

# カメラ台数が多いときに有効な 非線形距離指標学習に基づく複数カメラ間人物トラッキング

井尻 善久<sup>†,††</sup> Shihong Lao<sup>†</sup> 村瀬 洋<sup>††</sup>

<sup>†</sup> オムロン株式会社技術本部コアテクノロジーセンタ, 〒 619-0283 京都府木津川市木津川台 9 丁目 1 番

<sup>††</sup> 名古屋大学大学院情報科学研究科, 〒 464-8603 愛知県名古屋市千種区不老町

E-mail: <sup>†</sup>{joyport,lao}@ari.ncl.omron.co.jp, <sup>††</sup>murase@is.nagoya-u.ac.jp

**あらまし** 人物トラッキングは監視カメラの重要なアプリケーションの一つである。特に、一台の監視カメラにおける視野は限られていることから、複数のカメラによるトラッキングが必要となる。一方、カメラの台数が増加しても完全に死角のない領域を広く渡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。従って本論文では、複数の視野を共有しないカメラにより人物照合を行い、トラッキングをすることを考える。複数の視野を共有しないカメラにおいては、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、服装の色ヒストグラムを抽出し人物間照合する方法が提案されてきた。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に照合すると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うけにくい照合を目指す。また、高次元非線形空間への写像には、Jensen-Shannon カーネルの利用し、これを距離指標学習と組み合わせることにより、より高い精度が実現できることを実験により示す。

**キーワード** カメラ間人物照合 距離指標学習 Jensen-Shannon カーネル

## 1. はじめに

安心安全な社会の実現に向けて、大量の監視カメラが設置されるようになってきており、その監視対象の多くが人である。これに伴い、大量の監視カメラによ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。中でも人物トラッキングは、人物監視の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トラッキングし続けるためには、複数の監視カメラによるトラッキングが実用上重要となる。

一方、カメラの台数が増えても、死角のない監視エリアを構築するためには、相当な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。従って一般的には、視野を共有しない複数の監視カメラ間で、人物照合を行いトラッキングする必要性が生じる。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トラッキング問題を考える。

カメラ間人物トラッキングにより、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、被疑者の写真を元に場内監視カメラ映像と照合を行い、検挙等につなげることができる。また犯罪捜査等においては、犯人等の足取りを知るのにも有効である。

カメラ間人物トラッキングを実現するためには、既存の顔認識アルゴリズム [1] を用いてカメラ間人物照合を行うことが考えられる。しかしながら、図 1 に見られるように、一般的な監視条件下では、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。また監視



図 1 VIPeR データセットの一例; (a)(b) は ID 019 の人物から得られた登録用画像と照合用画像、以下同様に (c)(d) は ID 302 から、(e) (f) は ID 188 から、(g)(h) は ID 358 から得られた画像。

カメラは、シーン全体を撮影することに主眼をおいており、必ずしも焦点がトラッキング対象の人物に合わせられているとは限らない。従って、人物の顔がいつも鮮明に見えることを想定することはできない。こうした理由により、実用的な条件でカメラ間人物照合をするのに、顔認識だけでは不十分であることが多い。

従って本論文では、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、上記の多くのアプリケーションにおいて服装色は有効である。また多くの場合、服装の色は前後左右で同じ色が含まれることを想定できるので、様々な方向を向いた人物を照合するのに適している。

一方で、パターン認識問題として考えたときに、カメラ間の人物照合は困難な課題である。異なる監視カメラにおいては、各カメラの色校正が同じであることは仮定できない。加えて、撮影している環境の照明条件等はカメラ毎また時間帯毎に異なる。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。さらに、これらが複合条件として生じることを考えると、特に困難な課題であると言える。

こうした課題を解決するために、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) カメラ間の関係の陰的モデリングによる手法 [2]~[5] の3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 [6]~[20] 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。こうした冗長な特徴抽出により、単一の情報に依存するよりも高精度な個人同定が可能となるのは当然であるが、過度に多くの特徴を抽出することは処理速度の低下につながり、最適な組み合わせを選ぶための最適化が必要である。このために Gray ら [13] は、様々な特徴量を用いる弱識別器を Adaboost により選択することで、最適な特徴量の組み合わせを学習する方法を提案している。一方、様々な特徴の組み合わせが提案されている中で、服装の色による特徴は、人体の方向（もしくは視点）が変わっても、ある程度不変の特徴であることから、ほとんどの手法において用いられている。また従来手法においては、様々な色空間上で抽出された色ヒストグラムは、単純なヒストグラム間距離により照合される。しかしながら、様々なカメラにより撮影された画像では色校正が異なるので、そこから得られたヒストグラムの単純な距離はそうした変動の影響を受けやすく精度低下につながると考えられる。そこで、カメラ間の色校正の対応関係を推定することにより、高精度な照合を行う方法が提案されて

きた。(ii) のカメラ間輝度伝達関数 (Brightness Transfer Function) の推定方法 [2], [12], [21], [22] はその代表例である。ただし、これらの手法では2台からなるカメラの組合せ一つ一つについてその色ヒストグラム間の関係を表す輝度伝達関数を求める必要がある。輝度伝達関数は同じ色の物体がどのような色ヒストグラムとして観測されるかに基づいているので、これを求めるためには、実際に用いるカメラで一定色のカラーパネル等を撮影するか、同一（同じ色の服を装着した）人物を各カメラで撮影する等が必要であり、カメラ台数が増えたときには全てのカメラについてこれを行うのは手間がかかる。実際、カメラペアはカメラの台数  $N$  に対して  $N(N-1)/2$  に増大するので、カメラ台数が多い場合には現実的とは言えない。従って、カメラの台数が多いときには、カメラのペアに関係なく有効なカメラ間のヒストグラム照合手段が必要であると考えられる。

これに対し本論文では、特定のカメラの対を想定することなく、複数のカメラに写った人物の画像と人物ラベルを数百枚分用意してそれら複数カメラ間で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、実環境上で周辺光や対象自身の姿勢変動、カメラ間の色校正の違いにロバストである。また学習において、カメラ間の配置が既知であることを前提とせず、与えられた色ヒストグラム間の照合精度が最も高くなるような距離指標を学習することができるので、カメラ台数が多いときに特に柔軟性が高いフレームワークである。さらに、距離指標の学習においては、非線形なカーネル写像を用いるので線形写像で表現できないような距離指標を学習することができる。

本論文では、こうした特徴を備えたフレームワークとして距離指標学習の一つである Large Margin Component Analysis(LMCA) と Jensen-Shannon カーネル関数の組み合わせの有効性を実験により示す。なお提案手法は、人物の服装色という単一の情報を用いた照合における基本的な精度を向上させるものであり、上記に総括した多くの方法と補完的な関係にあるので、それらと組み合わせた精度向上を望めるという構造になっている。

従って、本研究の貢献は次のとおりである。人体の服装色による照合問題において、(a) カメラ台数が多いときに、その対応関係を意識することなく、従来の色ヒストグラム間距離に基づく照合より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、図2左部分に示した。従来手法の最も基本的な照合方法(上段)では、モデル登録用人物と照合対象人物から各色成分の色ヒストグラムを抽出しそれらを単純な距離により照合しそれら照合スコアを統合する。一方、提案手法(下段)では、色

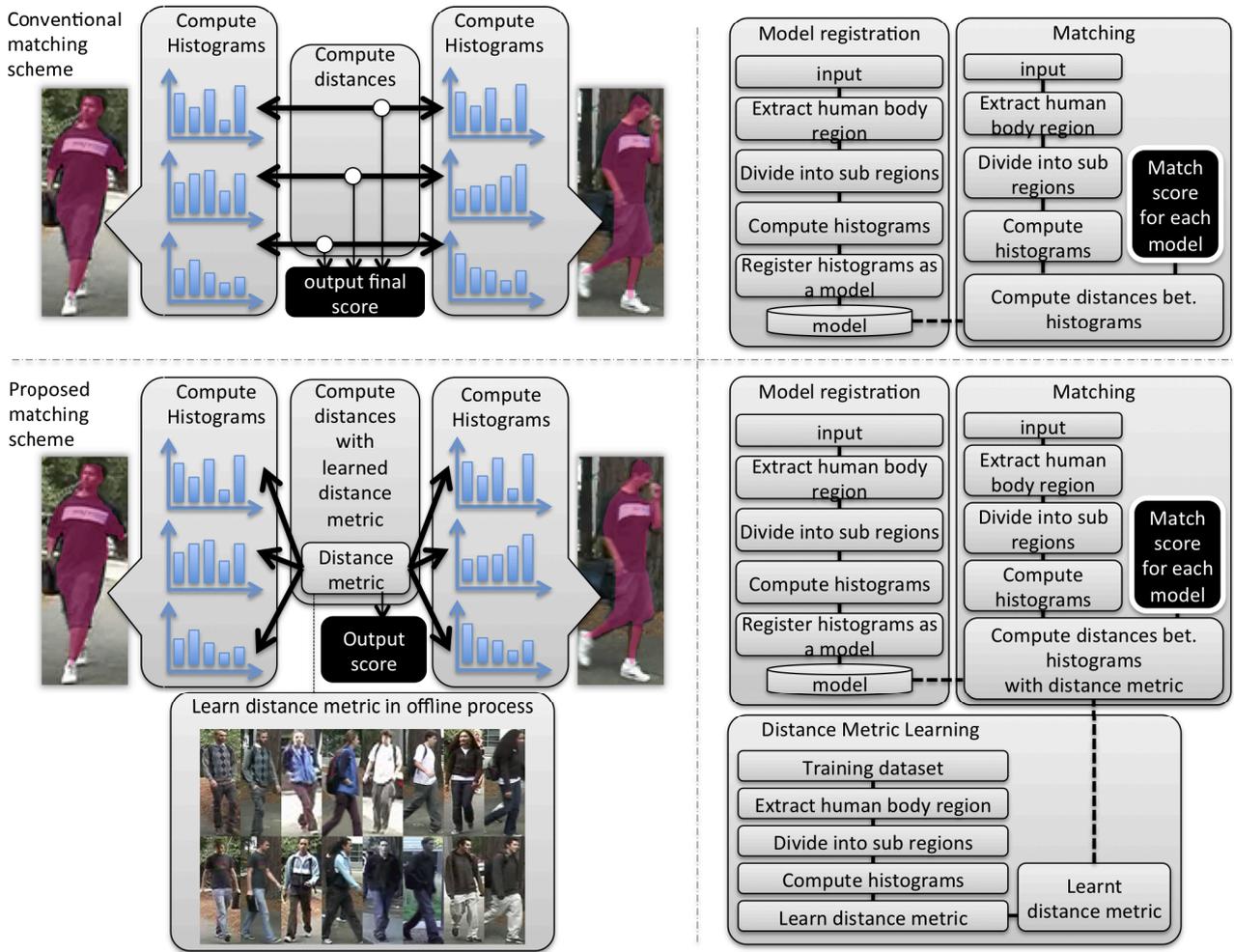


図2 従来手法(上段)と提案手法(下段)の比較:(左)基本的な考え方,(右)フロー

ヒストグラム抽出後、予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、良い距離指標を選ぶことで線形な距離指標を用いる場合に比べ柔軟かつ様々な変動にロバストな距離指標にすることができる。

本論文の以降の章は次のように構成される。まず、2章において提案手法について説明する。この中で、用いる距離指標学習とカーネル関数を説明する。次いで、3章において、提案手法の効果を実証するための実験を行う。最後に4章で結論づける。

## 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、特定されていることを仮定する。また学習のために、様々なカメラおよび様々な撮影条件で撮影された学習画像とそれらに含まれる人物のラベル情報が与えられる事を前提とする。前述のとおり、学習画像に含まれる様々な条件において、各画像において色の校正が異なるために、そうした画像に含まれる人物の色ヒストグラムを直接用いて距離(もしくは類似度)を計算すると精度が低下すると考えられる。そこで、様々なカメラお

よび様々な撮影条件で撮影された学習データを用いて、色ヒストグラム間の最適な距離指標を学習することを考える。提案手法の大まかな流れは、従来手法との比較と共に、図2右半分に示した。まず、距離指標学習時(図2右下 Distance Metric Learning 部分)においては、学習画像から人物領域を抽出し色ヒストグラム特徴量  $X = \{x_i; i = 1, \dots, n\}$  を抽出する。次いで、学習用色ヒストグラム特徴  $X$  およびラベル情報  $Y$  を用いて、これらの人物領域から得られた色ヒストグラムを照合するのに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う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、カメラの組み合わせ等によらず、最適な照合を行うことができると考えられる。登録時(図2右下 Model registration 部分)には、監視カメラに人物が撮影されるたびに登録していく。この際には、各人物領域から学習時と同様の方法で色ヒストグラム  $m_c$  を抽出する。ここで、登録された人物は  $C$  人であったとすると、全体で  $M = \{m_c; c = 1, \dots, C\}$  のモデル特徴量が存在することになる。一方、照合時(図2右下 Matching 部分)には、照合用に入力された人物領域に対して、学習時や登録時と同様の処理を行い、照合用の色ヒストグ

ラム特徴  $t_k$  を得る。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る。最後に、これら照合対象の色ヒストグラム  $T$  とモデル登録された色ヒストグラム  $M$  とを、予め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 (Support Vector Machine: SVM) 等で利用されるマージン最適化規範を用いて最適化を行う。最適化を行う際に、さらに高い精度を得るために、高次元非線形写像  $\phi(x_i)$  により変換を行う。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幾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 服装色ヒストグラム

人物の服装色照合に関する従来研究においては、RGB だけでなく HSV, Lab, YCbCr 色空間における、各チャンネル独立のヒストグラムを連結したものとジョイントヒストグラムが利用されている。本論文における提案手法は、理論的にはどのようなヒストグラムであっても適用できるが、実験においては HSV 色空間における HS ジョイントヒストグラムが低次元で高精度であったのでこれを利用した。また、様々な方向を向く人物の服装色を記述するのに、服装色が前後左右であり変化がないことを仮定すると、上下の服装色の組み合わせ等、服装色の垂直方向の組み合わせが、人物を特徴付ける情報と考えられる。このため、垂直方向に領域を分割し各領域において色ヒストグラムを計算する方法が提案されている [15]。従って本論文においても、人物領域を垂直方向に  $P$  分割して得られた各領域  $p, (p = 1, \dots, P)$  に対して HS ジョイントヒストグラム  $x_{ip} \in \mathcal{R}[0, 1]^{b_h \times b_s}$  を計算した。ただし、 $b_h, b_s$  はそれぞれ H, S 成分のヒストグラムを計算する際のビン数である。また HS ジョイントヒストグラムは通常 2 次元であるがそれをベクトルに並べ変えて利用した。次いで、人物領域全体についてこれらを連結した  $x_i = \{x_{i1}, \dots, x_{iP}\} \in \mathcal{R}[0, 1]^D$  を  $\sum_i x_i = 1$  となるように正規化したものを特徴量として利用した。ただし  $D = b_h \times b_s \times P$  である。これを、学習用、登録用および照合用の画像に対して行い、 $X, M$  および  $T$  を得る。

この手法は、服装色を特徴とする単純な手法の一例であるが、以降の記述では  $[0, 1]$  の値を取る正規化ヒストグラム特徴全般を対象として議論を進める。

## 2.2 距離指標の学習

距離指標の学習においては、教師無し学習法に加え、多くの教師付き学習法が提案されてきた [23]。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が、クラス内散布行列を求めるのに十分な数のデータ (各人物あたりの学習データ) がない場合や、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のランクが落ちるので十分な数の判別軸が取れない等、適用に制限が大きい。一方、線形な距離指標に加え、さらに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 があるが、基本的に二クラス識別のための手法であるので、本論文で対象とするのは複数の人物であり、これらを見分けなければならない他クラス識別問題に簡単には適用できない。こうした二値識別器を他クラス識別問題に適用するには、しばしば、同じラベルを持つペアの照合スコアと、異なるラベルを持つペア照合スコアとを分離する、二クラス問題に帰着する。しかしながら、このような問題設定で学習する際には、しばしば、同じラベルを持つデータ数に比べ異なるラベルを持つデータ数が極端に多くなり、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、大量の異なるラベルの中から、適当なサンプリングを行い、このようにして抽出された固定的な部分集合を用いたとしても、必ずしも学習に適したサンプルをサンプリングできているとは限らず、過適合する可能性が高くなる問題が生じる。一方、SVM のようにマージン最適化規範で他クラス間距離指標を最適化しながらもこうした問題を回避するのに、Large Margin Nearest Neighbor (LMNN) [24] が提案されている。この手法では、マージン最適化規範に基づきマハラノビス距離が Semi-definite problem を数値的に解くことにより最適化される。この方法は最近傍法に基づく手法であるので、本質的に他クラス識別問題であり、前述のような問題を生じない。一方、Torresani らはさらにこの方法を改良し、Large Margin Component Analysis (LMCA) [25] を提案した。これにより、入力データの次元数が多い時にも簡単に計算できるようになり、またカーネル法に基づく高次元・非線形な特徴空間における距離指標学習により高精度化が可能となった。従って、本論文ではマージン最適化規範に基づく距離指標を学習するのに、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:

色ヒストグラム特徴量  $X = \{x_i; i = 1, \dots, n\} \in \mathcal{R}^{D \times n}$  および対応する学習データのラベル  $Y = \{y_i; i = 1, \dots, n\} \in \{0, 1\}^n$  が与えられた時、LMCA は次の損失関数  $\epsilon(L)$  を最小化する。

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

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  $x_j$  が  $x_i$  が同じラベルを持ち ( $y_i = y_j$ ) かつ  $k$ -近傍である時に 1 を取る二値変数であり,  $c > 0$  は適当なバランスパラメータ,  $y_{il} \in \{0, 1\}$  は  $y_i = y_l$  の時に 1 をとる二値変数である. 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, 最初の項は同じラベルを持つデータ同士の距離を最小化し, 不変性 (invariance) を高める効果を持つ. もう一つの項は, 異なるクラスラベルを持つデータ同士 ( $x_l$  と  $x_i$ ) の距離と同じクラスラベルを持つデータ同士の距離の差が距離 1 のマージンを持って分離されるようにする項であり, 分離性 (discriminability) を高める効果を持つ. また, ヒンジ関数は, 十分なマージンを持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のみロスを生じさせる効果を持つ. これにより, 常に学習に適したデータが再サンプリングされ, 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. 最適化はこのロス関数の勾配を導出し勾配法を構築することで簡単に行え, これにより最適な線形写像  $L^\circ$  が得られ, 距離指標は  $L^{\circ T} L^\circ$  で与えられる.

#### 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  $\phi$  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. ここで, 上記線形 LMCA における  $L$  として  $L = \Omega \Phi$  の形をとる  $L$  を考えると, 次のような勾配法を構築することができる.

$$\Omega \leftarrow (\Omega - \lambda \Gamma), \quad (2)$$

$$\begin{aligned} \Gamma = 2\Omega \sum_{ij} \eta_{ij} [E_i^{(k_i - k_j)} - E_j^{(k_i - k_j)}] \\ + 2c\Omega \sum_{ijl} \eta_{ij} (1 - y_{il}) h'(s_{ijl}) \\ \cdot [E_i^{(k_i - k_j)} - E_j^{(k_i - k_j)} - E_i^{(k_i - k_l)} + E_l^{(k_i - k_l)}], \quad (3) \end{aligned}$$

$$s_{ijl} = \|\Omega(k_i - k_j)\|^2 - \|\Omega(k_i - k_l)\|^2 + 1 \quad (4)$$

ただし,  $E_i^v = [0, \dots, 0, v, 0, \dots, 0]$  は  $i$  番目の列のみ値  $v$  を持つ  $n \times n$  の行列であり,  $h'(s)$  はヒンジ関数  $h(s)$  の微分であるが, この関数の  $s = 0$  における不連続性を避けるために, 近似関数として滑らかなヒンジ関数を利用することが提案されている [26]. また, 仮定により, 任意のデータ  $x_q$  の写像は

$$L\phi_q = \Omega\Phi\phi_q = \Omega k_q \quad (5)$$

であるので, カーネル関数を計算することにより簡単に得られる. この際に距離指標は,  $k_q^T \Omega^T \Omega k_q$  となる.

## 2.3 カーネル関数

本節においては, 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  $a, b \in \mathcal{R}[0, 1]^D$  に有効と考えられるカーネル関数  $K(a, b)$  について考察する. 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.

### 2.3.1 従来研究における一般的なカーネル関数

#### Histogram Intersection カーネル:

分布間の距離の指標としてよく用いられてきたのがヒストグラムの重なりを評価する指標であるが, これをカーネル化した正定値カーネルとして Histogram intersection カーネルは次のように定義される [27], [28].

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

#### $\chi^2$ カーネル:

分布間の距離を測る指標として,  $\chi^2$  統計量もしばしば用いられるが, これに基づく正定値カーネルとして次の  $\chi^2$  カーネルが提案されている [29], [30].

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

#### Bhattacharyya カーネル:

分布間の距離として一般的である Hellinger 距離  $\frac{1}{2} \sum_i (\sqrt{a_i} - \sqrt{b_i})^2$  を考えると, 本論文では正規化色ヒストグラム ( $\sum_i a_i = 1$ ,  $\sum_i b_i = 1$ ) を用いるので,  $1 - \sum_i \sqrt{a_i b_i}$  となる. ここで  $\sum_i \sqrt{a_i b_i}$  は Bhattacharyya 係数と呼ばれ類似度の指標であるが, これをカーネル化した次のような正定値カーネルが提案されている [31].

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

### 2.3.2 情報理論に基づくカーネル関数

#### Jeffery divergence カーネル:

本論文における人物の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の照合が必要となる. また, 対象が大きく変動する対象であることから, 精密な空間的な照合を行うことはできない. 従って, 空間的な性質に大きく依存しない何らかの指標が必要であると考えられる. 本論文では, 特に色ヒストグラムを, 人の服装に関する観測情報として捉えたときに, その観測値が持つ情報量の差によりとらえることを提案する. 情報量の差を考える上で, Kullback-Leibler divergence (KLD) は一般的であるが,  $H(a||b) \neq H(b||a)$  であるため (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. 従って, これを直接カーネル関数として用いることはできない. そこで, 対称性を満たす Jeffrey divergence (JD)  $H(a||b) + H(b||a)$  を用い, さらにこの指数  $\exp(-JD/\sigma^2)$  をとったものを新たに Jeffrey

divergence カーネルとして定義する.

$$k(a, b) = \exp \left\{ \frac{-\sum_i \left( a_i \log \frac{a_i}{b_i} + b_i \log \frac{b_i}{a_i} \right)}{\sigma^2} \right\}.$$

このカーネル関数の正定値性は、実験によれば、 $\sigma$  の値に依存していたが、適当なパラメータを選ぶことにより正定値カーネルとなることが確認できた。従って、実験においては正定値カーネルとなるようなパラメータを選んだ。

#### 提案手法：Jensen-Shannon カーネル:

KLD は対称性を満たさないが、これを解決するため JD 以外にも Jensen-Shannon divergence (JSD) [32] が提案されている。JSD は距離の公理を満たし、情報量的な距離を評価するのに適していることから、様々なパターン認識の応用技術で用いられ有効性が示されている。これを利用した Jensen-Shannon divergence カーネルが提案されている [33].

$$k(a, b) = \exp \left\{ \frac{-\sum_i \left( \frac{a_i}{2} \log \frac{2a_i}{a_i+b_i} + \frac{b_i}{2} \log \frac{2b_i}{a_i+b_i} \right)}{\sigma^2} \right\}.$$

このカーネル関数は正定値カーネルであることも証明されている [33]. このように情報理論に基づく距離指標において、Jensen Shannon カーネルは最も理想的であると考えられる。一方でこのカーネル関数の有効性はあまり実証されていないのが実情である。本論文においては、色ヒストグラムによる人物照合において、特にこのカーネル関数と前述のカーネル距離指標学習の組み合わせが有効であることを示し提案する。

## 2.4 照 合

モデル特徴量と照合対象特徴量は、前節のカーネル関数  $K(a, b)$  を用いて、式 5 により、それぞれ  $\Omega K(M, X)$ ,  $\Omega K(T, X)$  として得られる。 $\Omega$  は、式 2 により学習される。これを用いて照合は次のように行う。

$$s(i, j) = f(\Omega K(M, X), \Omega K(T, X)) \quad (6)$$

ただし、 $f(\cdot)$  は任意の距離 (相違度) 関数もしくは類似度関数を表す。

$$\omega(j) = \min_i (or) \max_i (s(i, j)) \quad (7)$$

ここで、距離計算や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 (Viewpoint Invariant PEdestrian Recognition) データセット [13] を用いて実験を行う。このデータセッ

トはカメラ視点もしくは人物の向きに関係ない人物照合アルゴリズムを評価するために作られたデータセットである。このデータセット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、別々のカメラにおいて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。これらの対において、照合対象の体の向きが大きく異なり、それに応じて見え方も大きく異なることがわかる。特に (e)(f) の例等を見ると、体の向き変化により衣服の色も大きく変化している。(c)(d) や (g)(h) の例を見ると、照明条件が大きく異なるために、服装色の色も大きく異なって見える。実用においては、それぞれのカメラが設置される場所は異なるので、このように照明条件が異なることはしばしばである。また、このデータセットにおいては、カメララベルが入っていないために、どのカメラでどのデータが撮影されたのか特定することはできない。冒頭の輝度伝達関数に関する説明で簡単に述べたように、現状の技術ではカメラの台数が増えたときにカメラ特定情報を利用するのは難しい。従って、カメララベルが入っていない状況というのは、実用上においてはカメラ台数が多いときに対応していると考えることができる。

このデータセットには、モデル登録用および照合用として、それぞれ 628 人分の各人物あたり 1 枚のデータが含まれている。従って、このデータセットに含まれる画像枚数は合計 1256 枚である。この 628 人のデータから 200 人分をランダムに選び出して学習用とし、残りの 428 人分を照合用とした。従って、学習画像枚数は、200 人分それぞれにつき登録用と照合用があるので、合計 400 枚である。全ての実験において、10 回のクロスバリデーションを行い、平均および標準偏差を算出した。また、色ヒストグラム間の照合の方法による差を正確に比較するために、すべてのデータについて人体領域のセグメンテーションを人為的に行なった。実際の問題においては、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ことから、実用的な観点から見てもこれは妥当な設定と考えられる。

特徴抽出には、前章で書いたように、様々な論文において有効であることが報告されている HSV 色空間における HS 成分の joint ヒストグラムを正規化して利用した。HS 成分それぞれについてそれぞれビン数は 5 とした。局所領域分割としては図 3 の青線で示したような 8 領域分割を行い、その各領域で赤い色で示した人体領域に含まれる画素から色ヒストグラムを抽出した。このようにして得られた各領域における色ヒストグラムを連結して、合計  $5 \times 5 \times 8 = 200$  次元の特徴量とした。これに基づきカーネル LMCA を学習することで、最適な色ヒストグラム間距離指標を学習した。

評価基準には、Cumulative Match Characteristic (CMC) を用いた。この指標を用いると、正しい照合



図3 (a) 人体領域の分割; (b)(c): 提案手法による成功例 (同一人物が正しく照合されている); (d)(e): 提案手法の失敗例 (別の人物が同一人物として照合されている)

結果がある順位以内に照合される率を知ることができる。従って、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、実用的には上位で照合されることが重要であるので、特に一位照合率(CMC(1))と十位以内照合率(CMC(10))を求め様々な手法を比較した。

このようにして行なった実験の結果を表.1に示した。表中における各数値は平均照合率、±に続く数字は標準偏差を意味している。比較においては、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他、ヒストグラム間重なり(表中 Hist. Intersec.)や Bhattacharyya 距離 (Bhat. dist.) 等、色ヒストグラム間の距離算出に多用されてきた方法を試した。また、LMCAにおいては、線形の LMCA (表中 LMCA with lin.) の他、様々なカーネル関数 (Gauss, Bhattacharyya(Bhat.), Jeffrey(JD), Jensen-Shannon(JS),  $\chi^2$ ) を用いたカーネル LMCA を比較している。この結果を見ると、他の様々な手法と比べ、LMCA と Jensen-Shannon カーネルを用いた手法により高い精度が得られていることがわかる。一方で、一般的に用いられる Gauss カーネルにおいては、単純な距離指標や線形の LMCA より精度が低下しており、色ヒストグラム間の比較には適していないことがわかる。一方、全体の CMC 曲線は、図 4 に示した。視認性の都合上、全てを載せることができなかったが、CMC(1) や CMC(10) だけではなく、全体的に提案手法により高い精度が得られ、実質的に精度向上が得られていることがわかった。

照合結果の幾つかの例を図 3 の (b)-(e) に示した。(b) と (c) は同一人物であり、(d) と (e) は異なる人物である。(b)(c) の対においては、見え方がかなり異なるにも関わらず、正しく照合できている。これは、提案手法の距離指標学習により、効果的にカメラ間の違いが吸収されているからであると考えられる。一方、(d)(e) を識別するのは一見 (b)(c) の例より簡単そうに見えるが、異なる人物が同一人物として誤照合が起こった。(d)(e) の人物領域を観察すると、大まかに言って、服装の色の違いは上半身の白っぽい色と黒っぽい色の違いである。これは、色相・彩度のみからなる色ヒストグラムを、特徴量とし

表 1 CMC

	CMC(1)	CMC(10)
Baseline (Euclidean dist.)	11.4±0.8	31.1±1.5
Baseline (NCC)	13.7±0.7	32.7±1.4
Baseline (hist. intersec.)	13.3±1.0	36.1±1.0
Baseline (Bhat. dist.)	14.7±1.6	38.8±1.6
LMCA with lin.	13.7±0.6	35.8±1.3
LMCA with Gauss	12.5±1.1	39.1±1.2
LMCA with hist. intersec.	13.7±0.6	35.8±1.3
LMCA with $\chi^2$	14.7±0.8	44.0±0.5
LMCA with Bhat.	14.9±1.6	41.8±1.3
LMCA with JD	14.9±0.9	41.0±2.0
<b>LMCA with JS</b>	<b>16.9±1.8</b>	<b>45.5±1.8</b>

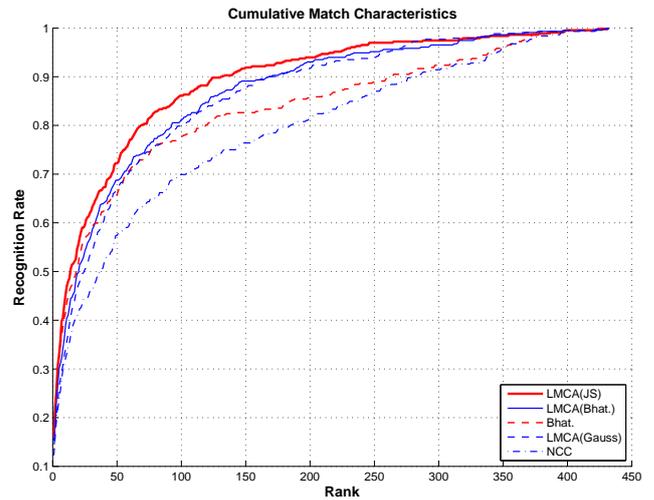


図 4 CMC 曲線：横軸は正しい照合結果が何番目に出現するかを表す、縦軸は認識率、(例えば 10 の際の認識率は、正しい照合結果が 10 位以内に出てくる率を表す)

て用いていることに起因していると思われる。つまり、白と黒の違いは HSV 色空間においては、V 成分に現れるが、V 成分を今回は用いなかったので、これらが判別できなかったのであろうと考えられる。本論文で用いた実験データでは学習データとして用意できる枚数が限られていることを考えると、過学習を防ぐにはこれ以上特徴量の次元数を増やすことができなかった。しかしながら、より良い特徴量を検討することにより、さらに高精度化を狙うことができるとと思われる。

なお、400 枚の学習データを用いて、LMCA の学習に要した時間は、Core2Duo 2.2GHz, Matlab 環境においておよそ 80[sec] であった。従って学習データさえ手に入れば、比較的短時間のうちに最適な距離指標を学習することができることが分かる。

#### 4. 結 論

本論文では、人物の服装色の照合によるカメラ間人物トラッキングにおいて、カメラの配置等を想定せず最適な色ヒストグラム間の非線形距離指標を学習する手法を提案した。また、人物照合においては人物の姿勢が常

に変化し、見え方も大きく変化するので、空間的に正確な照合をすることができないが、観測された色ヒストグラムの持つ情報量の差により比較する Jensen-Shannon カーネルを利用することにより、高精度な照合が行えることを明らかにした。

今後、より高い人物照合精度を実現するために、服装の色のみならず、服装のスタイルや模様、さらには各人物が身につけた所持品や装飾品等から、照合精度を向上させること等が考えられるが、これらは今後の課題である。

**謝辞** 本研究の一部は、独立行政法人 NEDO 次世代ロボット知能化技術開発プロジェクトおよび文部科学省科学技術振興調整費による「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, 村瀬: “実用化に向けた顔画像処理”, 電子情報通信学会技術研究報告 PRMU2010-003, pp. 205–210 (2010).
- [2] O. Javed, K. Shafique, Z. Rasheed and M. Shah: “Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views”, *CVIU*, **109**, 2, pp. 146–162 (2008).
- [3] T. Huang and S. Russell: “Object identification in a Bayesian context”, *Proc. of Joint Conf on AI & IJCAI* (1997).
- [4] H. Pasula, S. J. Russel, M. Ostland and Y. Ritov: “Tracking many objects with many sensors”, *Proc. of IJCAI* (1999).
- [5] B. Song and A. K. Roy-Chowdhury: “Stochastic Adaptive Tracking In A Camera Network”, *Proc. of ICCV* (2007).
- [6] X. Wang, G. Doretto, T. Sebastian, J. Rittscher and P. Tu: “Shape and Appearance Context Modeling”, *Proc. of ICCV* (2007).
- [7] Z. Lin and L. S. Davis: “Learning Pairwise Dissimilarity Profiles for Appearance Recognition in Visual Surveillance”, *Proc. of ISVC* (2008).
- [8] W. R. Schwartz and L. S. Davis: “Learning Discriminative Appearance-Based Models Using Partial Least Squares”, *Proc. of Brazil. Symp. on Comp. Graph. and Image Proc.* (2009).
- [9] A. Alahi, P. Vanderghenst, M. Bierlaire and M. Kunt: “Cascade of descriptors to detect and track objects across any network of cameras”, *CVIU*, **114**, 6, pp. 624–640 (2010).
- [10] S. Bak, E. Corvee, F. Brémond and M. Thonnat: “Person Re-identification Using Spatial Covariance Regions of Human Body Parts”, *Proc. of AVSS* (2010).
- [11] G. Berdugo, O. Soceanu, Y. Moshe, D. Rudoy and I. Dvir: “Object Reidentification in Real World Scenarios Across Multiple Non-overlapping Cameras”, *Proc. of Euro. Sig. Proc. Conf.* (2010).
- [12] B. Prosser, S. Gong and T. Xiang: “Multi-camera Matching under Illumination Change Over Time”, *Proc. of Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications* (2008).
- [13] D. Gray and H. Tao: “Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features”, *Proc. of ECCV* (2008).
- [14] C. Nakajima: “Full-body person recognition system”, *Pattern Recognition*, **36**, 9, pp. 1997–2006 (2003).
- [15] N. D. Bird, O. Masoud, N. P. Papanikolopoulos and A. Isaacs: “Detection of Loitering Individuals in Public Transportation Areas”, *IEEE Trans. on ITS*, **6**, 2, pp. 167–177 (2005).
- [16] 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. of ICSDC* (2008).
- [17] C.-H. Kuo, C. Huang and R. Nevatia: “Inter-camera Association of Multi-target Tracks by On-Line Learned Appearance Affinity Models”, *Proc. of ECCV* (2010).
- [18] C. C. Loy, T. Xiang and S. Gong: “Multi-camera activity correlation analysis”, *Proc. of CVPR* (2009).
- [19] M. Farenzena, L. Bazzani, A. Perina, V. Murino and M. Cristani: “Person Re-Identification by Symmetry-Driven Accumulation of Local Features”, *Proc. of CVPR* (2010).
- [20] L. Bazzani, M. Cristani, A. Perina, M. Farenzena and V. Murino: “Multiple-shot Person Re-identification by HPE signature”, *Proc. of ICPR* (2010).
- [21] B. Prosser, S. Gong and T. Xiang: “Multi-camera Matching using Bi-Directional Cumulative Brightness Transfer Functions”, *Proc. of BMVC* (2008).
- [22] A. Gilbert and R. Bowden: “Tracking Objects Across Cameras by Incrementally Learning Inter-camera Colour Calibration and Patterns of Activity”, *Proc. of ECCV* (2006).
- [23] L. Yang: “Distance metric learning: A comprehensive survey”, *Michigan State Univ. Tech. Report* (2006).
- [24] K. Q. Weinberger and L. K. Saul: “Distance Metric Learning for Large Margin Nearest Neighbor Classification”, *JMLR*, **10**, pp. 207–244 (2009).
- [25] L. Torresani and K.-c. Lee: “Large margin component analysis”, *NIPS* (2007).
- [26] J. D. M. Rennie and N. Srebro: “Fast maximum margin matrix factorization for collaborative prediction”, *In proc. of ICML* (2005).
- [27] F. Odone, A. Barla and A. Verri: “Building kernels from binary strings for image matching”, *IEEE Trans. on Image Proc.*, **14**, 2, pp. 169–180 (2005).
- [28] K. Grauman and T. Darrell: “The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features”, *Proc. of ICCV* (2005).
- [29] J. Zhang, M. Marszalek, S. Lazebnik and C. Schmid: “Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study”, *IJCV*, **73**, 2, pp. 213–238 (2006).
- [30] C. Fowlkes, S. Belongie, F. Chung and J. Malik: “Spectral grouping using the Nyström method”, *IEEE Trans. on PAMI*, **26**, 2, pp. 214–225 (2004).
- [31] T. Jebara and R. Kondor: “Bhattacharyya and Expected Likelihood Kernels”, *Proc. of Comp. Learn. Theory* (2003).
- [32] D. M. Endres and J. E. Schindelin: “A new metric for probability distributions”, *IEEE Trans. on Info. Theory*, **49**, 7, pp. 1858–1860 (2003).
- [33] A. Chan, N. Vasconcelos and P. Moreno: “A Family of Probabilistic Kernels Based on Information Divergence”, *Univ. of California, San Diego, Tech. Report* (2004).