

サーベイ論文：画像からの歩行者属性認識

川西 康友[†] 新村 文郷^{††} 出口 大輔^{†††} 村瀬 洋[†]

[†] 名古屋大学 大学院 情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町

^{††} 名古屋大学 未来社会創造機構 〒464-8601 愛知県名古屋市千種区不老町

^{†††} 名古屋大学 情報連携統括本部 情報戦略室 〒464-8601 愛知県名古屋市千種区不老町

E-mail: [†]kawanishi@is.nagoya-u.ac.jp

あらまし 防犯カメラや車載カメラによって取得した画像から、そこに写った歩行者が持つ属性を認識することは、歩行者の認識、照合、交通危険予測など様々な用途に有用である。しかし、人の違いや撮影環境の違いによって大きな変動がある歩行者画像から属性を安定して認識することは未だに困難な問題である。本サーベイでは、カメラによる観測から得ることの可能な歩行者の属性に焦点を当て、歩行者の照合や交通危険予測に有用な属性について整理を行い、身体属性、行動属性、及び外見属性に分類する。また、既存手法の整理・紹介だけでなく、評価に用いられているデータセットについての現状についても報告する。

キーワード 歩行者属性、属性認識、サーベイ

Pedestrian Attributes Recognition from a Image: A Survey

Yasutomo KAWANISHI[†], Fumito SHINMURA^{††}, Daisuke DEGUCHI^{†††}, and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University

Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan

^{††} Institute of Innovation for Future Society, Nagoya University

Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan

^{†††} Information Strategy Office, Information & Communications, Nagoya University

Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan

E-mail: [†]kawanishi@is.nagoya-u.ac.jp

Abstract Pedestrian attributes from an image captured by a camera are important keys for various applications such as person recognition, identification, and traffic accident risk prediction. However, it is still a hard task to recognize pedestrian attributes from an image due to large variations of pedestrians. In this survey, we focus on the attributes which can be obtained by a camera, especially, effective attributes for person identification and traffic accident risk prediction. We categorize them into physical, appearance and behavioral attribute groups. We review worth noting methods and also summarize existing datasets.

Key words Pedestrian Attributes, Attributes Recognition, Survey

1. はじめに

近年、画像中の物体認識の研究が盛んになっており、機械学習手法の発展に伴い様々な認識が可能となってきた。特に、物体が何かといった基本的な認識問題よりも、意味的に高度な概念を理解することを目的とする研究が増えてきた。しかし、そうした認識問題のクラスは問題設定によって様々であり、画像から得られる特徴量と各クラスとの対応関係を直接学習することは困難である。

この問題に対し、各クラスを複数の「属性」の組み合わせで

表現すること [1] ~ [3] により、各属性認識によって得られる中間表現を用いて各種認識問題を解くアプローチ [4] ~ [6] が提案されている。

例えば、防犯カメラ映像からの容疑者検索のためには、容疑者がどのような服装をしていたか、何を持っていたかといった属性が有用である [7]。また、高度道路交通システム (Intelligent Transportation Systems: ITS) の分野においては、歩行者が大人か子供か、車の方を見ているか否かといった属性が、車に対する危険度の推定に有用である。このように、画像を一旦「属性」という中間表現に置き換えることにより、より高レベルな

認識問題を解く研究が増えている。

こうしたなかで、どのようにして画像から属性自体を認識するかを扱う研究、得られた属性を元に様々な認識問題をどう解くかを扱う研究が発展してきている。しかし、歩行者の属性認識に関するまとまった資料は筆者の知る限り存在しない。このサーベイにより、この分野全体を見通し、研究の流れをつかむことを目的とする。また、各種研究の評価のために、どのようなデータベースが用いられているのかについて報告する。

本サーベイでは、カメラで観測可能な歩行者の属性を対象とし、その人に対して付与できる情報を属性として扱う。特に、自動運転などを目的とした車載カメラ映像、防犯などを目的とした監視カメラ映像に映る歩行者に関する属性を主な対象とする。

関連する問題として、人物検出 [8] [9]、姿勢推定 [10] [11]、ジェスチャ認識 [12]、行動認識 [13] などがあるが、これらは他のサーベイで十分に議論されているため本サーベイでは扱わない。また、人物以外の物体や一般の画像に対する属性認識も多数の研究 [14] ~ [16] がなされているが、これらについても本サーベイでは対象としない。

本サーベイの構成は以下のとおりである。2. 節では、歩行者に関する属性として、どのようなものが扱われているのかを、身体、外見、行動に関する属性といった分類で整理する。3. 節では、身体属性、外見属性、行動属性ごとに、それぞれに特化した手法を紹介する。4. 節では、複数の属性を複合的に認識する手法の研究事例を紹介する。5. 節では、評価に用いられるデータセットについて紹介する。最後に、6. 節で総括する。

2. 歩行者の属性と本サーベイの方針

本サーベイでは、主に車載カメラや防犯カメラに写った歩行者を対象とする。また歩行者だけでなく、一般的に人に関する属性についても一部扱う。歩行者の様々な属性認識に関する報告が数多くなされており、また属性の種類は非常に多様である。本節では、それらを分類して整理する。

車載カメラや防犯カメラに写った歩行者に対する画像処理の目的としては、人物認証やカメラ間追跡、そして交通安全が代表的である。まずは、これらの応用において利用される属性について述べる。

人物認証は、バイオメトリクス認証に代表されるような、人固有の属性に注目することで、入力された人物画像が誰であるかを判定する。例えば、性別や年齢、身長といった身体固有で変化することが少ないものが注目される。人物照合やカメラ間追跡の分野では、少し制約が緩まり、ある時間を限定した時に人を同定しうる属性に注目する。例えば、服装は、異なる日では変わるが、ある 1 日の中で変化することは稀であるため、人物同定の特徴として利用される。

車載カメラ映像を扱う高度道路交通システム (ITS) における歩行者の危険度推定のためには、人を同定するのではなく、その人がいまどういう状態なのか、どういう行動をするのかを知ることが重要である。例えば、体が車道側を向いていて車道に飛び出してきそうか否かなどに注目する。

そこで本サーベイでは、歩行者の属性を、以下の 3 つのグループに分類する。

- 身体属性：短期間で変化することが少ない人固有の属性
- 外見属性：服装など変化しうる外見上の属性
- 行動属性：観測時点での歩行者の行動で決まる属性

これらの属性を認識するためのアプローチとしては、各属性に特化して精度よく認識を行うアプローチと、上記のグループをまたぐ多様な属性を、それらの関係を元に複合的に認識することを狙うアプローチがある。そのため、前者のアプローチを「特化型」、後者のアプローチを「複合型」として分類する。3. 節では、「特化型」の研究について上記 3 グループに分類される属性を扱う手法を紹介する。次いで、4. 節では、「複合型」の研究について紹介する。

3. 「特化型」属性認識手法

この節では、前節で分類した属性に対し、各種属性に特化して精度よく推定する手法について、それぞれ代表的なものを紹介する。

3.1 身体属性

前節で身体属性と名付けた、人固有の属性について紹介する。これらは、一般に Soft Biometrics と呼ばれている [17] ~ [19]。それらの中でも、性別、年齢の推定、及び身長推定について詳述する。

3.1.1 性別

顔画像を対象とした性別認識の研究は古くから多数の研究が行われている。初期の研究では、顔パーツの位置関係などの幾何的な関係の特徴にした研究がなされてきた [20] ~ [22]。一方、機械学習の発展に伴い、顔画像を入力として、特徴選択や識別器の学習を行う手法も注目されるようになってきた [23] ~ [28]。古くは、Golomb ら [29] により、ニューラルネットワークを用いた顔画像から性別認識を行う SexNet という手法が提案されている。

Wang ら [30] は、局所特徴量の記述子として Local Binary Patterns (LBP) を改良し、量子化にクラスタリングを用いた Local Circular Patterns (LCP) を提案している。LCP によって識別性能を高めることにより、FERET database により 95.36% の認識率を達成している。

Shih ら [31] らは、Active Appearance Model (AAM) によって得た顔画像中のランドマーク周辺から、LBP のヒストグラムを作り、各ランドマークが属する顔パーツごとに重み付けした特徴量である、Precise Patch Histogram (PPH) を提案し、隠れに対する頑健性を高めている。

可視光カメラの映像は照明変化に弱いため、Wang ら [32] は、赤外線カメラと可視光カメラの観測結果を統合した性別認識手法を提案している。可視光カメラの画像から AAM によって得たパーツ位置と、赤外線カメラの画像から抽出した LBP 特徴を元に、F 検定によって特徴選択を行い、Bayesian Network によって性別認識を行っている。

多くの手法は、顔画像からの性別推定を行うものであるが、Cao ら [33] は、顔画像からではなく、全身の画像から性別を

推定する手法を提案している．全身写真，しかも後ろ姿からでも人間は性別推定ができることに着目し，Part-based Gender Recognition (PBGR) という手法を提案している． 6×19 ピクセルの HOG 特徴量を用いたアンサンブル学習により，MIT pedestrian database で 75% 程度の認識精度を達成している．

なお，Ng ら [34] [35] により，性別推定に特化した詳細なサーベイがなされている．このサーベイでは，顔画像から性別推定する手法，全身画像から性別推定する手法に加え，歩容特徴から性別推定する手法についてもサーベイがなされている．

3.1.2 年 齢

年齢推定の研究は，大まかな年代に分類する手法 (Age Classification) [36] ~ [38] と，年齢に回帰する手法 (Age Estimation) [39] ~ [49] に分けられる．

年代に分類する手法としては，大人，子供といった大まかな分類の研究から始まり，細かな年齢層の推定へと発展している．

Kwon ら [36] は，顔画像を元に，顔のパーツ，輪郭，及び顔のしわを特徴として，大人か子供かを分類する手法を提案している．独自データセットを用い，顔のしわが見えるぐらいの画像ならば 100% の認識率を達成している．

Davis [50] は，大人と子供の歩き方の違いに着目し，身長，歩幅，周期などの特徴を元に大人と子供を分類する手法を提案している．頭と足首のモーションキャプチャによりデータを収集し，線形識別器により独自データセットにおいて 95% の認識率を達成している．

Zhou [37] らは，ラドン変換による特徴抽出を行い，Scaling SVM による分類を行う手法を提案している．Scaling SVM で特徴選択と重み付けを行う際に，Rényi エントロピーを計算することで，分類に有効な特徴を選択している．

Ylioinas ら [38] は，顔だけを実験環境で撮影した画像ではなく，制限のない実環境で撮影されたデータに対して，子供から大人まで 7 段階の分類を行う手法を提案している．顔の細かな特徴を強調するために，LBP で 2 値化する際に失われる画素値の差の大きさを表現する CLBP_M ヒストグラムを導入している．Images of Groups database を用いて評価を行い，7 段階の分類について 51.7% の精度を達成している．

一方，回帰に基づく手法は，分類に比べ，詳細な年齢の推定を目的としている．評価には FG-Net Aging Database が利用されることが多いようである．

Geng ら [39] は，Aging Pattern Vector という，年齢順に並べた特徴量集合を人ごとに用意し，Aging Pattern の部分空間を学習する手法を提案している．顔画像が与えられた際に，その人の Aging Pattern を部分空間から再構成することにより年齢を推定する．FG-Net Aging Database で推定誤差 6.77 歳を達成している．

Guo ら [40] は，顔画像から多様体を構築してその上での回帰を行う手法を提案している．特に，回帰方法として，SVR で回帰した値を局所的に修正することで精度を高める Locally Adjusted Robust Regression (LARR) という回帰手法を提案している．この回帰法により，FG-Net Aging Database で推定誤差 5.07 歳を達成している．

Chen ら [48] は，Cumulative Attribute という中間表現を用い，画像特徴から Cumulative Attribute，Cumulative Attribute から目的変数への回帰を同時に学習する手法を提案している．顔画像からの年齢推定及び，群衆画像からの人数推定に適用し，評価を行っている．年齢推定では，FG-Net Aging Database で推定誤差 4.67 歳を達成している．

Nguyen ら [51] は，Multilevel LBP (MLBP) 特徴と Support Vector Regression (SVR) を用いた年齢推定手法を提案している．回帰の前処理として，性別や表情によってデータを分類しておくことの有効性を実験により調査している．学習データが多数ある状況であれば，表情や性別による分類が有効であることを示している．

これらの手法に対し，性別の推定と同様に，全身の画像から年齢を推定する手法も提案されている．Ge ら [47] は，SIFT 特徴のスパースコーディングによって特徴記述を行い，回帰を行っている．Web から収集した画像に対して 8.76 歳の精度を達成している．

なお，Fu ら [52] により，年齢推定のサーベイも発表されている．

3.1.3 身 長

遠方にいる歩行者など，顔画像が高解像度で得られないような対象に対して，年齢を正確に推定することは困難である．しかし，歩行者の身長がわかれば，大人か子供かといった，荒い粒度で年齢を推定することは出来る．

歩行者の身長は，カメラの姿勢が決まれば人物検出位置から推定することが出来る．そのため，人物検出位置から身長へのマッピングを得るために，カメラのパラメータ推定と回帰が用いられる [53] ~ [65] ．

Momeni-K ら [61] は，カメラ姿勢と消失点位置を既知として，物体の高さを精度よく推定する手法を提案している．対象物と画像平面が平行になるような擬似的な画像平面へと画像を射影することによりカメラの光軸を地面と並行にし，射影後の画像から高さ推定を行っている．

Lo ら [62] は，多数設置された視野重複のある防犯カメラ映像を用いて，歩行者検出と同時に身長推定をする手法を提案している．背景差分により得た各歩行者の領域ごとに，カメラから見てどの方向に人がいるかという情報を得る．各カメラから得られた情報を地図上にマッピングして統合することにより，精度よく歩行者検出を行う．検出位置に対して高さが既知の仮想物体を置き，画像平面へ逆投影することにより各歩行者の身長を推定する．

Li ら [65] は，足元位置と頭部位置から非線形回帰によって身長推定を行う手法を提案している．陽に消失線を求める従来手法に対し，ノイズに頑健であると主張している．

3.1.4 そ の 他

その他，身体に関する属性としては，他に人種や民族の推定 [66] ~ [68] ，髪色の推定 [69], [70] など様々な属性について研究がなされている．

また，Han ら [71] は，顔画像データベースに対し，Amazon Mechanical Turk を利用して多くの人間に年齢，性別，人種を

回答してもらい、人間であればどの程度の精度が出せるのかを調べ、結果をまとめている。

3.2 外見属性

この節では、前節で外見属性と名付けた、ある時間を限定した時に人を同定しうる属性について紹介する。カメラ間人物照合や、人物画像検索において、画像中の人物が同一か否かを判定する際、歩行者の全身の外見から特徴抽出を行うことが多い。以下では、特に歩行者の外見に大きな影響を与える、服装と所持品について詳述する。

3.2.1 服装

服装は、最も歩行者の外見に影響を与えるものである。服装を認識することにより、それを属性として人物の特徴記述を行い、服装による人物検索や人物同定が可能となる [72], [73]。また、近年では、ファッション分析の研究 [74] ~ [78] も注目を集めている。ファッション分析については、Liu ら [76] により研究の動向が報告されている。

上着、ズボンなどの服装をそれぞれ認識する場合、画像中の人体の領域分割が必要となる。上半身と下半身などおおまかな分割に基づくアプローチ [74], [77], [79] と、姿勢推定などに基づきさらに詳細な領域分割を行うアプローチ [80], [81] が存在する。

おおまかな分割に基づくアプローチとして、Bossard ら [77] は、上半身検出を用いて領域を限定したうえで、密な特徴抽出を行い、Random Forest と SVMs による認識を行っている。

Chen ら [79] は、Web 上のショッピングサイトから自動的に学習した服装の属性を、Region Proposal によって提案される領域候補矩形に対して認識する手法を提案している。

一方、詳細な領域分割に基づくアプローチとしては、Poselets [82] が提案されて以来、体の部位を詳細に分割してから個々の服装を認識するという方法がとられるようになった。近年では、Flexible Mixtures of Parts (FMP) という手法 [83] も体の部位の分割手法として注目されている。

服装は、上着とズボンなど、属性間の共起性が高いため、4. 節で述べる、複数の属性を複合的に推定するアプローチがとられることが多い。

3.2.2 所持品

所持品の有無を表す属性の認識は、一般に何かの物体を持っているか否かという判定をする手法 [84] ~ [88] と、リュックなど特定の物体を持っているか否かの判定をする手法 [89] ~ [94] に分けられる。

前者の手法は、特定の物体検出器を利用するのではなく、シルエットや、所持物と人との位置関係などに着目する。

Damen ら [85], [87] は、歩行者のシルエットを蓄積し、所持品を持っていない場合のシルエットとの差分を取ることで、何か物体を所持しているか否かを判定する手法を提案している。

Tavanai ら [88] は、人が持ち運んでいる物一般を検出するために、Geometric Shape Model を利用し、時空間の整合性をもとに所持物を追跡する手法を提案している。

一方、後者の手法はリュックやバッグなど所持品ごとに学習

した検出器を用い、歩行者画像をスキャンして検出ができれば所持、そうでなければ非所持と判断することが多い。

Schels ら [89] は、所持品の中で trolley, suitcase, backpack の所持に注目した人物画像検索手法を提案している。対象とする所持品の 3 次元モデルを用いて大量に所持品画像を生成し、所持品検出器を構築する。検出器を用いて取得した所持品領域の集合に対して、人物と所持品の 3 次元相対位置を考慮したフィルタリングを行うことで、精度の良い所持品に基づく人物画像検索を実現している。

浅井ら [90] は、所持品の中で空港で所持されていることが多いキャリーバッグ、リュックサック、ショルダバッグを対象とした手荷物所持判定法を提案している。被写体の方向によって所持品の見た目が変化することと、人物画像に対する所持品の相対位置がある程度決まることに着目し、学習データを人物画像の向きごとに別々に用意し、所持品が出現する領域の画像特徴量のみを用いて分類器を学習する。また、複数視点からの認識結果を統合することで、性能を向上させる手法 [91] も提案している。

Chua ら [92] は、Sling Bag と Backpack を対象として、所持品検出を行う手法を提案している。所持品の種類と人物画像の方向毎に特化したモデル化を行うことで、精度の良い所持品検出を行う。

Du ら [94] は、バッグの所持判定と、所持している場合どの位置に所持しているかの検出を行う手法を提案している。Random Forest に基づく手法により所持判定を行い、所持していると判定された場合のみ、Region Proposal で得られる領域に対して Convolutional Neural Network によるスコア付けを行い、最もスコアの高い領域を所持品領域として出力する。

他に、可視光カメラではなく、赤外線カメラの映像を対象として Kresnaraman ら [95] は、リュックの所持や、メガネ、帽子などの着用の有無を判定する手法について報告している。

以上の手法のように、特定の物体の検出に基づくアプローチでは、予め検出対象がバッグなどに限定されていれば学習データを収集することは容易であるが、任意の所持品に注目した所持の有無を判定する識別器を構築する場合、その都度定めた所持品に対する識別器を構築することは困難である。

井関ら [96] は、任意の所持品の所持の有無に基づく人物画像分類を目的とした、任意の所持品の所持の有無を判定する識別器を学習する枠組みを提案している。転移学習と能動学習を組み合わせた、能動的転移更新モデルを提案し、学習データがほとんど存在しない状況でも効率的な識別器の構築を実現している。

また、Dondera [97] らも、Multiple Instance Learning を用いて、できるだけ少ないアノテーションによってキャリーバッグの検出器を構築する手法を提案している。

3.3 行動属性

前述の 2 種類の属性に対し、本節では、ある瞬間における人の属性について述べる。ここでは、特に人の行動に起因する属性に着目し、特に ITS の研究において重視されている、歩行者向き推定及び、その他の代表的な研究事例を紹介する。

3.3.1 人体向き推定

歩行者は一般的に体の向きに従って移動するため、歩行者を検出し追跡することで体の向きを推定することができる。しかしこの方法では、歩行者が立ち止まっている場合や向きを変えた場合に正確な推定ができない。そのため、歩行者の画像から体の向きを推定する研究が行われてきた。

体の向きは、姿勢推定の研究において姿勢を表すパラメータの1つとして扱われていた [98] が、歩行者の進行方向の予測 [99], [100] や注意が向いている方向の推定 [101] において重要な属性であることから、向きを推定するための研究が行われてきた。向き推定のアプローチは、主に歩行者検出と同様に統計的学習の手法が用いられた [102]。向き推定では歩行者検出と比べ、方向による人体の見えの変化が大きくないため遠距離での推定が難しく、これを解決するための手法も提案されてきた [103], [104] が、遠距離での向き推定は依然として解決すべき課題の1つである。近年では、カメラの性能向上による解像度向上や高精度に人体の部位の位置を推定する手法が登場したこともあり、人体の各部位の情報を用いて性能向上を図る手法が提案されており [105] ~ [107]、成果を上げている。また、歩行者検出へ応用され、検出と向き推定を同時に行うことで性能を向上した手法も提案されている [108], [109]。画像に加えて距離情報を利用することで性能を向上した手法も提案されている [110], [111]。

向き推定の研究では、前処理として歩行者検出を想定し、歩行者を切り出した画像を入力とすることが多い。出力は、4方向や8方向のクラスに分類する研究と、角度を推定する研究がある。評価は、4方向や8方向の分類における正解率にて行われることが多い。

以下では、各手法について詳細に説明していく。

向きを分類する手法は、歩行者の向きごとに low level な特徴を学習して識別器を構築する手法が提案されている。構築した識別器から最大のスコアを得た方向を向き推定結果とする手法もあるが、向きの分類ではクラス間に関係があるため、それを考慮した手法が提案されている。例えば 45° ごとの8方向の場合、 0° と 45° は差が小さいが、 0° と 180° は差が大きい。そのため、近いクラスの識別器の出力を利用して向き推定を行う手法 [99] や、識別器の出力をもとに各方向の確率分布を求めて向き推定を行う手法 [100], [103] が提案されている。この改良により、向き推定の誤差を小さくし、分類の精度を向上させることに成功している。また、各方向の確率分布を利用して、向きを角度で推定する手法 [108] も提案されている。

Shimizu ら [99] は、方向ごとの複数の識別器を構築し、分類時に隣接クラスの識別器の出力も考慮した手法を提案している。画像から Haar wavelet 特徴を抽出し、向きごとに SVM の学習を行う。 22.5° ごとの16方向の識別器を構築し、隣接クラス (45° であれば 22.5° と 67.5°) の識別器の出力との重み付き和により向きを推定する。隣接クラスの識別器の出力を利用することで、1つの識別器の出力を用いる場合と比べて向き推定精度の向上を達成した。

Baltieri ら [103] は、複数の識別器の出力から方向の連続分布

を推定し、分布から向きを推定する手法を提案している。特徴量としてブロックサイズの異なる3段階の HOG (three level HOG) を利用し、ランダムフォレストを学習し、各クラスの確率を出力する識別器を構築する。識別器の出力から、提案された Mixture of Approximated Wrapped Gaussian (MoAWG) によって、方向の連続分布を推定する。MoAWG は識別器の出力の“補間”として働き、識別器の出力から向きを推定する手法と比べて精度向上を達成した。TUD データセットによるベンチマークでは、8方向の推定において 65% の認識率を達成した。

また、体の各部位の情報を利用することで精度向上を図る手法が提案されている。各部位を利用する手法は姿勢推定にしばしば用いられるが、詳細なモデルを利用して各部位の位置を特定する姿勢推定の手法とは異なり、向き推定では、一部の部位のみ用いる手法 [101] や各部位の low level な特徴を用いた手法 [105] が見られる。体の各部位の検出には Pictorial Structure [112] や Poselets [82] が利用される。

Hayashi ら [107] は、頭の位置から求めた上体姿勢 (spine pose) を利用した向き推定手法を提案している。提案された Poselet-regressor により、頭の位置から上体姿勢を予測し、上体の角度が得られる。あらかじめ上体の角度ごとに向き識別器を学習し構築しておくことで、向き推定の精度向上を達成している。

Tao ら [105] は、識別器の学習時に画像から大きさと位置をランダムに選択したパッチ (body parts) を利用した向き推定手法を提案している。識別器にランダムフォレストを利用し、各決定木の学習をランダムに選ばれたパッチで行うことにより、画像全体を用いて学習を行った場合よりも向き推定精度が向上することを報告している。TUD データセットによるベンチマークでは、4方向の推定において 66% の認識率を達成した。

体の向き推定に利用される特徴は、形状を表現する特徴が用いられる傾向があり、輪郭のエッジ [98], [101] や HOG 特徴 [100], [102], [103], [105], [107] が用いられている。Tosato ら [104] は、画像特徴の共分散を利用した特徴記述子を提案し、小さい画像やノイズの多い画像を対象とした体の向き推定において有効であることを報告している。Goto ら [109] は、HOG の近接するヒストグラムを掛け合わせた Feature Interaction Descriptor (FIND) を提案している。

別のアプローチとして、歩行者の検出と向き推定を同時に行う手法 [108], [109] が提案されている。歩行者の検出と向き推定を組合せて行うことで性能が向上することが報告されている [108]。Goto ら [109] は、カスケード型に構成した検出器の最終段において、向きごとに歩行者検出器を構築することで、検出と向き推定を同時に行う手法を提案している。歩行者は向きごとに見えの変化が大きいため、それぞれの向きについて別々に学習して検出器を構築することで、歩行者の検出性能の向上を達成し、同時に向き認識を行った。また、Goto らの手法では時系列情報も利用しており、向きを複数フレームの多数決により決定する。複数フレームの多数決を利用することで、利用しない場合に比べ精度向上を達成した。

体の向きと頭の向きには相関がある。Flohr ら [106] は、体の向きと頭の向きを双方向に利用し、両方の推定精度を向上させた手法を提案している。前フレームの体と頭の向き推定結果と、識別器を用いて得られた現フレームの身体と頭の向きの推定結果を用いて、Dynamic Bayesian network model により向きを推定する。頭の位置を Pictorial structure model を用いて求め、体と頭の向きを組み合わせることで、それぞれ独立で向きを推定した場合に比べ精度向上を達成した。

可視光画像 (RGB 画像) 以外の画像を用いた手法も提案されている。Shinmura ら [110], [113] は、RGB 画像と距離画像 (Depth 画像) を同時に取得可能なセンサを利用し、向き推定時の背景の影響を低減した手法を提案している。Depth 画像では人体と背景の分離が容易であることを利用し、RGB 画像からの特徴抽出時に背景を除去することで、画像特徴にあらわれる背景の影響を低減し、向き推定の精度向上を達成した。

Liu ら [111] は、時系列の RGB 画像と Depth 画像から得られる特徴を用いて向きを推定する手法を提案している。人体の形状と表面の凹凸を表す特徴とオプティカルフローに似た 3 次元の動きを表す特徴を抽出し、Dynamic Bayesian network を用いて向きを推定する。RGB 画像のみや Depth 画像のみを用いた他手法と比較し、RGB と Depth の情報を組み合わせることで精度向上を達成した。

3.3.2 その他

歩行者の周辺への注意不注意の度合いは、交通安全において非常に重要である。特に、ヘッドフォンやスマートフォンなどのモバイル・デバイスを使用中の歩行者は、周りへの注意度が低下することにより危険であることが報告されている [114], [115]。この問題に対し、新村ら [116] は、車載カメラ画像中の歩行者に対し、スマホ歩きをしているか否かの認識を行い、不注意度を推定する手法を提案している。

4. 「複合型」属性認識手法

歩行者の属性の間には、関連があるものがある。例えば、「半ズボン」の属性を持つ人は「半袖シャツ」の属性を持つ可能性が高く、「スカート」の属性を持つ人は「女性」の属性を持つことが多い。こうした属性間の関係を利用する手法のうち代表的なものを紹介する。

こうした考えに基づく初期の研究では、Neural Network を用いて複数属性を認識する手法が挙げられる。Cottrell ら [117] は、現在使われている Autoencoder に相当する特徴抽出用の Neural Network を用いて学習した顔画像特徴量を元に、識別用の Neural Network によって顔、感情、性別の認識を行っている。

それぞれの属性認識器で認識した結果を元に、それらの結果を統合して最終的な属性認識結果を再計算したり、別の属性を求めるアプローチとしては、属性認識器の出力結果を特徴量として分類する手法 [73] や、Bayesian Network を利用したものが挙げられる [80], [118], [119]。Yang ら [73] は、ポロノイ図を利用した画像の領域分割により、衣服の領域を分割して特徴抽出し、年齢、性別の推定結果とともに SVM によって服

装の種類を認識する手法を提案している。Bourdev ら [80] は、Poselets [82] を用いて検出した各人体部位に対して属性認識を行い、それらの結果を Bayesian Network で統合している。

個別の属性認識を、他の属性認識結果を加味して決定する、Conditional Random Field (CRF) を用いる手法も多い [74], [120] ~ [122]。特に、服装の認識など、相関が強い属性間の認識において、よく利用されている。Yamaguchi ら [78] は、姿勢推定と Superpixel 分割により画像を分割し、各領域から特徴量を抽出したうえで CRF によって各領域の服装を認識する手法を提案している。また、陽に姿勢推定をせずに、人体周辺の矩形領域に対する CRF により、各衣服のラベルを割り当てる手法 [74] も提案している。

近年では、Deep Learning の技術を用い、特徴抽出や分類の初期段階を共通化することで、共通の特徴や属性間の相関を学習するアプローチもとられている [79], [123]。Zhang ら [123] らは、Convolutional 層を共通化した CNN を学習し、多種の属性を認識できるネットワークを構築している。

複数の属性認識問題におけるその他のトピックとしては、属性の追加がある。一般に、新たな属性を追加する場合、これまでのデータセットに対し、新たに追加する属性のラベルを追加して、学習をする必要があるが、頻繁に属性が追加される場合、非常に手間がかかり現実的ではない。この問題に対し井尻ら [124] は、Codebook による中間表現をすることにより、少数のラベル付き学習データだけで新たな属性を属性認識器に追加することを可能とする手法を提案している。

5. 評価用データセット

顔画像は比較的長く研究が行われているため、様々なデータセットが存在する。性別認識のためのデータセットとしては、Color FERET dataset [125], Labeled Face in the Wild (LFW) dataset [126], USTC-NAIVE database [127] などがよく利用されるようである。年齢推定用のデータセットとしては、FG-Net Aging Database [128] や MORPH [129] が有名である。また、可視光カメラだけでなく、赤外線カメラの画像も含む、CBSR NIR Face Dataset [130] も公開されている。

服装などの外見属性に関するデータセットの多くは、既に公開されている人物画像データセット (人物検出・追跡や人物照合を対象としたもの: H3D [82], VIPeR [131], ETHZ [132] など) に対し、自前で、もしくは Amazon Mechanical Turk によって属性情報のアノテーションを行ったものである。代表的なものとして、Dataset of Human Attribute (HAT) [133], Attributes of People dataset [80], Attributes for Re-identification [134], APiS 1.0 Database [135], PEdesTrian Attribute (PETA) [136] などが公開されている。特に服装に限定すれば、Yamaguchi ら [137] らによる、ファッション SNS である Chictopia の写真に対するアノテーションをつけたもの、Vittayakorn ら [138] らによる 15 年間のファッションショーの写真にアノテーションをつけたものが公開されている。

歩行者の体の向きといった行動属性について、ITS 関係の研究機関からベンチマークが公開されており、精度の向上が競われて

いる [139] . また , 人物検出用の公開データセット [140] に対し , 独自に付与したアノテーションを公開している Pedestrian Direction Classification (PDC) [105] もある . その他 , TUD [141] , Human Orientation Classification Dataset (HOC) [104] など も利用されている .

6. おわりに

「属性」は画像の意味的に高度な理解のための重要なキーであると近年注目を集めている . 本サーベイでは , 歩行者に付与される属性を , 身体属性 , 外見属性 , 行動属性の 3 つに分類し , それらに特化した手法及び , 複数の属性を汎用的に認識する手法についてまとめた . また , 各手法で評価に利用されるデータセットについて幾つか紹介した .

本サーベイでは紙面の都合上 , 大きな分類と代表的な手法の簡潔な説明にとどめた . また , 幾つかの分類については十分な議論ができていない部分がある . 今後 , より詳細なサーベイを論文誌等で発表していきたい .

謝辞 本研究の一部は国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーションプログラム (名古屋 COI:高齢者が元気になるモビリティ社会)」の支援によって行われた .

文 献

- [1] V. Ferrari and A. Zisserman: “Learning visual attributes”, *Proceedings of Advances in Neural Information Processing Systems*, pp. 433–440 (2008).
- [2] A. Farhadi, I. Endres, D. Hoiem and D. Forsyth: “Describing objects by their attributes”, *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1778–1785 (2009).
- [3] D. Parikh and K. Grauman: “Relative Attributes”, *Proceedings of the 13th International Conference on Computer Vision* (2011).
- [4] C. H. Lampert, H. Nickisch and S. Harmeling: “Learning to detect unseen object classes by between-class attribute transfer”, *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 951–958 (2009).
- [5] A. Farhadi, I. Endres and D. Hoiem: “Attribute-centric recognition for cross-category generalization”, *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2352–2359 (2010).
- [6] Y. Wang and G. Mori: “A discriminative latent model of object classes and attributes”, *Proceedings of the 11th European Conference on Computer Vision*, pp. 155–168 (2010).
- [7] T. Nortcliffe: “People analysis cctv investigator handbook”, *Home Office Centre of Applied Science and Technology*, **2**, p. 3 (2011).
- [8] M. Enzweiler and D. M. Gavrilu: “Monocular Pedestrian Detection: Survey and Experiments”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**, 12, pp. 2179–2195 (2009).
- [9] 山内, 山下, 藤吉: “[サーベイ論文] 統計的学習手法による人検出”, *信学技報* (2012).
- [10] E. Murphy-Chutorian, S. Member and M. M. Trivedi: “Head Pose Estimation in Computer Vision: A Survey”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**, 4, pp. 607–626 (2009).
- [11] Z. Liu, J. Zhu, J. Bu and C. Chen: “A survey of human pose estimation: The body parts parsing based methods”, *Journal of Visual Communication and Image Representation*, **32**, pp. 10–19 (2015).
- [12] S. Mitra and T. Acharya: “Gesture Recognition : A Survey”, *IEEE Transactions On Systems, Man, And Cybernetics - Part C: Applications And Reviews*, **37**, 3, pp. 311–324 (2007).
- [13] R. Poppe: “A survey on vision-based human action recognition”, *Image and Vision Computing*, **28**, 6, pp. 976–990 (2010).
- [14] R. Farrell, O. Oza, V. I. Morariu, T. Darrell and L. S. Davis: “Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance”, *Proceedings of the 13th International Conference on Computer Vision*, pp. 161–168 (2011).
- [15] A. Kovashka, D. Parikh and K. Grauman: “WhittleSearch: Image search with relative attribute feedback”, *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2973–2980 (2012).
- [16] S. Shankar, V. K. Garg and R. Cipolla: “DEEP-CARVING : Discovering Visual Attributes by Carving Deep Neural Nets”, *Proceedings of the 2015 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2015).
- [17] A. Dantcheva, C. Velardo, A. D’Angelo and J. L. Dugelay: “Bag of soft biometrics for person identification: New trends and challenges”, *Multimedia Tools and Applications*, **51**, 2, pp. 739–777 (2011).
- [18] D. A. Reid, M. S. Nixon and S. V. Stevenage: “Soft Biometrics; Human Identification Using Comparative Descriptions”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **36**, 6, pp. 1216–1228 (2014).
- [19] P. Tome, J. Fierrez, R. Vera-Rodriguez and M. S. Nixon: “Soft biometrics and their application in person recognition at a distance”, *IEEE Transactions on Information Forensics and Security*, **9**, 3, pp. 464–472 (2014).
- [20] R. Brunelli and T. A. Poggio: “HyberBF Network for Gender Classification”, *DARPA Image Understanding Workshop*, pp. 311–314 (1992).
- [21] A. M. Burton, V. Bruce and N. Dench: “What’s the difference between men and women? Evidence from facial measurement”, *PERCEPTION-LONDON-*, **22**, p. 153 (1993).
- [22] J. M. Fellous: “Gender discrimination and prediction on the basis of facial metric information”, *Vision Research*, **37**, 14, pp. 1961–1973 (1997).
- [23] A. J. O’Toole, N. Troje and H. H. Bühlhoff: “Sex classification is better with three-dimensional head structure than with image intensity information”, *Perception*, **26**, 1, pp. 75–84 (1997).
- [24] B. Moghaddam and M.-h. Yang: “Gender Classification with Support Vector Machines”, *Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition* (2000).
- [25] B. Moghaddam and M. H. Yang: “Learning gender with support faces”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, pp. 707–711 (2002).
- [26] G. Bebis and S. Louis: “Genetic feature subset selection for gender classification: a comparison study”, *Workshop on Applications of Computer Vision*, pp. 165–170 (2002).
- [27] S. Buchala, N. Davey, T. M. Gale and R. J. Frank: “Analysis of linear and nonlinear dimensionality reduction methods for gender classification of face images”, *International Journal of Systems Science*, **36**, 14, pp. 931–942 (2005).
- [28] H.-c. Lian, B.-l. Lu, E. Takikawa and S. Hosoi: “Gender Recognition Using a Min-Max Modular Support Vector Machine”, *Advances in Natural Computation*, **3611**, pp. 438–441 (2005).
- [29] B. Golomb, D. Lawrence and T. Sejnowski: “Sexnet: A

- Neural Network Identifies Sex from Human Faces” (1990).
- [30] C. Wang, D. Huang, Y. Wang and G. Zhang: “Facial image-based gender classification using local circular patterns”, *International Conference on Pattern Recognition*, pp. 2432–2435 (2012).
- [31] H.-C. Shih: “Robust gender classification using a precise patch histogram”, *Pattern Recognition*, **46**, 2, pp. 519–528 (2013).
- [32] S. Wang, Z. Gao, S. He, M. He and Q. Ji: “Gender recognition from visible and thermal infrared facial images”, *Multimedia Tools and Applications* (2015).
- [33] L. Cao, M. Dikmen, Y. Fu and T. S. Huang: “Gender Recognition from Body”, *Proceedings of the 16th ACM international conference on Multimedia*, pp. 725–728 (2008).
- [34] C. B. Ng, Y. H. Tay and B. M. Goi: “Vision-based Human Gender Recognition: A Survey”, *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 30 (2012).
- [35] C. B. Ng, Y. H. Tay and B.-m. Goi: “Recognizing Human Gender in Computer Vision: A Survey”, *Proceedings of the 12th Pacific Rim International Conference on Artificial Intelligence*, Vol. 7458, pp. 335–346 (2012).
- [36] Y. H. K. Y. H. Kwon and N. D. V. Lobo: “Age classification from facial images”, *Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1994).
- [37] H. Zhou, P. Miller and J. Zhang: “Age classification using Radon transform and entropy based scaling SVM”, *Proceedings of the British Machine Vision Conference*, pp. 1–28 (2011).
- [38] J. Ylioinas, A. Hadid and M. Pietikäinen: “Age Classification in Unconstrained Conditions Using LBP Variants”, *Proceedings of the 21st International Conference on Pattern Recognition*, pp. 1257–1260 (2012).
- [39] X. Geng, Z. H. Zhou and K. Smith-Miles: “Automatic age estimation based on facial aging patterns”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**, 12, pp. 2234–2240 (2007).
- [40] G. Guo, Y. Fu, C. R. Dyer and T. S. Huang: “Image-based human age estimation by manifold learning and locally adjusted robust regression”, *IEEE Transactions on Image Processing*, **17**, 7, pp. 1178–1188 (2008).
- [41] Y. Fu and T. S. Huang: “Human Age Estimation With Regression on Discriminative Aging Manifold” (2008).
- [42] G. Guo, G. Mu, Y. Fu and T. S. Huang: “Human Age Estimation Using Bio-Inspired Features”, *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 112–119 (2009).
- [43] K.-Y. C. K.-Y. Chang, C.-S. C. C.-S. Chen and Y.-P. H. Y.-P. Hung: “A Ranking Approach for Human Ages Estimation Based on Face Images”, *Proceedings of the 20th International Conference on Pattern Recognition* (2010).
- [44] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park and J. Kim: “Age estimation using a hierarchical classifier based on global and local facial features”, *Pattern Recognition*, **44**, 6, pp. 1262–1281 (2011).
- [45] K. Y. Chang, C. S. Chen and Y. P. Hung: “Ordinal hyperplanes ranker with cost sensitivities for age estimation”, *Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 585–592 (2011).
- [46] C. Li, Q. Liu, J. Liu and H. Lu: “Learning ordinal discriminative features for age estimation”, *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2570–2577 (2012).
- [47] Y. Ge, J. Lu, X. Feng and D. Yang: “Body-based human age estimation at a distance” (2013).
- [48] K. Chen, S. Gong, T. Xiang and C. C. Loy: “Cumulative attribute space for age and crowd density estimation”, *Proceedings of the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2467–2474 (2013).
- [49] D. R. Gadbail, S. Dhande and K. M. Pimple: “Age Synthesis and Estimation From Face Image”, *International Journal of Engineering and Computer Science*, **3**, 4, pp. 5462–5466 (2014).
- [50] J. W. Davis: “Visual Categorization of Children and Adult Walking Styles”, *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*, pp. 295–300 (2001).
- [51] D. T. Nguyen, S. R. Cho, K. Y. Shin, J. W. Bang and K. R. Park: “Comparative study of human age estimation with or without preclassification of gender and facial expression.”, *The Scientific World Journal*, **2014**, p. 905269 (2014).
- [52] Y. Fu, G. Guo and T. S. Huang: “Age Synthesis and Estimation via Faces: A Survey” (2010).
- [53] A. Bovyrin and K. Rodyushkin: “Human height prediction and roads estimation for advanced video surveillance systems” (2005).
- [54] I. Kispal and E. Jeges: “Human Height Estimation using a Calibrated Camera”, *Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2008).
- [55] C. BenAbdelkader and Y. Yacoub: “Statistical body height estimation from a single image”, *Proceedings of the 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1–7 (2008).
- [56] E. Jeges, I. Kispal and Z. Hornak: “Measuring human height using calibrated cameras” (2008).
- [57] B. Hoogeboom, I. Alberink and M. Goos: “Body height measurements in images.”, *Journal of forensic sciences*, **54**, 6, pp. 1365–75 (2009).
- [58] A. Utasi and C. Benedek: “Multi-Camera People Localization and Height Estimation using Multiple Birth-and-Death Dynamics”, *Proceedings of The 10th International Workshop on Visual Surveillance*, pp. 74–83 (2010).
- [59] J. Shao, S. K. Zhou and R. Chellappa: “Robust Height Estimation of Moving Objects From Uncalibrated Videos” (2010).
- [60] N. Ramstrand, S. Ramstrand, P. Brolund, K. Norell and P. Bergström: “Relative effects of posture and activity on human height estimation from surveillance footage.”, *Forensic science international*, **212**, 1-3, pp. 27–31 (2011).
- [61] M. Momeni-K, S. Diamantas, F. Ruggiero and B. Siciliano: “Height Estimation from a Single Camera View”, *Proceedings of International Conference on Computer Vision Theory and Applications*, pp. page 358–364. (2012).
- [62] K. Lo and C. Wang: “Acceleration of vanishing point-based line sampling scheme for people localization and height estimation via 3D line sampling”, *Proceedings of the 21st International Conference on Pattern Recognition*, pp. 2788–2791 (2012).
- [63] S. Diamantas and P. Dasgupta: “An Active Vision Approach to Height Estimation with Optical Flow”, *Advances in Visual Computing SE - 17* (Eds. by G. Bebis, R. Boyle, B. Parvin, D. Koracin, B. Li, F. Porikli, V. Zordan, J. Klosowski, S. Coquillart, X. Luo, M. Chen and D. Gotz), Vol. 8033 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 160–170 (2013).
- [64] S. X. M. Yang, P. K. Larsen, T. Alkjær, B. Juul-Kristensen, E. B. Simonsen and N. Lynnerup: “Height estimations based on eye measurements throughout a gait cycle”, *Forensic Science International*, **236**, pp. 170–174 (2015).
- [65] S. Li, V. H. Nguyen, M. Ma, C.-B. Jin, T. D. Do and H. Kim: “A simplified nonlinear regression method for human height estimation in video surveillance”, *EURASIP Journal on Im-*

- age and Video Processing, **2015**, 1, p. 32 (2015).
- [66] S. Gutta, H. Wechsler and P. Phillips: “Gender and ethnic classification of face images”, International Conference on Automatic Face and Gesture Recognition (1998).
- [67] D. Huang, H. Ding, C. Wang, Y. Wang, G. Zhang and L. Chen: “Local circular patterns for multi-modal facial gender and ethnicity classification”, Image and Vision Computing, **32**, 12, pp. 1181–1193 (2014).
- [68] E. Boutellaa, A. Hadid, M. Bengherabi and S. Ait-Aoudia: “On the use of Kinect depth data for identity, gender and ethnicity classification from facial images”, Pattern Recognition Letters, **68**, pp. 270–277 (2015).
- [69] N. Kumar, A. C. Berg, P. N. Belhumeur and S. K. Nayar: “Attribute and simile classifiers for face verification”, Proceedings of the 12th IEEE International Conference on Computer Vision, pp. 365–372 (2009).
- [70] R. Feris, R. Bobbitt, L. Brown and S. Pankanti: “Attribute-based People Search: Lessons Learnt from a Practical Surveillance System”, Proceedings of the 2014 International Conference on Multimedia Retrieval (2014).
- [71] H. Han, C. Otto, X. Liu and A. Jain: “Demographic Estimation from Face Images: Human vs. Machine Performance”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **37**, 6, pp. 1148–1161 (2014).
- [72] E. S. Jaha and M. S. Nixon: “Soft Biometrics for Subject Identification using Clothing Attributes”, Proceedings of the 2014 IEEE/IAPR International Joint Conference on Biometrics (2014).
- [73] M. Yang and K. Yu: “Real-Time Clothing Recognition in Surveillance Videos”, Proceedings of the 18th International Conference on Image Processing, pp. 2998–3001 (2011).
- [74] K. Yamaguchi, T. Okatani, K. Sudo, K. Murasaki and Y. Taniguchi: “Mix and Match: Joint Model for Clothing and Attribute Recognition”, Proceedings of the 2015 British Machine Vision Conference (2015).
- [75] H. M. Kiapour, K. Yamaguchi, A. C. Berg and T. L. Berg: “Hipster Wars : Discovering Elements of Fashion”, Proceedings of the 13th European Conference on Computer Vision, pp. 472–488 (2014).
- [76] S. Liu, L. Liu and S. Yan: “Fashion Analysis: Current Techniques and Future Directions”, IEEE Computer Society Magazine, **21**, 2, pp. 72–79 (2014).
- [77] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack and L. Van Gool: “Apparel classification with style”, Proceedings of the 11th Asian Conference on Compute, pp. 321–335 (2013).
- [78] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz and T. L. Berg: “Parsing clothing in fashion photographs”, Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3570–3577 (2012).
- [79] Q. Chen, J. Huang, R. Feris, L. M. Brown, J. Dong and S. Yan: “Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes”, Proceedings of the 2015 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 5315–5324 (2015).
- [80] L. Bourdev, S. Maji, J. Malik, U. C. Berkeley, A. Systems, P. Ave and S. Jose: “Describing People : A Poselet-Based Approach to Attribute Classification”, Proceedings of the 13th International Conference on Computer Vision (2011).
- [81] X. Liang, S. Liu, X. Shen, J. Yang, L. Liu, J. Dong, L. Lin and S. Yan: “Deep Human Parsing with Active Template Regression”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **XX**, X, pp. 1–14 (2015).
- [82] L. Bourdev, J. Malik, U. C. Berkeley, A. Systems, P. Ave and S. Jose: “Poselets : Body Part Detectors Trained Using 3D Human Pose Annotations”, Proceedings of 2009 IEEE International Conference on Computer Vision, pp. 1365–1372 (2009).
- [83] Y. Yang and D. Ramanan: “Articulated Human Detection with Flexible Mixtures-of-Parts.”, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–15 (2012).
- [84] I. Haritaoglu, R. Cutler, D. Harwood and L. S. Davis: “Backpack: detection of people carrying objects using silhouettes” (1999).
- [85] D. Damen and D. Hogg: “Detecting Carried Objects in Short Video Sequences”, Proceedings of the 10th European Conference on Computer Vision, pp. 154–167 (2008).
- [86] T. Senst, A. Kuhn, H. Theisel and T. Sikora: “Detecting people carrying objects utilizing lagrangian dynamics”, Proceedings of the 9th IEEE International Conference on Advanced Video and Signal-Based Surveillance, pp. 398–403 (2012).
- [87] D. Damen and D. Hogg: “Detecting Carried Objects from Sequences of Walking Pedestrians”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **34**, 6, pp. 1056–1067 (2012).
- [88] A. Tavanai, M. Sridhar, F. Gu, A. G. Cohn and D. C. Hogg: “Carried Object Detection and Tracking using Geometric Shape Models and Spatio-Temporal Consistency”, Computer Vision Systems SE - 23, **7963**, pp. 223–233 (2013).
- [89] J. Schels, J. Liebelt and R. Lienhart: “Self-calibrating 3D context for retrieving people with luggage”, Proceedings of the 13th IEEE International Conference on Computer Vision Workshops, pp. 1920–1927 (2011).
- [90] 浅井, 高橋, 出口, 井手, 村瀬: “手荷物の種類と向きを考慮した人物画像からの手荷物の所持判定に関する検討”, 信学技報, pp. 115–120 (2013).
- [91] 浅井, 西堀, 高橋, 出口, 井手, 村瀬: “複数方向から撮影された人物の手荷物所持判定に関する検討”, 信学技報, pp. 21–24 (2015).
- [92] T. W. Chua, K. Leman, H. L. Wang, N. T. Pham, R. Chang, D. D. Nguyen and J. Zhang: “Sling bag and backpack detection for human appearance semantic in vision system”, Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2130–2135 (2013).
- [93] B. Yuan, Q. Ruan and G. An: “Carried object detection in short video sequences”, Proceedings of 2014 12th International Conference on Signal Processing, pp. 1311–1316 (2014).
- [94] Y. Du, H. Ai and S. Lao: “A Two-Stage Approach for Bag Detection in Pedestrian Images”, Proceedings of the 12th Asian Conference on Computer Vision (2014).
- [95] B. Kresnaraman, D. Deguchi, T. Takahashi, I. Ide and H. Murase: “A Preliminary Study on Human Attribute Classification in Thermal Image”, 2015 年電子情報通信学会総合大会講演論文集, p. 66 (2015).
- [96] 井関, 川西, 椋木, 美濃: “所持品に基づく人物画像分類のための学習データが少ない問題に対する分類器学習”, 情報処理学会第77回全国大会講演論文集, pp. 547–549 (2015).
- [97] R. Dondera, V. Morariu and L. Davis: “Learning to Detect Objects with Minimal Supervision”, Proceedings of the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 759–766 (2013).
- [98] G. Shakhnarovich, P. Viola and T. Darrell: “Fast pose estimation with parameter-sensitive hashing”, Proceedings of the 9th IEEE International Conference on Computer Vision, p. 750 (2003).
- [99] H. Shimizu and T. Poggio: “Direction estimation of pedestrian from multiple still images”, Proceedings of 2004 IEEE Intelligent Vehicles Symposium, pp. 596–600 (2004).
- [100] T. Gandhi and M. M. Trivedi: “Image based estimation of pedestrian orientation for improving path prediction”, Proceedings of 2008 IEEE Intelligent Vehicles Symposium, pp. 506–511 (2008).
- [101] O. Ozturk, T. Yamasaki and K. Aizawa: “Estimating hu-

- man body and head orientation change to detect visual attention direction”, 10th Asian Conference on Computer Vision Workshop (2010).
- [102] C. Weinrich, C. Vollmer and H. M. Gross: “Estimation of human upper body orientation for mobile robotics using an SVM decision tree on monocular images”, Proceedings of 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2147–2152 (2012).
- [103] D. Baltieri, R. Vezzani and R. Cucchiara: “People orientation recognition by mixtures of wrapped distributions on random trees”, Proceedings of the 12th European Conference on Computer Vision, pp. 270–283 (2012).
- [104] D. Tosato, M. Spera, M. Cristani and V. Murino: “Characterizing Humans on Riemannian Manifolds”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **35**, 8, pp. 1972–84 (2013).
- [105] J. Tao and R. Klette: “Part-based RDF for Direction Classification of Pedestrians, and a Benchmark”, Proceedings of the 12th Asian Conference on Computer Vision Workshop (2014).
- [106] F. Flohr, M. Dumitru-Guzu, J. F. P. Kooij and D. M. Gavrila: “Joint probabilistic pedestrian head and body orientation estimation”, Proceedings of 2014 IEEE Intelligent Vehicles Symposium, pp. 617–622 (2014).
- [107] M. Hayashi, K. Oshima, M. Tanabiki and Y. Aoki: “Upper Body Pose Estimation for Team Sports Videos Using a Poselet-Regressor of Spine Pose and Body Orientation Classifiers Conditioned by the Spine Angle Prior”, IPSJ Transactions on Computer Vision and Applications, **7**, pp. 121–137 (2015).
- [108] M. Enzweiler and D. M. Gavrila: “Integrated pedestrian classification and orientation estimation”, Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 982–989 (2010).
- [109] K. Goto, K. Kidono, Y. Kimura and T. Naito: “Pedestrian detection and direction estimation by cascade detector with multi-classifiers utilizing feature interaction descriptor”, Proceedings of 2011 IEEE Intelligent Vehicles Symposium, pp. 224–229 (2011).
- [110] F. Shinmura, D. Deguchi, I. Ide, H. Murase and H. Fujiyoshi: “Estimation of Human Orientation using Coaxial RGB-Depth Images”, Proceedings of 10th International Conference on Computer Vision Theory and Applications, pp. 113–120 (2015).
- [111] W. Liu, Y. Zhang, S. Tang, J. Tang, R. Hong and J. Li: “Accurate estimation of human body orientation from RGB-D sensors”, IEEE Transactions on Cybernetics, **43**, 5, pp. 1442–1452 (2013).
- [112] P. F. Felzenszwalb and D. P. Huttenlocher: “Pictorial structures for object recognition”, International Journal of Computer Vision, **61**, 1, pp. 55–79 (2005).
- [113] F. Shinmura, Y. Kawanishi, D. Deguchi, I. Ide and H. Murase: “Pedestrian Orientation Classification Utilizing Single-Chip Coaxial RGB-ToF Camera”, Proceedings of Workshop on Environment Perception for Automated On-road Vehicles in conjunction with 2015 IEEE Intelligent Vehicles Symposium, pp. 7–11 (2015).
- [114] R. D. Lang: “Don’t Text, Talk, and Walk: The Emerging Smartphone Defense in Personal Injury Litigation”, Albany Law Review, **77**, 2, p. 425 (2014).
- [115] C. H. Basch, D. Ethan, S. Rajan and C. E. Basch: “Technology-related distracted walking behaviours in Manhattan’s most dangerous intersections”, Injury Prevention, **20**, 5, pp. 343–346 (2014).
- [116] 新村, 川西, 出口, 井手, 村瀬, 藤吉: “車載カメラ画像からの「スマホ歩き」認識に基づく歩行者の不注意度推定”, 信学技報, pp. 83–88 (2015).
- [117] G. W. Cottrell and J. Metcalfe: “EMPATH: Face, emotion, and gender recognition using holons”, Proceedings of the 1990 conference on Advances in neural information processing systems 3, pp. 564–571 (1990).
- [118] J. Dong, Q. Chen, W. Xia, Z. Huang and S. Yan: “A Deformable Mixture Parsing Model with Parselets”, Proceedings of the 14th IEEE International Conference on Computer Vision, pp. 3408–3415 (2013).
- [119] J. Zhu, S. Liao, Z. Lei and S. Z. Li: “Improve Pedestrian Attribute Classification by Weighted Interactions from Other Attributes”, Proceedings of the 12th Asian Conference on Computer Vision, pp. w16–p14 (2014).
- [120] A. C. Gallagher and T. Chen: “Estimating Age, Gender, and Identity using First Name Priors”, Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2008).
- [121] H. Chen, A. Gallagher and B. Girod: “Describing clothing by semantic attributes”, Proceedings of the 12th European Conference on Computer Vision, Vol. 3, pp. 609–623 (2012).
- [122] E. Simo-serra, S. Fidler, F. Moreno-noguer and R. Urtasun: “A High Performance CRF Model for Clothes Parsing”, Proceedings of the 12th Asian Conference on Computer Vision, pp. P2–11 (2014).
- [123] N. Zhang, M. Paluri, M. Ranzato, T. Darrell and L. Bourdev: “PANDA: Pose Aligned Networks for Deep Attribute Modeling”, Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1637–1644 (2014).
- [124] 井尻, 勞, 村瀬: “多様な属性に柔軟に対応できる人物属性認識の準教師付き学習フレームワーク”, 信学技報, pp. 97–102 (2009).
- [125] P. Phillips, Hyeonjoon Moon, S. Rizvi and P. Rauss: “The FERET evaluation methodology for face-recognition algorithms”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **22**, 10, pp. 1090–1104 (2000).
- [126] G. B. Huang, M. Ramesh, T. Berg and E. Learned-Miller: “Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments”, University of Massachusetts Technical Report (2007).
- [127] S. Wang, Z. Liu, Z. Wang, G. Wu, P. Shen, S. He and X. Wang: “Analyses of a multimodal spontaneous facial expression database”, Affective Computing, IEEE Transactions on, **4**, 1, pp. 34–46 (2013).
- [128] FG-NET: “The Fg-Net Aging Database” (2014).
- [129] K. R. Jr. and T. Tesafaye: “MORPH: A Longitudinal Image Database of Normal Adult Age-Progression”, Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, pp. 341–345 (2006).
- [130] S. Z. Li, R. Chu, S. Liao and L. Zhang: “Illumination invariant face recognition using near-infrared images”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **29**, 4, pp. 627–639 (2007).
- [131] D. Gray, S. Brennan and H. Tao: “Evaluating appearance models for recognition, reacquisition, and tracking”, Proceedings of the 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Vol. 3, pp. 41–47 (2007).
- [132] A. Ess, B. Leibe, K. Schindler, L. V. Gool and L. Van Gool: “A mobile vision system for robust multi-person tracking”, Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, AK, IEEE, pp. 1–8 (2008).
- [133] G. Sharma and F. Jurie: “Learning discriminative spatial representation for image classification”, Proceedings of the 2011 British Machine Vision Conference, pp. 6.1–6.11 (2011).
- [134] R. Layne and T. Hospedales: “Person Re-identification by Attributes”, Proceedings of the 2012 British Machine Vision Conference (2012).
- [135] J. Zhu, S. Liao, Z. Lei, D. Yi and S. Z. Li: “Pedestrian

- attribute classification in surveillance: Database and evaluation”, Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops, Sydney, NSW, IEEE, pp. 331–338 (2013).
- [136] Y. Deng, P. Luo, C. C. Loy and X. Tang: “Pedestrian Attribute Recognition At Far Distance”, 22nd ACM international conference on Multimedia, pp. 789–792 (2014).
- [137] K. Yamaguchi, S. Brook, T. L. Berg, C. Hill and L. E. Ortiz: “Chic or Social: Visual Popularity Analysis in Online Fashion Networks”, Proceedings of the 22nd ACM International Conference on Multimedia, pp. 773–776 (2014).
- [138] S. Vittayakorn, K. Yamaguchi, A. C. Berg and T. L. Berg: “Runway to Realway: Visual Analysis of Fashion”, Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, IEEE, pp. 951–958 (2015).
- [139] A. Geiger, P. Lenz and R. Urtasun: “Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite”, Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3354–3361 (2012).
- [140] S. Munder and D. M. Gavrila: “An experimental study on pedestrian classification”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **28**, 11, pp. 1863–1868 (2006).
- [141] M. Andriluka, S. Roth and B. Schiele: “Monocular 3d pose estimation and tracking by detection”, Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, IEEE, pp. 623–630 (2010).