人物の服装色を用いた照合方法を検討する。比較的短時間であれば服装は変化しないことを仮定できるので、前記のようなアプリケーションにおいてこれは有効である。また多くの場合、服装色には前後左右で、ある程度、同色が使われていることを仮定できるので、様々な方向を向いた人物を照合するのに適していると考えられる。

一方で、カメラ間の人物照合には様々な課題がある。一般に異なる監視カメラにおいては色校正が異なる。加えて、撮影している環境の照明条件等はカメラ毎、時間帯ごとに異なる。また対象となる人物は様々な方向を向いており、同じ人物であっても見え方が異なる。更にこれらが複合条件として生じることを考えると、様々な変動を受けるので高精度な照合が難しいことが分かる。

こうした課題を解決しより高い照合精度を実現するために、様々な先行研究がなされてきた。それらは、(i) 見え方に基づく手法、(ii) カメラ間の色校正に基づく手法、(iii) カメラ間の関係モデリングによる手法の三つに大別できる。

(i) の見え方に基づく手法に関しては、服装の色や模様をうまく取り出す特徴量抽出方法 [3]～[15] が数多く提案されている。これらの方法に共通する特徴は、人物の見え方が大きく変動する中で精度を維持するため、複数の特徴量を冗長に抽出することである。特に、服装の色や模様の特徴を抽出するために、様々な色空間における色ヒストグラムや、局所特徴量が組み合わせて用いられている。中でも、Farenzena ら [14] は、服装の色情報を抽出するのに重み付き HSV ヒストグラムと Maximally Stable Color Region (MSCR)、服装の模様特徴を抽出するのに Recurrent High-Structured Patches (RHSP) を用い、それらによる照合結果を統合して現在最高精度の人物照合精度を実現している。更にこの手法は、人物の対称である軸（左右）と非対称な軸（上下）により人物領域を分割し、分割された各領域において特徴抽出することから、Symmetric Drive Accumulation of Local Features (SDALF) と呼ばれている。この手法における重み付き HSV ヒストグラムは、単純な Bhattacharyya 距離により照合されるが、異なるカメラでは色校正が異なるので、こうした単純な距離指標で照合すると精度低下につながると考えられる。

一方、カメラ間の色校正が異なる課題に対しては、カメラ間でのカメラ間の色校正（前記(ii)）による高精度化が提案されている。中でも輝度伝達関数(Brightness Transfer Function; BTF) [9], [16]～[18] が最も一般的である。BTFの推定はカメラの組合せそれぞれについて行う必要があるので比較的手間がかかる。実際カメラペア数はカメラの台数 N に対して $N(N-1)/2$ となり、カメラ台数が多い場合には現実的とはいえない。また、ネットワーク内の各カメラの視野への人物の出入りの時間が統計的な関係をもつと仮定し、カメラ間の人物移動の統計的モデルを学習し、異なるカメラで撮影された人物のペアが同一人物である確率を計算する方法が提案されている（前記(iii)）。こうした考え方をを用いた例 [13], [16], [19] は多数あるが、人物がほぼ一定の速度で移動することを仮定している点や、全てのカメラペアの通り方に対して一定以上の人物の対応付けデータを集める手間がかかる点等が課題である。

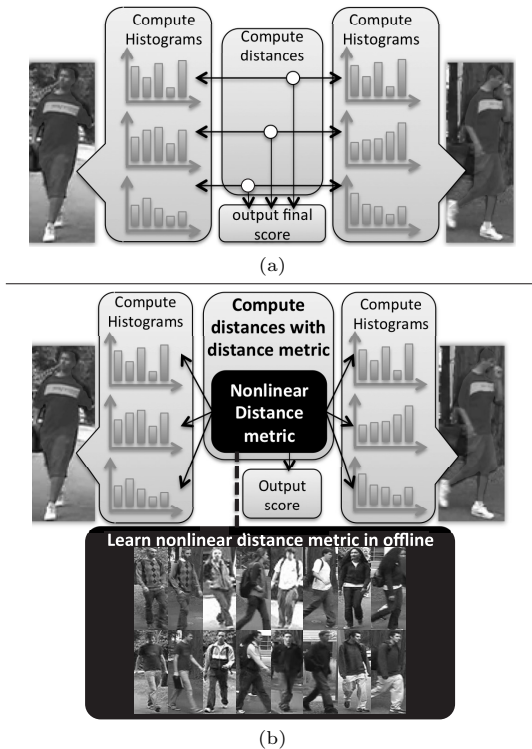


図2 従来手法 (a) と提案手法 (b) の比較

Fig. 2 (a) Conventional and (b) proposed color matching schemes.

こうした既存研究を俯瞰すると、服装色は最も基本的かつ重要な特徴であることが分かるが、異なる色校正のカメラから得られた色ヒストグラムが単純な距離指標で照合されている点が課題である。このため本論文では、複数カメラ間で照合を行うのに最適な色ヒストグラム間距離指標を学習することを提案する。提案する手法は、カメラの台数が多い場合にその対応関係を意識することなく手軽に最適化ができ、カメラの増減や変更に対しても簡単に精度向上を望める手法である。更に、距離指標の学習において非線形カーネル関数を用いるので、非線形な距離指標を学習することができる。本論文では、こうした特徴をもった手法として、距離指標学習の一つである Large Margin Component Analysis (LMCA) と Jensen-Shannon カーネル関数の組合せを提案する。

したがって、本研究の貢献は次のとおりである。複数カメラ間での人物服装色による照合問題において、(a) 従来の色ヒストグラム間の単純な距離指標と比べ高い照合精度を得るため、非線形な距離指標を学習す

ることを提案する。(b) その具体的手法として、代表的な距離指標学習である Large Margin Component Analysis (LMCA) と Jensen-Shannon カーネル関数の組合せが有効であることを示す。

本手法の基本的な考え方は、従来手法との比較とともに、図2に示した。SDALFのHSVヒストグラム照合に代表される従来手法の最も基本的な照合方法(図中(a))では、モデル及び照合対象の各色成分の色ヒストグラムを単純な距離指標により照合し、各色成分における照合スコアを統合する。一方、提案手法(図中(b))では、色ヒストグラム抽出後、あらかじめ学習しておいた最適な距離指標を用いて色照合を行う。この距離指標は非線形な距離指標であり、適当な非線形距離指標を選択することで線形な距離指標を用いる場合に比べより高い照合精度を得ることができる。

本論文の以降の章は次のように構成される。まず、2.において提案手法について説明する。この中で、基本となる既存手法を振り返り、精度向上のために用いる Jensen-Shannon カーネルを説明する。次いで、3.において、提案手法の効果を実証するための実験を行う。最後に4.で結論づける。

2. 提案手法

以降の記述では、人物領域が人体検出や背景差分等により特定されており、学習データとして様々なカメラ及び様々な撮影条件で撮影された学習画像と人物のラベル情報が与えられることを前提とする。提案手法の大まかな流れを説明する。まず、距離指標学習時には、人物が N 人含まれる学習画像から人物領域を抽出し N 個の正規化色ヒストグラム $X = \{x_n; n = 1, \dots, N\}$ を抽出する。正規化色ヒストグラムは色ヒストグラムを全体の度数で正規化したものである。次いで、この正規化色ヒストグラム X 及びラベル情報 $Y = \{y_n; n = 1, \dots, N\}$ を用いて、正規化色ヒストグラム間の最適な距離指標を学習する。学習時に、照合時に実際に起こり得る変動を含んだ十分な枚数の画像から学習しておくことにより、カメラの組合せ等によらず、最適な照合を行うことができると考えられる。

一方、人物の登録は、各監視カメラを一人の人物が通過するたびに行われる。登録時には、各人物に対して学習時と同様の方法で正規化色ヒストグラム m_c を抽出する。ここで、登録された人物が C 人であったとすると、全体で $M = \{m_c; c = 1, \dots, C\}$ のモデル特

微量を得る。

また照合時には、通過する人物一人ひとりに、学習時や登録時と同様の方法で正規化色ヒストグラム t_k を抽出する。照合用の画像が K 枚であったとすると、全体で $T = \{t_k; k = 1, \dots, K\}$ の照合用特徴量を得られるが、これら照合対象の正規化ヒストグラム T とモデル登録された正規化ヒストグラム M とを、距離指標学習時に学習しておいた距離指標を用いて照合を行う。

提案手法では、最適な距離指標を学習するために、サポートベクタマシン (SVM) 等で利用されるマージン最大化規範を用い最適化を行う。最適化においては、より高い精度を得るために、高次元非線形写像 $\phi(x_n)$ を用いる。こうした非線形写像を行うためのカーネル関数として、ヒストグラムの照合に適していると考えられるいくつかのカーネル関数を検討し、次章の実験により、最適なカーネル関数を明らかにする。

2.1 距離指標の学習

距離指標の学習においては、教師無しの学習法に加え、多くの教師付きの学習法が提案されてきた [20]。本論文においては、ラベル付きのデータ（各人のデータは少数）が得られることを前提としているので、教師付きの学習を行うことができる。

教師付き学習の中で最も一般的な手法として、線形判別分析がある。しかしこの方法では、クラス内散布行列を求めるのに十分な数のデータ（各人物当りの学習データ）がない場合には解の汎化性が欠け、またクラス数（学習データに含まれる固有な人物の人数）が少ない場合にはクラス間散布行列がランク制約から十分な数の判別軸をとれない等、適用に制限も大きい。これらの制限は学習データの与え方に影響するが、十分な人数の人物に対し十分な枚数の学習データを整えるのは手間がかかる。

一方、線形な距離指標に加え、より高い汎化精度を得るため、カーネル関数を用いて非線形な距離指標を学習する手法も提案されてきた。カーネル法の代表的な手法として SVM が提案されているが、SVM は基本的に 2 クラス識別のための手法であるので、今回のような多クラス識別問題に直接適用することはできない。2 値識別器を多クラス識別問題に適用するには、しばしば、同じラベルをもつペアの照合スコアと、異なるラベルをもつペアの照合スコアとを分離する、2 クラス問題を解けばよい。しかしながら、このような問題設定で学習する際には、多くの場合、同じラベルをもつデータ数に比べ異なるラベルをもつデータ数

が極端に多くなる、いわゆる不均衡問題 (imbalanced problem) が生じてしまう。仮にこれを回避するのにサンプリングを行い抽出された固定的な部分集合を学習に用いたとしても、必ずしも学習に適したサンプルを選択できるとは限らず、過適合の可能性が高くなる。

こうした問題を回避しながらも SVM のようにマージン最大化規範で多クラス間距離指標を最適化する方法として、Large Margin Nearest Neighbor (LMNN) [21] が提案されている。この方法は最近傍法に基づく手法であるので、本質的に多クラス識別問題であり、前述のような問題を生じない。汎化精度の観点からも、LMNN は距離指標学習においては最も優れた手法の一つと考えられている [22]。一方、Torresani らは LMNN を改良し、入力データの次元数が多いときにも簡単に最適化でき、またカーネル法による非線形距離指標を学習できるようにするための方法として、Large Margin Component Analysis (LMCA) [23] を提案した。そこで本論文では、マージン最大化規範に基づく距離指標を学習するのに最新、最適、かつ簡易であると考えられる LMCA を用いることにする。以下に簡潔に LMCA を振り返り、これをどのように本手法に応用するかを示す。

線形の LMCA: 正規化色ヒストグラム X 及び対応する学習データのラベル Y が与えられたとき、LMCA は次の損失関数 $\epsilon(L)$ を最小化する。

$$\epsilon(L) = \sum_{ij} \eta_{ij} \|L(x_i - x_j)\|^2 + c \sum_{ijl} \eta_{ij} (1 - y_{il}) \cdot h(\|L(x_i - x_j)\|^2 - \|L(x_i - x_l)\|^2 + 1). \quad (1)$$

ただし、 $L \in \mathcal{R}^{d \times D}$ は X に対する線形写像、 $\eta_{ij} \in \{0, 1\}$ は x_j と x_i が同じラベルをもち ($y_i = y_j$) かつ k -近傍であるときに 1 をとる 2 値変数、 $y_{il} \in \{0, 1\}$ は $y_i = y_l$ のときに 1 をとる 2 値変数、 $c > 0$ は適当なバランスパラメータである。また、 $h(s)$ は $h(s) = \max(s, 0)$ で定義されるヒンジ関数であり、SVM 等で使われるソフトマージンを与えるための損失関数と同じである。この目的関数のうち、最初の項は同じラベルをもつデータペア間の距離を最小化し、不変性 (invariance) を高める効果をもつ。もう一つの項は、異なるクラスラベルをもつデータペア間 (x_i と x_l) 距離が、同じクラスかつ k -近傍のデータ間 (x_i と x_j) 距離と比べ、1 のマージンをもちながら離れるようにし、分離性 (discriminability) を高める効果をもつ。またヒンジ関数は、十分なマージンをもってい

るデータに対しては何も作用せず、マージンを侵したデータだけにロスを発生させる効果をもつ。これにより学習に適したデータが常に再サンプリングされるので、事前に固定的なペアを選択するよりも高い汎化性を期待することができる。なお最適化はこう配法により行う。

カーネル LMCA：入力データ x_i をより高次元の非線形空間に写像するための、内積がカーネル関数 $K(x_i, x_j) = \phi_i^T \phi_j$ で表されるような写像 ϕ を考える。以降の記述では、 $\phi_i = \phi(x_i)$ 、 $\Phi = [\phi_1, \dots, \phi_N]^T$ 、 $k_i = \Phi \phi_i = [k(x_1, x_i), \dots, k(x_N, x_i)]$ と表す。この非線形写像を用いて、写像された空間におけるロス関数は次のように表される。

$$\epsilon(L) = \sum_{ij} \eta_{ij} \|L(\phi_i - \phi_j)\|^2 + c \sum_{ijl} \eta_{ij} (1 - y_{il}) \cdot h(\|L(\phi_i - \phi_j)\|^2 - \|L(\phi_i - \phi_l)\|^2 + 1). \quad (2)$$

したがって、ロス関数の L に関するこう配は次のようになる。

$$\begin{aligned} \frac{\partial \epsilon(L)}{\partial L} &= \sum_{ij} \eta_{ij} L(\phi_i - \phi_j)(\phi_i - \phi_j)^T \\ &\quad + c \sum_{ijl} \eta_{ij} (1 - y_{il}) h'(s_{ijl}) \\ &\quad \cdot L[(\phi_i - \phi_j)(\phi_i - \phi_j)^T - (\phi_i - \phi_l)(\phi_i - \phi_l)^T], \quad (3) \end{aligned}$$

ただし、

$$s_{ijl} = \|L(\phi_i - \phi_j)\|^2 - \|L(\phi_i - \phi_l)\|^2 + 1. \quad (4)$$

である。ここで、 $L = \Omega \Phi$ となるように L を選ぶことにより、式 (3) は次のようにカーネル関数のみを用いた表現にすることができる。

$$\begin{aligned} \frac{\partial \epsilon(L)}{\partial L} &= 2\Omega \sum_{ij} \eta_{ij} [E_i^{(k_i - k_j)} - E_j^{(k_i - k_j)}] \Phi \\ &\quad + 2c\Omega \sum_{ijl} \eta_{ij} (1 - y_{il}) h'(s_{ijl}) \\ &\quad \cdot [E_i^{(k_i - k_j)} - E_j^{(k_i - k_j)} - E_i^{(k_i - k_l)} + E_l^{(k_i - k_l)}] \Phi, \quad (5) \end{aligned}$$

ただし、

$$s_{ijl} = \{|\Omega(k_i - k_j)|^2 - |\Omega(k_i - k_l)|^2 + 1\}, \quad (6)$$

であり、 $E_i^v = [0, \dots, 0, v, 0, \dots, 0]$ は i 列でのみ v の値をとりそれ以外の成分は零である $n \times n$ 行列である。

また、 $h'(s)$ の $h(s)$ 微分である。 $h'(s)$ は 0 付近で不連続であり微分が定義されないで、平滑化ヒンジ関数を用いることが提案されている [24]。式 (5) を用いて最急降下法による写像 L の更新則は次のように与えられる。

$$L \leftarrow L - \lambda \frac{\partial \epsilon(L)}{\partial L}. \quad (7)$$

ただし、 λ は適当な係数である。ここで式 (5) において、 $\frac{\partial \epsilon(L)}{\partial L}$ の全ての項に Φ が掛かっていることと、左辺の $L = \Phi \Omega$ と仮定していることを考えると、式 (7) における全ての Φ は消える。したがって最急降下法の更新則は、次の Ω に関する簡単なこう配法に帰着される。

$$\Omega \leftarrow \Omega - \lambda \Gamma, \quad (8)$$

$$\begin{aligned} \Gamma &= 2\Omega \sum_{ij} \eta_{ij} [E_i^{(k_i - k_j)} - E_j^{(k_i - k_j)}] \\ &\quad + 2c\Omega \sum_{ijl} \eta_{ij} (1 - y_{il}) h'(s_{ijl}) \\ &\quad \cdot [E_i^{(k_i - k_j)} - E_j^{(k_i - k_j)} - E_i^{(k_i - k_l)} + E_l^{(k_i - k_l)}]. \quad (9) \end{aligned}$$

仮定により、非線形空間への写像は次の式で簡単に計算することができる。

$$L\phi_q = \Omega \Phi \phi_q = \Omega k_q. \quad (10)$$

この場合、最適な距離指標は $k_q^T \Omega^T \Omega k_q$ として得られる。

2.2 従来研究におけるカーネル関数

本論文において利用する正規化色ヒストグラムのような二つの分布 $a, b \in \mathcal{R}[0, 1]^D$ に有効なカーネル関数 $K(a, b)$ として、次のようなカーネル関数が提案されてきた。なお以降の記述では $a = \{a_i; i = 1, \dots, D\}$ 、 $b = \{b_i; i = 1, \dots, D\}$ とする。

Histogram Intersection カーネル [25]：

$$K(a, b) = \sum_i \min(a_i, b_i) \quad (11)$$

χ^2 カーネル [26]：

$$K(a, b) = \exp \left\{ - \frac{\left(\sum_i \frac{a_i - b_i}{\frac{1}{2}(a_i + b_i)} \right)^2}{\sigma^2} \right\} \quad (12)$$

Bhattacharyya カーネル [27]：

$$K(a, b) = \sum_i \sqrt{a_i b_i} \quad (13)$$

2.3 提案手法で用いるカーネル関数：Jensen-Shannon カーネル

本節においては、特に確率分布や本論文での正規化色ヒストグラムの照合に適したカーネル関数を検討する。服装色ヒストグラムの照合においては、様々な変動により不確実性を有するヒストグラム間の照合が必要となる。これを実現するため、本論文では、照合すべき二つの服装色ヒストグラムという観測情報が与えられたときに、それら観測情報をもつ情報量の差に注目することを提案する。情報量の差を考える上で、Kullback-Leibler divergence (KLD) は一般的であるが、対称性を満たさない（また三角不等式も満たさないため）に距離の公理を満たさない。したがって、KLD を直接カーネル関数として用いることはできない。これを解決するために、距離の公理を満たす Jensen-Shannon divergence (JSD) [28] が提案されている。更に、この JSD を用いた次のようなカーネル関数が提案されている [29]。

$$K(a, b) = \exp(-\text{JSD}/\sigma^2)$$

$$= \exp \left\{ - \frac{\sum_i \left(\frac{a_i}{2} \log \frac{2a_i}{a_i+b_i} + \frac{b_i}{2} \log \frac{2b_i}{a_i+b_i} \right)}{\sigma^2} \right\}. \quad (14)$$

このカーネル関数は正定値カーネルである [29]。一方、基本となる JSD は情報量的な距離を評価するのに適していることから、本論文に類似の課題をもつ様々なパターン認識に応用され有効性が示されている [30] ので、本論文で目的とするような正規化色ヒストグラムの照合にも適していると思われる。

この仮説を実証するため、ガウス混合モデルにより生成された人工的なデータを用いて予備実験を行った。HSV 色空間における HSV joint histogram を利用すると仮定すると、本色ヒストグラム照合問題に特有の点として、色分布が循環性をもつことが挙げられる。すなわち、色相 (hue) 成分は極座標系において表現されているので、照明変動が起こると極座標的すなわち循環的なシフトを起こしやすい。また、関連して、彩度 (saturation) や輝度 (value) もシフトを起こしやすい。このような場合にも、正しく同じ分布として認識されなければならない。このような同じ人物の変動を表現するため、元の分布とそれを循環的にシフトさせた分布のペアを用いた（以降、同一人物内変動データ）。また別の人物から得られた色ヒストグラム

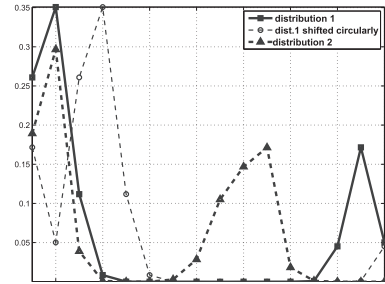


図 3 生成したテスト分布
Fig. 3 Synthetic test distributions.

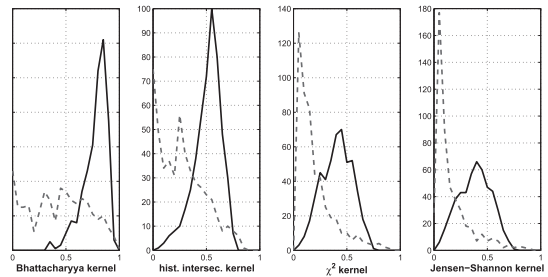


図 4 各カーネル関数による類似度の分布
Fig. 4 Distributions of similarity scores in each kernel function.

を表現するには、生成した個別の分布のペアを用いた（以降、人物間変動データ）。図 3 は用いたデータの一例を示している（太い線が生成した分布 (a)、細い点線はシフトした分布 (b)、太い点線は (a) と異なる別の生成した分布 (c)、(a)(b) のペアが同一人物内変動データ、(a)(c) のペアが人物間変動データ）。このような 500 ペアのデータを生成し、それらそれぞれに対してカーネルの出力する類似度の分布を描画した結果を図 4 に示した（左から順に Bhattacharyya, Histogram Intersection, χ^2 , Jensen-Shannon カーネル、またグラフにおいて通常線は同一人物内変動データの類似度分布、点線は人物間変動データの類似度分布）。これを見ると Jensen-Shannon カーネルにおいて高い分離性が得られていることが分かる。また結果を定量的に表現するために、同一人物内変動データの平均類似度と人物間変動データの平均類似度の比を計算した。結果を表 1 に示した。この結果は大きいほど優れたカーネル関数であることを示している。この結果においても、Jensen-Shannon カーネル関数が、本論文に特有の課題である循環性をもつ分布からなる特徴量間の照合問題に対して優れていることが分かる。

表 1 同一人物間平均類似度と、人物間平均類似度の比
Table 1 Ratios of a mean similarity score within simulated same-person data pairs and that between different-person data pairs.

	Bhat.	HI	$\chi^2(40)$	JS(2)
$\frac{m_{shift}}{m_{diff.}}$	1.8059	1.9113	2.0271	2.3351

2.4 照 合

最終的なモデル特徴量と照合用特徴量は、正規化色ヒストグラム M と T 及び前節のカーネル関数 $K(a, b)$ を用いて、式 (10) により得られる。このようにして得られたモデル及び照合用特徴量をそれぞれ $\Omega K(M, X)$, $\Omega K(T, X)$ とすると、照合用の正規化色ヒストグラム T の各要素の t_j に対して照合結果 ω_j は次のように得られる。

$$s_{ij} = f(\Omega K(m_i, X), \Omega K(t_j, X)) \quad (15)$$

$$\omega(j) = \arg \max_i s_{ij} \quad (16)$$

ただし、 $f(\cdot)$ は任意の類似度関数、 m_i は M の要素である各モデル特徴量を表す。類似度計算には任意の関数を用いることができるが、本検討においては実験で最も精度が良かった単純な相関関数を用いた。

3. 実験結果

提案手法の有効性を確認するために、Gray らによる VIPeR DB [10] を用いて実験を行う。この DB に含まれる人物のサンプルを図 1 に示した。図中において、(a) と (b), (c) と (d), (e) と (f), (g) と (h) は、異なるカメラで撮影された同一人物であり、これらが正しく照合されなければならない。これらのペアを見ると、照合対象の体の向きが異なり、照明条件も異なるため、これらに応じて見え方の変動も大きいことが分かる。更に、この DB ではどのカメラでどのデータが撮影されたのかに関する情報が与えられておらず、BTF 推定等ができないが、これはカメラ台数が多いときにカメラペアごとの最適化が困難である状況と類似である。

この DB には、モデル登録用及び照合用として各 1 枚ずつ 632 人分のデータが含まれているので画像枚数は合計 1264 枚である。この 632 人のデータから 200 人分 (400 枚) をランダムに選び学習用とし、残りの 432 人分を照合用とした。

実験では、色ヒストグラム間照合法単体による差異を比較するために、人体領域のセグメンテーション



図 5 人物領域の分割
Fig. 5 Subregion division.

表 2 累積照合特性
Table 2 Cumulative match characteristics.

	VIPeR 全て		向き限定	
	CMC(1)	CMC(10)	CMC(1)	CMC(10)
Baseline (Euc.)	8.7±0.8	24.2±1.0	14.4±3.7	35.8±3.8
Baseline (NCC)	10.8±1.0	28.6±0.8	14.9±3.6	47.0±4.2
Baseline (HI)	13.2±1.1	35.3±1.0	24.6±4.5	53.7±4.5
Baseline (Bhat)	17.2±0.8	39.2±1.0	32.2±4.2	57.5±5.0
LMCA (lin)	13.0±1.0	37.3±2.4	20.5±3.9	50.7±9.0
LMCA (Ga(5))	15.0±1.2	43.8±1.9	23.0±4.2	61.6±4.4
LMCA (HI)	17.1±2.4	51.1±2.3	28.5±5.7	67.6±5.7
LMCA ($\chi^2(40)$)	19.6±1.5	53.6±1.5	33.8±3.8	68.1±4.9
LMCA (Bhat)	16.2±1.2	46.4±2.1	32.1±5.0	67.6±5.0
LMCA (JS(4))	20.5±1.5	55.7±1.5	35.2±3.6	70.5±5.6

を人為的に行った。実用的には、背景差分等により比較的簡単にセグメンテーションが行えるが、セグメンテーションの結果が完全でない場合には、色ヒストグラムに背景の成分が混入したり逆に人物の領域が欠損して一部の情報が失われたりして、精度が低下することが考えられるので、下記の実験結果は、理想的な場合の結果と考えることができる。

特徴抽出には、HSV 色空間における、正規化 HSV joint histogram を利用したが、その際の色相 (H) 及び彩度 (S) のビン数はそれぞれ 5、輝度 (V) のビンは 3 とした。また人体領域は図 5 の横線で示したような 8 領域に分割し、その各領域で正規化色ヒストグラムを抽出した。これらを連結して特徴量としたので、一人物の特徴量の次元数は $5 \times 5 \times 3 \times 8 = 600$ 次元である。

評価基準として累積照合特性 (Cumulative Match Characteristics; CMC) を用いた。この指標を用いると、正しい照合結果がある順位以内に出る確率が分かる。CMC の横軸には登録された人物の総数まで考えることができるが、実用的には上位に照合されることが重要であるので、特に 1 位照合率 (CMC(1)) と 10 位以内照合率 (CMC(10)) を求め比較した。

10 回のクロスバリデーションによる結果を、表 2 に示した。表中における各数値は平均照合率、± に

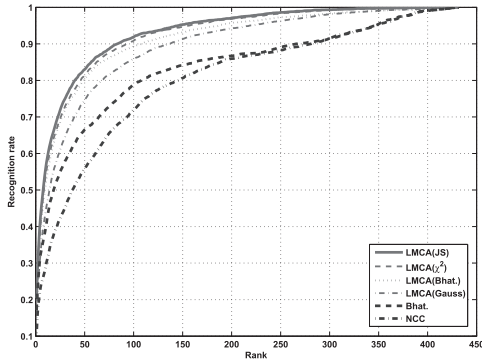


図 6 CMC 曲線：横軸は順位，縦軸は認識率（正しい照合結果が各順位以内に出る確率）

Fig. 6 CMC curve: horizontal axis represents rank, and vertical axis recognition rate (the probability that correct matches appear within each rank).

続く数値は標準偏差を意味している。また，各項目内における括弧内の数値は，用いたカーネル関数のパラメータ値 (σ^2) を表している。左欄は VIPeR のデータ全てを用いた場合を示している。VIPeR DB にはカメラの違いによる照明変動や色変動のみならず人物向きなど様々な変動が含むので，特に色変動に対する精度を見るために，登録及び照合画像間の人物向きの差が 45 度以内のもののみを選び出して照合を行った。この結果が，右欄の「向き限定」の結果である。比較対象として，最も基本的と考えられるユークリッド距離や正規化相関のほか，ヒストグラム間重なり（表中 HI）や Bhattacharyya 距離（Bhat）等，ヒストグラム間照合に適していると思われる単純な距離指標を試した。また LMCA においては，線形の LMCA（表中 LMCA (lin)）のほか，様々なカーネル関数（Gauss (Ga), Bhattacharyya (Bhat), Jensen-Shannon (JS), Histogram Intersection (HI), χ^2 ）を用いたカーネル LMCA を比較した。この結果を見ると，他の様々な手法と比べ LMCA と Jensen-Shannon カーネルを用いた提案手法が高精度であることが分かる。一方で，一般的に用いられる Gauss カーネルは単純な距離指標や線形の場合より低い精度であり，ヒストグラム照合には適していないことが分かった。

次いで，全体の CMC 曲線を，図 6 に示した。視認性の都合上全ての手法を載せることはできなかったが，CMC(1) や CMC(10) だけではなく全体的に提案手法の精度が高いことが分かった。照合結果のいくつかの例を図 7 の (a)～(d) に示した。図中においては，

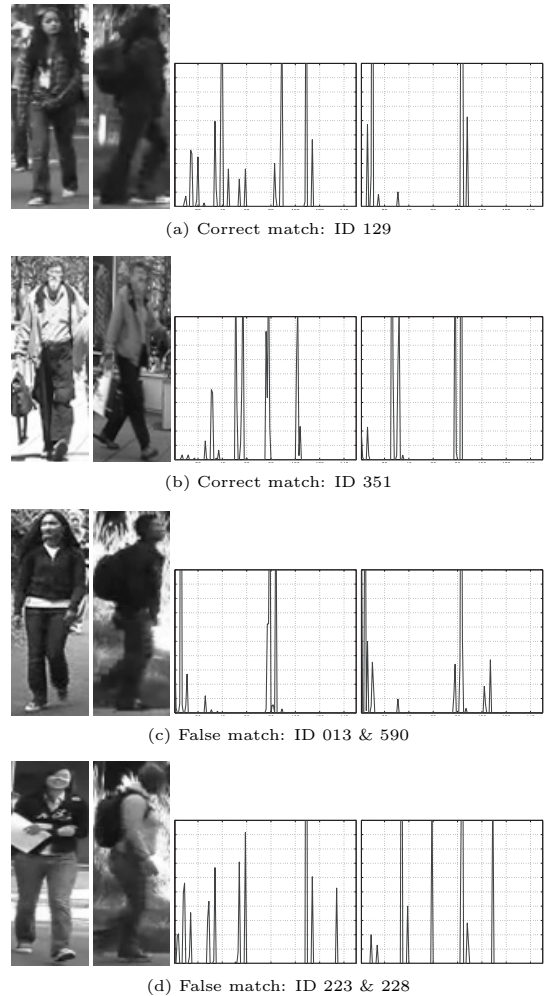


図 7 提案手法による成功例 (a)(b) 及び失敗例 (c)(d)
Fig. 7 Successful (a)(b) and failure (c)(d) cases by the proposed method.

各行左から，一位に照合された登録画像，テスト画像，一位に照合された画像の正規化色ヒストグラム，テスト画像の色ヒストグラムを示している。また正規化色ヒストグラムは次元が高いため視認性の関係上，上半身・下半身の代表として図 5 に示した 3 番目と 7 番目の領域を連結してプロットした。各正規化色ヒストグラムにおいて，横軸はヒストグラムのビンであり縦軸が正規化された度数である。また，前半部分が上半身，後半部分が下半身と対応している。(a) と (b) はそれぞれ同一人物の対であり，(c) と (d) はそれぞれ異なる人物の対である。(a)(b) においては，見え方がかなり異なるにもかかわらず，正しく照合できている。正規化色ヒストグラムを見ても大きな度数をもつビンが

表 3 最新のアルゴリズムとの組合せ評価結果
Table 3 Combination with state-of-the-art algorithm.

	CMC(1)	CMC(10)
SDALF (wHSVhist)	9.8±0.5	28.2±1.7
SDALF (MSCR)	7.8±0.7	23.2±0.8
SDALF (MSCR/wHSVhist.)	14.7±1.1	41.1±1.6
LMCA (JS)+SDALF (MSCR)	21.3±1.2	57.7±2.6

照合対象データ間で HSV 各方向にシフトしているが、このような場合でもある程度照合がうまくいくことが分かった。一方 (c) の対においては、人体の大部分の色が類似しており、正規化色ヒストグラム上でも大きな差異がない。こうした場合には当然ではあるが誤照合が起きた。(d) も誤照合の例であるが、上半身が同じ色 (白/黒) を含んでおり、各色の空間的配置のみが異なっている。図中でも照合対象の二つの正規化色ヒストグラムの上半身 (前半分) 部分で共通する部分があることが分かる。本実験では、各領域ごとにヒストグラムを構築するが、その領域内での空間的な情報を用いない。これは人物の向きに対する不変性を与える上で重要であるが、これと空間的局所性はトレードオフとなる。したがって、こうした例での誤照合は本手法の限界といえる。

最後に、本手法が他の手法と組み合わせてより高い精度を実現できることを示すために、人物照合問題において現在最高精度であると考えられる SDALF [14] と組み合わせて実験を行った。結果は表 3 に示した。実験条件は、上記の実験と全く同じである。最初の三つの実験結果は SDALF において色特徴 (MSCR, 重み付き HSV ヒストグラム, 及びその組合せ) のみを用いた結果を示している。提案手法は、重み付きヒストグラム照合の置き換えと考えることができるので、それを置き換え、提案手法と MSCR の組合せで評価した。その結果が最後の行である。結果的に、距離指標学習に基づく照合を用いることで SDALF の精度は劇的に向上しており、色特徴のみで VIPeR DB における既存手法の最高レベル (SDALF において色特徴及び模様特徴の両方を用いた場合) の精度が実現できている。これは提案手法の有効性を如実に示している結果といえる。

4. む す び

本論文では、人物の服装色の照合によるカメラ間人物追跡において、カメラの違いによる影響を受けにくい照合を目指し、色ヒストグラム間の最適な非

線形距離指標を学習する手法を提案した。またその手法として、Large Margin Component Analysis と Jensen-Shannon カーネルの組合せが優れていることを示した。提案手法は、カメラ間の位置関係を意識しなくても最適化ができるので、特にカメラの台数が多いときに、簡単に精度向上が望める。一方、実応用の観点からは、最高精度を見ても一位照合率は 20% 程度であり、自動人同定システムというよりは人支援システムとしての応用が望まれる。例えば、捜査や特定人物の検索においては、大勢の通行人の中から特定の服装的特徴をもった人物を絞り込み最終的に人が同定する等の支援システムが考えられる。こうした応用例ではある程度候補をだしその中に本人が含まれていればよいことから例えば 10 位以内照合率等の精度が高ければ効果を発揮することができる。今後の精度向上のためには、より良い特徴量を検討することが必要であるが、それは今後の課題である。

謝辞 本研究の一部は NEDO 次世代ロボット知能化技術開発プロジェクト及び文科省による「安全・安心な社会のための犯罪・テロ対策技術等を実用化するプログラム」の一環として実施されました。

文 献

- [1] 井尻善久, 村瀬 洋, “実用化に向けた顔画像処理,” 信学技報, PRMU2010-003, March 2010.
- [2] Y. Ijiri, S. Lao, T.X. Han, and H. Murase, “Efficient facial attribute recognition with a spatial codebook,” ICPR, pp.1461–1464, 2010.
- [3] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, “Shape and appearance context modeling,” Proc. ICCV, pp.1–8, 2007.
- [4] Z. Lin and L.S. Davis, “Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance,” Proc. ISVC, pp.23–34, 2008.
- [5] W.R. Schwartz and L.S. Davis, “Learning discriminative appearance-based models using partial least squares,” Proc. Brazil. Symp. on Comp. Graph. and Image Proc., pp.322–329, 2009.
- [6] A. Alahi, P. Vanderghenst, M. Bierlaire, and M. Kunt, “Cascade of descriptors to detect and track objects across any network of cameras,” CVIU, vol.114, no.6, pp.624–640, 2010.
- [7] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, “Person re-identification using spatial covariance regions of human body parts,” Proc. AVSS, pp.435–440, 2010.
- [8] G. Berdugo, O. Soceanu, Y. Moshe, D. Rudoy, and I. Dvir, “Object reidentification in real world scenarios across multiple non-overlapping cameras,” Proc. Euro. Sig. Proc. Conf., pp.1806–1810, 2010.

- [9] B. Prosser, S. Gong, and T. Xiang, "Multi-camera matching under illumination change over time," Proc. Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, pp.199–205, 2008.
- [10] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," Proc. ECCV, pp.262–275, 2008.
- [11] 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," Proc. ICDS, pp.1–6, 2008.
- [12] C.-H. Kuo, C. Huang, and R. Nevatia, "Inter-camera association of multi-target tracks by on-line learned appearance affinity models," Proc. ECCV, pp.383–396, 2010.
- [13] C.C. Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," Proc. CVPR, pp.1988–1995, 2009.
- [14] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," Proc. CVPR, pp.2360–2367, 2010.
- [15] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino, "Multiple-shot person re-identification by HPE signature," Proc. ICPR, pp.1413–1416, 2010.
- [16] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views," CVIU, vol.109, no.2, pp.146–162, 2008.
- [17] B. Prosser, S. Gong, and T. Xiang, "Multi-camera matching using bi-directional cumulative brightness transfer functions," Proc. BMVC, pp.1–6, 2008.
- [18] A. Gilbert and R. Bowden, "Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity," Proc. ECCV, pp.125–136, 2006.
- [19] B. Song and A.K. Roy-Chowdhury, "Stochastic adaptive tracking in a camera network," Proc. ICCV, pp.1–8, 2007.
- [20] L. Yang, "Distance metric learning: A comprehensive survey," Technical Report, Michigan State University, 2006.
- [21] K.Q. Weinberger and L.K. Saul, "Distance metric learning for large margin nearest neighbor classification," JMLR, vol.10, pp.207–244, 2009.
- [22] B. Kulis, "ICML 2010 tutorial on metric learning," Proc. ICML, 2010.
- [23] L. Torresani and K.-C. Lee, "Large margin component analysis," NIPS, pp.1385–1392, 2007.
- [24] J.D.M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," Proc. ICML, pp.713–719, 2005.
- [25] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," Proc. ICCV, pp.1458–1465, 2005.
- [26] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the Nyström method," IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.2, pp.214–225, 2004.
- [27] T. Jebara and R. Kondor, "Bhattacharyya and expected likelihood kernels," Proc. Comp. Learn. Theory, pp.57–71, 2003.
- [28] D.M. Endres and J.E. Schindelin, "A new metric for probability distributions," IEEE Trans. Inf. Theory, vol.49, no.7, pp.1858–1860, 2003.
- [29] A. Chan, N. Vasconcelos, and P. Moreno, "A family of probabilistic kernels based on information divergence," Technical Report, Univ. of California San Diego, 2004.
- [30] X. Huang, S.Z. Li, and Y. Wang, "Jensen-Shannon boosting learning for object recognition," Proc. CVPR, pp.144–149, 2005.

(平成 23 年 7 月 13 日受付, 11 月 23 日再受付)



井尻 善久 (学生会員)

平 12 京都工繊大・工学・機械シス卒, 平 14 同大大学院修士課程了, 同年オムロン(株)入社. 平 21 より名大・情・メディア博士後期課程にも所属. 主に人・顔画像処理の研究に従事. 平 21 SSII 高木賞受賞. 平 21 本会 PRMU 研究会研究奨励賞受賞.



勞 世竝

1984 中国浙江大・電気工学卒. 昭 60～平 4 京大・工・電気留学. 同年オムロン(株)入社. 平 23 よりオムロンソーシアルソリューションズ(株)に異動. 人工知能, 画像処理, パターン認識等の研究開発に従事. 平 21 SSII 高木賞受賞.



村瀬 洋 (正員:フェロー)

昭 53 名大・工・電気卒. 昭 55 同大大学院修士課程了. 同年日本電信電話公社(現 NTT)入社. 平 4 から 1 年間米国コロムビア大客員研究員. 平 15 から名古屋大学大学院情報科学研究科教授. 文字・図形認識, コンピュータビジョン, マルチメディア認識の研究に従事. 昭 60 本会学術奨励賞, 平 6 IEEE-CVPR 最優秀論文賞, 平 7 情報処理学会山下記念研究賞, 平 8 IEEE-ICRA 最優秀ビデオ賞, 平 13 高柳記念奨励賞, 平 13 本会ソサイエティ論文賞, 平 14 本会業績賞, 平 15 文部科学大臣賞, 平 16 IEEE Trans. MM 論文賞, 他受賞. IEEE フェロー, 情報処理学会会員.