

サーベイ論文：視野を共有しない複数カメラ間での人物照合

井尻 善久[†] 川西 康友^{††} 美濃 導彦^{††} 村瀬 洋^{†††}

[†] オムロン株式会社技術本部コアテクノロジーセンタ 京都府木津川市木津川台9丁目1番

^{††} 京都大学学術情報メディアセンター 京都府京都市左京区吉田二本松町

^{†††} 名古屋大学大学院情報科学研究科 愛知県名古屋市千種区不老町

E-mail: †joyport@ari.ncl.omron.co.jp, ††{kawanishi,minoh}@mm.media.kyoto-u.ac.jp,

†††murase@is.nagoya-u.ac.jp

あらまし 人物監視において、特定の人物を追跡し続けるためには、一台の監視カメラの視野は限られていることから、視野を共有しない複数カメラ間で人物照合を行い、人物追跡することが提案されてきた。本サーベイにおいては、視野を共有しないカメラ間での人物照合の既存手法について総括し、基本的な評価などに関する現状も振り返る。

キーワード サーベイ、複数カメラ間人物照合、カメラ間人物トラッキング

Person Re-identification Algorithms: A Survey

Yoshihisa IJIIR[†], Yasutomo KAWANISHI^{††}, Michihiko MINOH^{††}, and Hiroshi MURASE^{†††}

[†] Corporate R&D, OMRON Corporation 9-1 Kizugawadai, Kizugawa, Kyoto, Japan

^{††} Academic Center for Computing and Media Studies, Kyoto University Yoshida Nihonmatsucho, Sakyoku, Kyoto, Japan

^{†††} Graduate School of Information Science, Nagoya University Furocho, Chigusa, Nagoya, Japan

E-mail: †joyport@ari.ncl.omron.co.jp, ††{kawanishi,minoh}@mm.media.kyoto-u.ac.jp,

†††murase@is.nagoya-u.ac.jp

Abstract In human surveillance, tracking of people is one of essential tasks. Since a field of view in a camera is limited, human tracking across cameras with non-overlapping field of views, i.e., human re-identification have been studied. In this paper, we survey existing work in the field of human re-identification, and summarize evaluation dataset and methodologies as well.

Key words survey, human re-identification, human tracking across cameras

1. はじめに

安心安全な社会の実現に向けて、大量の監視カメラによる人物監視が重要となっている。大量の監視カメラからくる大量の情報を処理するのは、人間の労力の限度を超えることがある。したがって、こうした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、特定の人物を追跡し続ける技術は必要とされている技術の一つである。追跡においては、一台のカメラ視野内の追跡と、複数のカメラによる一台の監視カメラ視野より広い領域における追跡がある。前者に関しては特に多くの研究がなされてきているが、通常施設内など広い領域において監視カメラが利用されることから、後者も実用上重要である。

一方、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。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間で、人物照合を行い追跡する必要性が生じる。本論文で扱うのは、こうした視野を共有しないカメラ間での人物照合問題 (Human re-identification or Human tracking across cameras) である。

したがって、本サーベイにおけるカメラ間での人物照合問題とは、次のように定義される: 「あるカメラで撮影された複数の人物と、別のカメラで撮影された複数の人物が与えられたとき、それら人物間で、1対1対応付けを行うこと。もしくは、対応付けの候補を複数提示すること」

カメラ間人物照合により、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などにおける万引き・窃盗などの際には、被疑者の写真を元に場内監視カメラ映像と照合を行い、検挙などにつなげることができる。また犯罪捜査

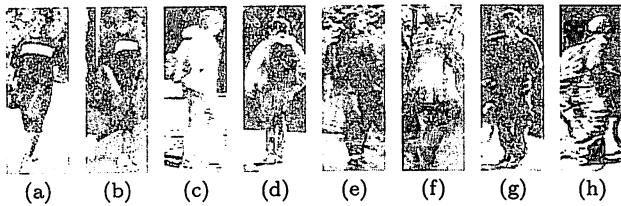


図1 Examples of VIPeR DB: (a)(b)are from ID19. Similarly, (c)(d)from 302, (e)(f)from 188, and (g)(h)from 358.

などにおいては、犯人などの足取りを知るのにも有効である。

類似の問題として、カメラ間で人流を解析し、大まかな人物軌跡を求める問題があるが、必ずしも1対1の人物照合を求める必要があるわけではないので、本サーベイでは取り扱わない。また、別の類似問題として、人物以外を対象とした、カメラ間照合問題が存在する。例えば、対象が車両である場合などである。手法において、カメラ間の人物照合問題と共通する部分が多いと考えられるが、本サーベイでは人以外を対象としたカメラ間照合問題に関する文献の総括は行わない。

カメラ間人物照合を実現するには、顔認識[1]や顔属性[2]、歩容認識[3]の既存手法を利用することも考えられる。顔認識や顔属性を用いた方法は、基本的には顔が正面を向いている必要がある。また歩容認識は、動画像に基づく方法であるので大量カメラを用いたときに計算量が増える。このようにこうした従来手法には課題や制限が大きい。また他のサーベイなどで十分に議論されていることから、本サーベイでは取り扱わない。

既存の人物照合手法は、通常のパターン認識としての解決法と、多数のカメラという問題設定を積極的に利用する解決法とに大きく分類される。前者は主に(i)特徴抽出と(ii)識別手法を工夫する。また後者は、(iii)カメラ間の色校正や(iv)カメラの配置やカメラ間の関係を陰的にモデリングする。上記の(i)-(iv)に基づき既存手法の分類を進める。

本論文の以降の構成は次のとおりである。まず上記分類に従い、2章では見え方に基づく手法を振り返り、特に人物照合において用いられる特徴抽出手法を総括する。3章では識別手法を振り返り、特に学習に基づく手法に焦点を当てる。4章では、カメラ間の色校正に基づく手法を振り返る。5章では、カメラ間の関係の陰的モデリングに基づく手法について総括する。6章においては、既存の評価データおよび評価手法について説明する。最後に7で結論を述べる。

2. 特徴抽出手法：見え方に基づく手法

ここに分類される手法においては、主に服装の色や模様をうまく取り出す特徴量抽出方法が数多く提案されている。ほとんどの手法に共通する特徴は、人の向きや視点変動に対応するために、そうした変動に比較的頑健と思われる服装色の色ヒストグラムを利用する点である。また、様々な変動に対応するため、前記の服装の色と模様の情報に対応する複数の特徴量を冗長に抽出される。

まず、服装の色の特徴を抽出する方法としては、Nakajimaら[4]が、カメラ間の人物照合ではないが人物の全身に見え方を

用いて人物照合する方法を提案している。特徴量としてRGB色空間において各色の和でRGB各チャネルを正規化したものから計算した正規化連結RGヒストグラムを利用している。識別器には、各人物とそれ以外の人物を識別するone vs. all SVMを用いている。同様に赤塚ら[5]は正規化色ヒストグラムを、本田ら[6]はRGB連結ヒストグラムを利用することを提案している。ヒストグラム処理は各色がどの領域で発生したのにかんする空間的情報を無視する。しかしながら、服装の上下の組合せなど空間的な情報も有用であることも多い。このため、Birdら[7]は、HSL色空間において人物領域を縦に10分割した各領域でのメディアン値を利用している。Linら[8]も、正規化RGB色空間、すなわちRGS($S=(R+G+B)/3$)色空間と、Color Rank特徴量(各色チャネルにおいて各値を相対的な輝度順位に置き換えたもの)を用いて、各色空間と高さ方向のジョイントヒストグラムを計算することで、空間的な情報を考慮した色ヒストグラムを用いる事を提案している。また、この手法においては、色の順位情報を用いてカメラによる諧調表現の違いを吸収しようとしているところも特徴的である。一方、別の拡張として、色ヒストグラムは色空間をなど間隔に分割して各ビン毎に頻度を取る処理であるが、Chengら[9]は、画像内から用いられている特徴的な色としてMajor Colorを取り出し、それらをビンとしたヒストグラムを作る(Major Color Spectrum Histogram)。この方法では照合すべき画像内に存在する色のみでヒストグラムを計算するので、RGBなどの色空間でなど間隔に切り出したビンを用いてヒストグラムを計算するより効率的な特徴量表現ができる。類似の方法として、Caiら[10]は、HSV色空間における色相(hue)と、opponent color空間における色比の逆正接により計算されるopponent color angle(ここで色相は彩度が零付近での不安定性を改善するために通常の色相と違う定義がされているが詳細は論文を参照)を、予め学習したコードブックによりヒストグラム化する。先ほどの手法と共通するのは一定の幅のビン上でヒストグラムをとるのではなく生起頻度が高い色をビン(コードブックを構成するコードワード)とするヒストグラムで表すところである。

次いで、服装の模様の特徴を抽出する方法としては、近年一般的である特徴点ベースの照合手法がある。Hamdounら[12]は、特徴点ベースの特徴抽出手法を用いて動画ベースの人物照合手法を提案している。特徴点抽出には、Camelliaライブラリで実装されている、SURFを改良し整数演算のみにした手法を用いて高速な特徴抽出および照合を行っている。特徴点ベースの手法は、対象の勾配情報に敏感であることから服装の模様に関する情報を取り出していると考えられる。同様に、Bakら[13]は、人物領域内におけるHaar-like特徴量のうち判別性の高いものをAdaboostにより選択して識別器を構築することを提案している。また、Wangら[14]は、特徴量としてLogをとったRGB色空間上で方向勾配ヒストグラム(Histogram of Oriented Gradients: HOG)を計算したものと、Lab色空間上でのHOGに対しさらにshape contextを計算したのを用い、これらの共起性を特徴量としたものであるShape and Appearance contextを用いることを提案してい

る。しかしながら、最も直感的な服装模様に着目した特徴抽出手法は、Alahi ら [15] が用いている「共分散」である (Alahi のみならず他手法でも多用されているが、他手法は色情報などと組合せて用いる)。共分散は x, y 方向に関する共起性を見ることから輝度/色や勾配の空間的な配置を表現する。この手法では様々な評価の結果、特に優れている特徴として、テクスチャ x/y 方向の一次勾配と二次勾配/その強度と方向それぞれに関する共分散を用いている。Bak ら [16] は、RGB の各チャンネルの輝度および勾配強度と方向の共分散を抽出する。また共分散行列の類似度 (相違度) としては、比較対象となる二つの共分散行列により定義される一般化固有値問題における固有値を用いる。

服装の色情報のみに基づく手法や、模様情報のみに基づく手法は、計算的観点から言えば効率が良いが、精度的には不十分であることが多い。したがって、これらを組み合わせた手法が数多く提案されてきた。例えば、上村ら [17] は RGB 色空間における RGB ヒストグラムとオートコリログラムを併用することを提案している。また、Gheissari ら [11] は、色情報として HSV 色空間における正規化 HS ヒストグラムを抽出し、watershed アルゴリズムを用いて分割した領域ごとに照合し、模様情報としては特徴点ベースの一手法である Hessian affine 不変特徴量と decomposable triangulated graph により人物の手足などの領域毎に対応付けを行うことを提案している。また、Berdugo ら [18] は、Lin ら [8] を拡張し、正規化色特徴量のような色情報のみならず、服装の模様に関する情報を取り出すために ranked color ratio, ranked oriented gradients, ranked saliency map 用いることを提案している。特に ranked color ratio は上下方向のピクセルの商を特徴としており、服装の模様を抽出するのに効果的であると考えられる。別の手法として、Farenzena ら [19] 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レベルの精度の人物照合精度を実現している。さらにこの手法は、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。他の手法として、Bazzani ら [20] は、全体的な服装色情報と局所的な模様の情報を抽出するため、HSV ヒストグラムおよび局所的な繰り返しパターンとしてエピトメ (epitome) を抽出している。エピトメとは、得られた画像を複数の特徴的なパッチ要素を平均とするガウス混合モデルで表現するものである。一方複数の特徴量を組み合わせる際に、どの特徴をどの程度用いるのかが問題となる。このため、Gray ら [21] は、色情報として RGB, YCbCr, HSV 色空間、および模様情報として輝度に対する Schmid および Gabor フィルタ出力において、Adaboost により人物の判別に役立つ局所特徴を学習し選択することで高精度な照合を可能としている。この手法は局所特徴集合 (Ensemble of Localized Features: ELF) と呼ばれる。興味深いことに、選択された特徴をみると、各色および

模様情報が同程度に選択されており、強いていえば HSV 色空間における H(色相) 情報をもっとも選択されているが、決定的な特徴が無いことが報告されている。したがって前述のような多くの特徴量およびその組合せが提案されていることの裏付けと成っていると考えることができる。同様に、Schwartz ら [22] も、性能向上のために様々な特徴量を用い、部分的最小二乗法 (Partial Least Square: PLS) により次元削減および各特徴量に対する重み付けを行う。用いる特徴量は、服装の模様、エッジ情報、および色情報を捉えるのに適した特徴量を抽出する。服装の模様を捉えるために、人物領域を一定の大きさの重なり合うブロックに分割し各ブロックにおいてテクスチャの解析にしばしば用いられる様々な統計量を計算する。このようにして得られた各ブロックに対する統計量の共起行列 (共分散) を用いることによりテクスチャの繰り返しパターンなどを抽出する。また模様情報のうち特にエッジ情報を捉えるのに HOG を用いる。さらに、色情報として、RGB の各チャンネルにおいて、相対順位を表した Color Rank に変換した色分布においてヒストグラムを計算したものをを用いている。同様の手法として、Kuo ら [23] も、色ヒストグラム、共分散行列、および HOG 特徴量を用いるが、各カメラにおいて得られる動画像から人物をトラッキングし、その一連の動画から同じ人物の集合と、同じ時間に別のカメラにおいて写るペア (絶対に同じ人物でありえないペア) の集合を、それぞれ positive bag および negative bag として、Multiple Instance Learning (MIL) により識別器を構築するところが目新しい。

より最近では時系列情報を用いる方法として、川合ら [24] による ST-HOG を用いた人物照合があるが、この手法はどちらかと言えば歩容認識に近い。

もう一つの異なるパラダイムとしてグループによるコンテキスト情報を用いた人物照合がある。Zheng ら [25] は、特定の個人を同定するのに同時に付近に存在する同行者の特徴を用いる手法を提案している。特徴量として、各局所領域で RGB の各カラーチャンネルにおける SIFT (Scale-Invariant Feature Transform) と正規化平均 RGB ベクトルを連結したのからコードワードヒストグラムを抽出する。さらにこのヒストグラムを用いて人物領域の中心からリング状に分割した各領域間でのヒストグラムの比を算出する。この特徴はグループの左右の位置関係などが変化しても不変性を有する。また、各領域をさらに細かくグリッド状に分割し、各領域においてその内部間での比を求める特徴によりさらに詳細な解析を行なっている。最終的にこうした特徴量を L1 ノルムを用いることで照合する。この手法は、数人からなるグループが入力であることを仮定しているが、実用的にはどのようにグループを切り出すかなどの問題があると思われる。また、グループの組成が変わったときなどには原理的に問題を生じる。Group の同定に関しては、Cai ら [26] が、さらに共分散を用いた手法を提案している。共分散においては色情報や勾配方向や強度を用いるが、特に opponent color を用いているのが特徴である。共分散行列間の距離には、固有値を用いている。同様に芥ら [27] も、人物グループ間対応付けの方法として、特に (他と識別する決定的な情報となりうる) 出

現頻度の低い色に着目した対応付け方法を提案している。

既存の手法を振り返ると、様々な特徴量が用いられている一方で、色ヒストグラムを用いることは基本的である。色空間は様々なものが用いられており、優劣は付けがたい。また、動画を利用するものと静止面を利用するものが存在する。動画を利用する方法においては、単にトラッキングのみを行うもの、クラスタリングなどを行い代表的な特徴を抽出しようとするもの、学習サンプルを増やし特徴選択や識別モデル構築を行うもの、スコアレベルの統合を行うもの、利用方法は様々である。定性的な特徴として、動画を利用するものは高い精度を望める一方で、大量のカメラを用いる場合には転送すべきデータが多くなり計算機負担が高くなる欠点があることが予想される。既存の手法において用いられている色空間、特徴量、距離指標や識別器を、表1にまとめた。表中「入力」の欄には静止面入力(S)か動画入力(V)を示した。また「学習」の欄には、学習ベース(Y)と非学習手法(N)の区別をした。

3. 識別手法

多くの既存手法においては、特徴量抽出に焦点が当てられているために、識別手法には十分な焦点が当てられていないことが多い。そうした既存手法においては、一般的に色照合問題などで用いられている距離指標などが用いられる。こうした事前知識に頼らない手法(非学習手法)は、学習がいらないという手軽さがある一方で、カメラやその色校正などの特性などを考慮した識別基準ではないので比較的精度が低い傾向にある。

非学習手法においては、ユークリッド距離、正規化相関、ヒストグラムの重なり、Bhattacharyya 距離などが用いられている。その例については、表1の距離指標の部分にまとめている。

一方、学習に基づく手法においては、異なるカメラから得られる同一人物の異なる特徴量間の距離を小さくしながら、異なる人物間の距離を大きくするような特徴空間を、事前に学習データセットを用いて最適化しておく。識別には個人特化モデル(Person-specific model)と一般判別モデル(Global/Generic discriminative model)の二つが存在する。

まず個人特化モデルに関しては、前述の Nakajima ら [4] らは SVM により、Schwartz ら [22] の PLS, Kuo ら [23] の MIL, Bak ら [13] の Adaboost がある。より最近では、Hirzer ら [28] が、前章で説明したような共分散に基づく一般的な照合手法を元に、各個人ごとにスコアをソートし類似度が高いものを positive、低いものを negative として、Haar-like 特徴量と RGB 色空間における共分散を特徴量とした adaboost により個人特化モデルを構築することを提案している。これにより一般的な照合手法をさらに強化した個人特化モデルを構築でき高い精度が得られることを報告している。個人特化モデルは、事前に学習するのではなく登録人物が与えられる度に学習することになり、新たな人物が次々と入ってくるような条件においては、実用上負荷が高いという欠点がある。また、one vs. all の他クラス識別器を学習している場合には、新たな人物が登録される度に全登録人物を学習しなければならない課題がある。したがって、照合すべき人物がある程度限られているときなど、用途が

限られてくると考えられる。

これらに対し、前述の Bird ら [7] は、フィッシャー判別分析(FLD)により、クラス内分散とクラス間分散の観点で、最適な一般判別モデルを構築する。FLD の学習には各人物につき比較的多くの枚数の学習データが必要であり、個別の人物の人数も比較的多くなければならぬなどの本質的な制約がある。同様の手法は、本田ら [6] にも見られ、判別に役立つ部分の特徴のみを適応的に取り出そうとする。別の手法として、前述の Gray ら [21] は、様々な色および模様の特徴量から、Adaboost により人物の判別に役立つ局所特徴量の組合せを最適化しておく。この手法は、同一人物間距離(類似度) vs. 異なる人物間距離(類似度)の二値識別により、一般判別モデルを構築する。Adaboost は逐次最適化手法であり、また各弱識別器は単一の局所特徴量を用いる。したがって、人物照合問題のような難しい識別問題に適用した場合には、一定の識別率を持つ弱識別器を確保しにくくなり、同じような局所特徴量のみが選択される可能性が高くなる。

一般に人物照合問題においては、見え方が大きく変化するの本人かどうかに関わりなく全体的に照合スコア低下が起こる。したがって、本質的に本人と他人を切り分ける絶対的マージンを確保するのが困難となる。こうした状況において、固定的識別境界を想定した従来の二値識別器の拘束条件は強すぎ、結果として理想的な識別器を学習できない可能性がある。したがって人物照合のように比較的難しい認識問題においては、拘束条件を緩和し、本人間の距離が他人間の距離より相対的に小さくすることを目指す目的関数を用いることが提案されている。このような相対的な距離関係を学習する手法として、Prosser ら [29] は、アンサンブル RankSVM により、人物照合問題を Ranking 問題に帰着して学習を行うことを提案している。Ranking 学習を行う手法としては、RankBoost と RankSVM が知られていたが、RankBoost においては最適化しやすく全体最適性に課題があるので、特に RankSVM を用いることを提案している。なお実装においては、PR SVM を用いている。これに対し、井尻ら [30] は、人物照合問題を距離指標学習のフレームワークで解くことにより高精度化を狙っている。距離指標学習には Kernel Large Margine Component Analysis を用い、これと Jensen-Shannon カーネル関数を組み合わせることで、非線形な距離指標を学習し、カメラ間の様々な影響を受ける色ヒストグラム間の距離を最適化している。類似の手法として、Zheng ら [31] は、独自の距離指標学習により人物照合問題を定式化し直している。この手法は、本人間の距離が他人間の距離より小さくなる確率を最大化する目的関数から成り立つ。確率は本人間距離と他人間距離との差を元にした、ロジスティック関数により定義されるが、このロス関数が、オーバーフィッティングを防ぐ効果を持っているものと思われ、前述のアンサンブル RankSVM より特に学習データ数が少ない時などに効果があることを報告している。

4. カメラ間色補正

カメラ間人物照合において、色に基づく特徴量は人物の向き

表 1 Variations of features, distance metric, classifiers described in articles

手法	入力	学習	色空間	特徴量	距離指標/識別器	評価 DB
Nakajima ら [4]	S* ³	Y	RGB	正規化 RG ヒストグラム	SVM	Own DB
Bird ら [7]	S	Y	HSL	10 領域メディアン HSL	Euclidean 距離	Own DB
Gheissari ら [11]	S	N	HSV* ⁵	正規化 HS ヒストグラム	Hist. Intersec.	Own DB
Cheng ら [9]	S/V	N	RGB	Major Color ヒストグラム	共通色存在確率	Own DB
Wang ら [14]	S	N	Log-RGB, Lab	HOG, shape context	L1 ノルム	Gheissari DB
Lin ら [8]	S	Y* ²	正規化 RGB(RGS)	Color Rank	KL divergence	Honeywell DB
Gray ら [21]	S	Y	RGB, YCbCr, HSV	輝度, Schmid/Gabor フィルタ	Adaboost	VIPeR
Hamdoun ら [12]	V	N	RGB	(Camellia lib 実装の)SURF	SAD, KD-tree	CAVIAR
Schwartz ら [22]	S	Y	RGB	共分散行列, HOG, ColorRankHist.	PLS	ETHZ
Alahi ら [15]	S	N		共分散行列 (輝度, 勾配, 勾配強度)	固有値	VIPeR 他
Bak ら [16]	S	N	RGB	色, 勾配方向, 勾配強度の共分散行列	固有値	iLIDS
Berdugo ら [18]	V	N	正規化 RGB(RGS)	ColorRank, RankedColorRatio, RankedOrientedGradients, RankedSaliencyMap	KL-divergence	Own DB
Kuo ら [23]	V	Y	RGB?	色ヒストグラム, 共分散行列, HOG	MIL	Own DB
Farenzena ら [19]	S/V	N	HSV	加重色ヒストグラム, MSCR, RHSP	Bhat. 距離	VIPeR, iLIDS, ETHZ
Bazzani ら [20]	V	N	HSV	加重色ヒストグラム, Epitome	Bhat. 距離	ETHZ, iLIDS
Bak ら [13]	S* ³	Y	RGB?	Haar-like 特徴量	Adaboost	CAVIAR, TRECVID
Cai ら [10]	S	Y* ⁴	HSV* ⁵	量子化 Hue/OpponentAngle 空間分布	χ^2 距離	ETHZ
Zheng ら [25]	S	Y* ⁴	RGB	SIFT, 平均 RGB ベクトル	L1 ノルム	iLIDS
Cai ら [26]	S	N	opponent color	色, 勾配方向, 勾配強度の共分散	固有値	CASIA, OULU DB
Hirzer ら [28]	S/V	Y	Lab, RGB	共分散, Haar-like 特徴量	固有値, Aadboost	VIPeR, Own DB

*1 入力は人物検出およびトラッキングが動画ベースであるものを除く。人物照合自体が静止画ベースか動画ベースかを表す。*2 この学習手法は相対 KL を予め登録データ間で計算しておくものである。識別器などを学習しているわけではないが事前に登録データを元にある程度データ間の関係を計算しておく必要があるという意味で著者らは学習と記述している。*3 少なくとも登録プロセスにおいては登録したい人物の複数枚の画像が必要である。*4 コードブックの学習のみ。*5 独自に修正を加えた HSV 色空間

や姿勢に頑健な特徴量として多くの研究で用いられている。色特徴量は人間の直観的にも有用な特徴量であると考えられる。しかし広範囲を複数のカメラを用いて観測する場合、カメラの色校正具合や照明環境の違いに応じて同じ色を観測していても異なった色として観測されてしまうため、特徴量が大きく変動してしまうという欠点がある。この原因としては、カメラの色校正と照明環境の違いによる影響の2つが考えられる。色校正については、多数のカメラを用いて視野重複のない範囲を観測する場合、通常あらかじめ厳密に全てのカメラの色校正を揃えることは困難である。また、照明環境については、暗い廊下に設置されたカメラと明るい室内など、観測している環境に応じて照明環境が異なることが挙げられる。さらに、照明環境の違いは異なるカメラ間だけで生じるものではなく、特に屋外環境においては、1台のカメラ内でも時刻によって照明環境が異なる。こうした問題に対し、色特徴比較にロバスト性をもたせる手法、複数のカメラの間での色の差異を補正する手法によって人物照合の精度を上げる研究が行われている。色比較にロバスト性を持たせる手法としては、Cheng ら [9] および Madden ら [32] による Incremental Major Color Spectrum Histogram Representation(IMCSHR)を用いる方法などがある。以下では、主にカメラ間での色の差異を補正する手法について振り返る。

カメラ間の色の差異を補正するのは、一方のカメラで観測さ

れた映像の輝度分布をもう一方のカメラで観測された映像の輝度分布と対応付ける関数を求めることによって実現されることが多く、こうした関数は輝度伝達関数 (Brightness Transfer Function: BTF) と呼ばれる。BTF には様々な表現方法があり、ガンマ補正やヒストグラム平坦化なども BTF とみなすことができる。

Porikli ら [33] は双方のカメラで得られた色ヒストグラムの関係を求めるため、双方のカメラで得られた色ヒストグラムの相互相関行列を作り、それらのヒストグラム間の BTF を動的計画法により求める方法を提案している。これにより、各カメラペアに対して非線形な変換関数をノンパラメトリックに推定することを実現している。しかし、この変換関数はカメラパラメータや照明環境の変化によって変動してしまうため、それらに変化するような環境では正しく求めることはできない。

Javed ら [34] [35] は、累積ヒストグラム同士の関係をノンパラメトリックに表現している。この論文で BTF は、2台のカメラ間に対して定義され、一方のカメラで観測された輝度値 B_i ともう一方のカメラでの観測値 B_j を対応付ける関数 f_{ij} である。これは次のように定義される。まず、あるカメラに映った物体 O_i の、画素数で正規化した輝度値 B_m の累積ヒストグラム値を $H_i(B_m) = \sum_{k=1}^m I_i(B_k)$ とする。ここで、あるカメラに映る物体 O_i と、別のカメラに映る物体 O_j が同一物体であるとき、その領域中に含まれる輝度値の順序関係は同一であるという仮定よ

り、 O_i 中の輝度値 B_i 以下の領域の割合と O_j 中の輝度値 B_j 以下の領域の割合が等しい時、 $H_i(B_i) = H_j(B_j) = H_j(f_{ij}(B_i))$ から BTF は、 $f_{ij}(B_i) = H_j^{-1}(H_i(B_i))$ となる。ただし、 H_i^{-1} は累積ヒストグラムの逆関数である。この時、ヒストグラムを 256 次元のベクトルとして表現すると、BTF も 256 次元のベクトルで表現することができる。こうして得られる 2 台のカメラで観測された BTF には、人ごとに日照変動などによりばらつきが生じている。そこで、BTF を多数集め、確率的主成分分析を行うことにより BTF の変化の部分空間を求めることでばらつきに対処している。そして、ある 2 枚の人物画像間の BTF と確率的主成分分析により求めた部分空間を比較することで、その 2 枚の人物画像が同一人物か否かを判断する手法を提案している。この手法は照明変化を含むカメラ間においても、うまく部分空間を構築することができていれば頑健な人物照合が可能である。しかし、BTF の部分空間を求めるための訓練集合中に、色空間中に広く分布する人物画像ペアが必要である。また、多くの訓練集合がないと、うまく部分空間を構築することは難しい。

少ないサンプル数でもうまく色ヒストグラムの変換関数を求める手法として、Prosser ら [36] は累積輝度伝達関数 (Cumulative BTF: CBTF) を提案している。各ペアから計算した BTF の部分空間を求める代わりに、全ての訓練集合中の画像を用いて一つの BTF を構築する。N 枚の画像 I_1, I_2, \dots, I_N から計算した累積ヒストグラム $\hat{H}_i(B_m) = \frac{1}{N} \sum_{k=1}^m \sum_{L=1}^N I_L(B_k)$ を用い、CBTF は $cf_{ij}(B_i) = \hat{H}_j^{-1}(\hat{H}_i(B_i))$ と定義できる。これにより、複数人の色情報をもとに広い範囲に分布する色ヒストグラムから BTF を作るができる。

これらの手法は、日照変化を想定していない、もしくは訓練集合に含まれる日照変化には対応することができるが、訓練集合に含まれないような日照変化があった場合対応することはできない。Prosser ら [37] はまた、適応的に CBTF を更新する手法を提案し、日照変化に対処している。画素毎の直近数フレームのメディアンから背景画像を作って現在の照明変化を推定したうえで照明変化前後の BTF により人物の色特徴を補正することで、各カメラ内での大きな照明変化を吸収している。しかし、背景の輝度値変化から BTF を作っているため、背景色が単純な場合は難しいと考えられる。

Gilbert ら [38] は、ひとつのカメラ内で観測した一人の人物軌跡から得られる多数のヒストグラムをメディアン処理したヒストグラムを用い、そのヒストグラムをカメラ間で比較することにより、カメラ間の色変換行列を漸進的に計算する手法を提案している。この手法は、計算コストが大きいということが報告されている。

これまでに取り上げた研究では、シーン全体の照明が一様に変化するようなものを対象としている。照明とシーン中の物体によって、観測画像中に大きな陰影が生じるようなシーン、照明が複雑に反射するようなシーンでは、場所により照明変化の影響が異なるため場所ごとの色変換を考慮することが今後の課題として考えられる。

5. カメラ間の関係モデリング

視野重複のない多数のカメラによって広いシーンを観測する状況において、カメラ間の人物照合では、人物の特徴量を比較するだけでなく、そのシーンに関する情報を導入することで人物照合を精度よく行う研究がある。シーンに関する情報として、どのカメラ間に移動経路が存在するのかといったカメラトポロジがよく用いられている。人物照合においては、これに加え、各移動経路がどのような確率で利用されるかというカメラ間の遷移確率や、経路を移動するのにかかる移動時間分布からなる時空間遷移確率 (spatio-temporal transition probability) が用いられることが多い。本節ではこうしたカメラトポロジを用いた研究について振り返る。

Kettner ら [39] は、ベイジ的な枠組みでカメラ間人物照合を定式化した。予め与えられたカメラ間の移動時間の確率分布を用い、その確率が最大となる人物の組み合わせを人物照合の答えとした。

Javed ら [40] [35] は、複数のカメラで観測した同一人物のペアから、カメラトポロジ及び、カメラ間の遷移確率を Parzen 窓推定で求める手法を提案した。また、人物照合対象である 2 枚の画像の特徴量比較により得られる、そのペアが同一人物である確率に対してカメラ間の遷移確率をかけた確率を最大化することにより、人物照合の精度を上げる定式化を行った。Porikli ら [41] は、観測範囲間に通路が存在するカメラ間にリンクを持つようなベイジアンネットワークでカメラ間の遷移確率をモデル化した。Dick ら [42] は、各映像を複数のブロックへと分割し、すべてのブロック間の遷移確率をモデル化した。この手法では、マーカ物体を持った人物が環境内を歩きまわることにより、遷移確率を学習するための訓練集合を与えた。

前述の手法は、手動で対応付けられた大量の同一人物ペアが教師信号つきの訓練集合として学習時に必要であった。しかし、広範囲における多数のカメラ間での人物照合には、このような訓練集合を用意することが難しい。そこで、どの人物が同一人物であるかという教師信号なしでカメラトポロジを自動的に学習する手法が提案されている。Ellis ら [43]、Makris ら [44] は、カメラの観測範囲とその相対的な位置関係は既知とした上で、カメラ内の人物追跡の開始点と終了点から出入口を推定し、複数カメラ映像の出入口間の対応関係を推定する手法を提案している。Tieu ら [45] は、遷移時間の確率分布を一般化して、複雑な分布形状に対応した。浮田 [46] は、誤対応を含んだ大量の入出データを上手く分類して連結推定を行い、視野間の重なり、経路の有無、経路の使用確率などの確率的表現を得た。

これまでの処理は予め用意した訓練集合を用いて学習をするバッチ処理で合ったが、人物照合を実行しながら逐次的にカメラトポロジや遷移確率を学習する手法が提案されている [47] [48] [49]。野田ら [50] は、逐次学習により、再計算を行うことなくカメラ台数の増減に対処した。Chen ら [51] は出入口のクラスタリング結果と移動時間の確率分布の更新を漸進的に行うことにより、カメラトポロジだけでなくカメラ間の移動時間の推定も同時に行った。

ここで取り上げたカメラトポロジを用いた手法は、一般の人物画像照合、検索に用いることは難しいが、カメラネットワークにおける特定の人物の移動経路抽出などを目的としたカメラ間の人物照合において非常に有用であると考えられる。

6. 評価データ, 評価指標

6.1 評価データ

各手法において用いられているデータセットは、表.1の右欄にまとめたが、このうち公開されているのは、CAVIAR [52], VIPeR [53], iLIDS [54], TRECVID [55], ETHZ [56]である。このうち、最も入手しやすいものは、VIPeR, CAVIAR, ETHZである。ETHZは移動カメラによる人物トラッキングの学習/評価DBであり、本来視野を共有しないカメラ間での人物照合の問題設定にはなっていない欠点がある。また、VIPeRは、各人物について静止画のみが用意されており、動画ベースの手法を評価することはできない欠点がある。また、VIPeRは変動が大きく、様々な変動が含まれているために、一位照合率は最新の手法を用いても20%程度と、かなり難易度が高いDBとなっている。iLIDSとTRECVIDは提供元は同じ(UK Home Office)であり、提供元へ同意書などの提出が必要であり、DBのサイズが大きいためHDDを送付するなどの手間がかかる。

6.2 評価指標

評価指標として最も一般的であるのは、累積照合特性(Cumulative Match Characteristics: CMC)である[57]。CMCは、 n 番目以内に正しい照合結果が候補として抽出される確率を求めるものである。すなわち $CMC(n)$ は次のように定義される。

$$ID(g_i) = ID(p_j), ID(g_i) \neq ID(p_j),$$

$$s_{ij} = f(g_i, p_j),$$

$$CMC(n) = \frac{|\{p_j; |\forall g_i; s_{ij} > s_{ij}| \leq n\}|}{|\{g_i; ID(g_i) = ID(p_j)\}|}.$$

ただし、 $|\cdot|$ は集合の要素数を表し、 $ID(\cdot)$ はラベルを求めるための演算子、 $f(\cdot)$ は何らかの類似度指標もしくは確率などを表している。また、 s_{ij} は同一人物についての登録データ g_i と照合用データ p_j の照合結果のスコアであり、 s_{ij} は異なる人物についての照合スコアである。また、CMCの分子は、各照合データ p_j に対して同一人物についてのスコア s_{ij} より大きな値を与える、異なる人物の登録データ g_i の数(照合スコアが何位にくるのか)が、 n 以内になっている照合データ p_j 数を表している。分母は同一人物のデータ数である。

CMCの曲線において、登録データの数が横軸の限界となる。また、類似度指標ではなく距離指標を用いるときには式.1の分子における $s_{lj} > s_{ij}$ を $s_{lj} < s_{ij}$ のように逆にすればよい。

実用的には上位で照合されることが重要である。例えばアプリケーションにおいて、対応付けの候補を出力する際に、10人程度の候補に絞り込むことが出来れば、あとは人が正しい対応付けを与えるなどが考えられるが、絞りこまれた状況でも100人程度の候補がいるとなると人による対応付けも困難となる。したがって、 $CMC(1)$ や $CMC(10)$ などが切り出した性能指標として用いられることもしばしばある。

7. 結 論

視野を共有しないカメラ間での人物照合は、監視カメラにおける人物同定の総合問題(顔・歩容・服装等)として近年脚光を浴び、特にセキュリティ応用の観点で既に数多くの研究例がある。一方、実応用観点では課題が多く、今後コンピュータビジョンアルゴリズムの進化とその応用が期待される分野である。特に実応用においては、服装に基づく照合自体の精度向上のみならず、顔や歩容といった情報との組合せも期待される。

今後の研究の発展のために既存の研究を整理する必要性に駆られ今回は本総括論文を執筆した。書面の都合上、多くの手法について簡潔な説明にとどまった。より詳細について論文誌等においてさらにページを割いて紹介できれば幸いである。また本総括が今後の研究の踏み台となれば本懐である。

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、および文科省による「安全・安心な社会のための犯罪・テロ対策技術など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”, 信学技報 (PRMU), pp. 205–210 (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] H. Murase and R. Sakai: “Moving object recognition in eigenspace representation: gait analysis and lip reading”, Pattern Recognition Letters, 17, pp. 155–162 (1996).
- [4] C. Nakajima, M. Pontil, B. Heisele and T. Poggio: “Full-body person recognition system”, Pattern Recognition, 36, 9, pp. 1997–2006 (2003).
- [5] 赤塚, 高橋, 井手, 村瀬: “不連続な撮影領域間の人物軌跡の対応付けに関する検討”, 電子情報通信学会総合大会講演論文集, 2006, 2, p. 198 (2006-03-08).
- [6] 本田, 出口, 高橋, 井手, 村瀬: “入退室解析のための色特徴を用いた人物対応付けに関する検討”, 電子情報通信学会総合大会講演論文集, p. 225 (2009-03-04).
- [7] 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).
- [8] Z. Lin and L. S. Davis: “Learning Pairwise Dissimilarity Profiles for Appearance Recognition in Visual Surveillance”, Proc. of ISVC, pp. 23–34 (2008).
- [9] E. D. Cheng and M. Piccardi: “Matching of Objects Moving Across Disjoint Cameras”, Proc. of ICIP, pp. 1769 – 1772 (2006).
- [10] Y. Cai and M. Pietik: “Person Re-identification Based on Global Color Context”, The Tenth International Workshop on Visual Surveillance (2010).
- [11] N. Gheissari, T. B. Sebastian and R. Hartley: “Person Re-identification Using Spatiotemporal Appearance”, Proc. of CVPR, pp. 1528 – 1535 (2006).
- [12] 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 ICDCS, pp. 1 – 6 (2008).
- [13] S. Bak, E. Corvee, F. Brémond and M. Thonnat: “Person Re-identification Using Haar-based and DCD-based Signature”, AMCSS 2010, pp. 1 – 8 (2010).
- [14] X. Wang, G. Doretto, T. Sebastian, J. Rittscher and P. Tu: “Shape and Appearance Context Modeling”, Proc. of ICCV,

- pp. 1–8 (2007).
- [15] 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).
 - [16] S. Bak, E. Corvee, F. Brémond and M. Thonnat: “Person Re-identification Using Spatial Covariance Regions of Human Body Parts”, *Proc. of AVSS*, pp. 435–440 (2010).
 - [17] 上村, 池亀, 下山, 玉木, 山本: “ネットワーク上の複数カメラを用いた実時間人物照合システム”, *信学技報 (PRMU)*, **103**, 659, pp. 67–72 (2004-02-13).
 - [18] 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.*, pp. 1806–1810 (2010).
 - [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*, pp. 2360–2367 (2010).
 - [20] L. Bazzani, M. Cristani, A. Perina, M. Farenzena and V. Murino: “Multiple-shot Person Re-identification by HPE signature”, *Proc. of ICPR*, pp. 1413 – 1416 (2010).
 - [21] D. Gray and H. Tao: “Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features”, *Proc. of ECCV*, pp. 262–275 (2008).
 - [22] 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.*, pp. 322–329 (2009).
 - [23] 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*, pp. 383–396 (2010).
 - [24] 川合, 楨原, 八木: “Sthog 特徴を用いた複数カメラ間での人物照合”, *情報処理学会研究報告 (CVIM)*, **2011**, 10, pp. 1–8 (2011-05-12).
 - [25] W.-S. Zheng, S. Gong and T. Xiang: “Associating Groups of People”, *Proc. of BMVC* (2009).
 - [26] Y. Cai, V. Takala and M. Pietikainen: “Matching groups of people by covariance descriptor”, *ICPR*, pp. 2744–2747 (2010).
 - [27] 芥, 新田, 馬場口: “疎分散カメラ間の人物グループの対応付けによる人流解析”, *信学技報告 PRMU2010-272*, pp. 205–210 (2011).
 - [28] M. Hirzer, C. Beleznaï, P. M. Roth and H. Bischof: “Person Re-identification by Descriptive and Discriminative Classification”, *Scandinavian Conference on Image Analysis*, pp. 91–102 (2011).
 - [29] B. Prosser, S. Zheng, S. Gond and T. Xiang: “Person Re-Identification by Support Vector Ranking”, *Proc. of BMVC*, pp. 1–11 (2010).
 - [30] 井尻, S. Lao, 村瀬: “非線形距離指標学習によるカメラ間人物照合”, *信学技報 (PRMU)*, pp. 139–146 (2011).
 - [31] W.-S. Zheng, S. Gong and T. Xiang: “Person Re-identification by Probabilistic Relative Distance Comparison”, *Proc. of CVPR* (2011).
 - [32] C. Madden, E. D. Cheng and M. Piccardi: “Tracking people across disjoint camera views by an illumination-tolerant appearance representation”, *Machine Vision and Applications*, **18**, 3-4, pp. 233–247 (2007).
 - [33] F. Porikli: “Inter-camera color calibration by correlation model function”, *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, IEEE, pp. II-133–6 (2003).
 - [34] O. Javed, K. Shafique and M. Shah: “Appearance Modeling for Tracking in Multiple Non-Overlapping Cameras”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, Ieee, pp. 26–33 (2005).
 - [35] 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).
 - [36] B. Prosser, S. Gong and T. Xiang: “Multi-camera Matching using Bi-Directional Cumulative Brightness Transfer Functions”, *BMVC* (2008).
 - [37] B. Prosser, S. Gong and T. Xiang: “Multi-camera Matching under Illumination Change Over Time”, *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, pp. 1–12 (2008).
 - [38] A. Gilbert and R. Bowden: “Tracking Objects Across Cameras by Incrementally Learning Inter-camera Colour Calibration and Patterns of Activity”, *9th European Conference on Computer Vision ECCV*, Vol. 3952, Springer, pp. 125–136 (2006).
 - [39] V. Kettner and R. Zabih: “Bayesian multi-camera surveillance”, *Proc. of CVPR*, pp. 253–259 (1999).
 - [40] O. Javed, Z. Rasheed, K. Shafique and M. Shah: “Tracking across multiple cameras with disjoint views”, *Proc. of ICCV*, Vol. 2, Ieee, pp. 952–957 (2003).
 - [41] F. Porikli and A. Divakaran: “Multi-camera calibration, object tracking and query generation”, *International Conference on Multimedia and Expo. ICME ’03. Proceedings (Cat. No.03TH8698)*, pp. I-653 (2003).
 - [42] A. R. Dick and M. J. Brooks: “A Stochastic Approach to Tracking Objects Across Multiple Cameras”, *Australian Conf. on Art. Intel.*, pp. 160–170 (2004).
 - [43] T. J. Ellis, D. Makris and J. K. Black: “Learning a multi-camera topology”, *Proc. of VSPETS*, pp. 165–171 (2003).
 - [44] D. Makris, T. Ellis and J. Black: “Bridging the gaps between cameras”, *Proc. of CVPR*, pp. 205–210 (2004).
 - [45] K. Tieu, G. Dalley and W. L. Grimson: “Inference of Non-Overlapping Camera Network Topology by Measuring Statistical Dependence”, *Proc. of ICCV*, Ieee, pp. 1842–1849 (2005).
 - [46] 浮田: “広域分散カメラ群の確率的連結関係推定法”, *電子情報通信学会論文誌 D*, **J89-D**, 7, pp. 1523–1533 (2006).
 - [47] a. Gilbert and R. Bowden: “Incremental, scalable tracking of objects inter camera”, *Computer Vision and Image Understanding*, **111**, 1, pp. 43–58 (2008).
 - [48] K. Shafique, A. Hakeem, O. Javed and N. Haering: “Self Calibrating Visual Sensor Networks pro fashion . at Tentry sstem 3 Recovering Topology of Overlap-”, *IEEE Workshop Applications of Computer Vision*, pp. 1–6 (2008).
 - [49] X. Wang, K. Tieu and W. E. L. Grimson: “Correspondence-free activity analysis and scene modeling in multiple camera views”, *IEEE trans. on PAMI*, **32**, 1, pp. 56–71 (2010).
 - [50] 野田, 島田, 有田, 谷口: “撮影領域に重ならないカメラ群の逐次の連結関係推定に基づく実時間物体追跡”, *火の国情報シンポジウム* (2007).
 - [51] K.-W. Chen, C.-C. Lai, P.-J. Lee, C.-S. Chen and Y.-P. Hung: “Adaptive Learning for Target Tracking and True Linking Discovering Across Multiple Non-Overlapping Cameras”, *IEEE Trans. on Multimedia*, **13**, 4, pp. 625–638 (2011).
 - [52] “CAVIAR DB” (<http://homepages.inf.ed.ac.uk/rbf/CAVIARDB/>).
 - [53] “VIPeR DB” (<http://users.soe.ucsc.edu/~dgray/VIPeR.v1.0.zip>).
 - [54] “iLids DB” (<http://tna.europarchive.org/20100413151426/scienceandresearch.homeoffice.gov.uk/hosdb/cctv-imaging-technology/i-lids/index.html>).
 - [55] “iLids DB” (http://www.itl.nist.gov/iad/mig/tests/trecvid/2010/index.html#eval_data).
 - [56] “ETHZ DB” (<http://www.vision.ee.ethz.ch/~aess/iccv2007/>).
 - [57] D. Gray, S. Brennan and H. Tao: “Evaluating Appearance Models for Recognition , Reacquisition , and Tracking”, *Proc. of PETS*, Vol. 3, pp. 41–49 (2007).