

修士論文

車載単眼カメラによる歩行者向き推定
に関する研究

Pedestrian Orientation Detection using
Monocular onboard camera

指導教官 上條 俊介 准教授



東京大学大学院
情報理工学系研究科
電子情報学専攻

学籍番号・氏名 48-126443 矢野 進也

提出日 平成 26 年 2 月 6 日

概要

本論文では、車載単眼カメラを用いた、歩行者の姿勢推定に関する手法について扱う。推定する歩行者の姿勢情報としては、歩行者の身体向き及び頭部向きの二つである。身体向き情報からは歩行者がどの方向へ移動しうる（歩くまたは走る）かを推定でき、頭部向き情報からは歩行者がどの方向に注意を向けているかを推定することができる。提案手法では、向き推定の学習データに対して半教師あり学習を用いて半自動的に学習データを生成することにより、より精度の高い向き推定器を生成することが可能である。また、歩行者の身体向きと頭部向きとに制約を設け、同時に推定を行うことにより身体向き・頭部向きとで判別しにくい向きについても安定した推定を行う手法についても扱う。これらの提案手法について、実際の車載単眼カメラから得られた映像データを用いて評価実験を行い、歩行者の向き推定精度向上に寄与することを確認した。

目次

第1章	序論	1
1.1	研究背景	2
1.2	本研究の目的	3
1.3	本論文の構成	4
第2章	関連研究	5
2.1	2次元姿勢推定	5
2.2	3次元姿勢推定	6
2.3	人体の各部位の向き推定	7
2.3.1	頭部の検出及び向き推定	7
2.3.2	身体の向き推定	9
2.4	本研究の位置づけ	10
第3章	歩行者向き推定	11
3.1	頭部位置・向き推定	11
3.1.1	推定するパラメータ	11
3.1.2	サポートベクトルマシンを用いた頭部向き推定	12
3.1.3	サポートベクトルマシンを用いた頭部位置・向き推定	13
3.1.4	CoHOG 特徴量	14
3.2	身体向き推定	17
第4章	半教師あり学習を用いた向き推定	20
4.1	向き推定における問題点	20
4.1.1	解像度による問題	20
4.1.2	学習データによる問題	21
4.2	解決手法の検討	23
4.2.1	複数人によるアノテーション	23
4.2.2	撮影時に他のセンサを用いたアノテーションの生成	24
4.2.3	仮想的な歩行者モデルを用いた学習データ生成	24
4.2.4	統計的手法を用いたアノテーションの修正	24
4.3	半教師あり学習	25
4.3.1	Self-Training	25
4.4	半教師あり学習を用いた向き学習データのアノテーション	27
4.4.1	学習データのアノテーションの細分化	27
4.4.2	Self-Training による自動アノテーション	28

4.5	評価実験.....	29
4.5.1	実験に用いたデータセット.....	29
4.5.2	実験結果.....	33
第5章	パーティクルフィルタを用いたトラッキング.....	36
5.1	歩行者画像列に対する向き推定.....	36
5.2	パーティクルフィルタ.....	37
5.2.1	パーティクルの設計.....	37
5.2.2	尤度関数の設計.....	38
5.2.3	パーティクルフィルタの更新.....	39
5.2.4	向き推定器の尤度のみを用いた場合の問題点.....	40
5.3	頭部向きと身体向きを考慮した尤度.....	41
5.4	評価実験.....	42
5.4.1	実験に用いたデータセット.....	43
5.4.2	実験結果.....	43
第6章	結論.....	52
6.1	本研究のまとめ.....	52
6.1.1	半教師あり学習を用いた向き推定の高精度化.....	52
6.1.2	頭部向きと身体向きを考慮した尤度関数の設計による向き推定の高精度化.....	52
6.2	今後の課題.....	53
6.2.1	歩行者の向き推定精度.....	53
6.2.2	歩行者の状態推定.....	53
6.2.3	子供の検出, 姿勢推定.....	53
6.2.4	歩行者検出手法との統合.....	54
6.2.5	歩行者のオクルージョン.....	54
謝辞	56
発表文献	57
参考文献	58

目次

図 1	Pictorial Structures による人体の姿勢表現[8]	6
図 2	3次元姿勢推定(左:3次元姿勢モデル 右:3次元姿勢モデルフィッティング) [4]	7
図 3	入力画像からの視線推定[21]	8
図 4	頭部検出と向き推定のフレームワーク[24]	9
図 5	車載カメラ映像からの歩行者向き推定[18]	10
図 6	入力画像(左)と推定する頭部位置・向き(緑矩形・緑矢印), 身体向き(赤矢 印)	11
図 7	8方向に量子化された頭部向き(数値はその方向クラスを表すラベル)	12
図 8	入力画像中の頭部領域の探索	13
図 9	歩行者の頭部探索範囲	14
図 10	CoHOG 特徴量における勾配量子化(a)とその共起ペア(b)[31]	15
図 11	CoHOG 特徴量におけるピクセルのペア(a)と勾配方向の共起(b)およびその共 起行列(c)[31]	16
図 12	CoHOG 特徴量の抽出[31]	16
図 13	CoHOG 特徴量におけるピクセルのペアのオフセット[31]	17
図 14	CoHOG 特徴量におけるペアのオフセットの対称性[31]	17
図 15	8方向に量子化された身体向き(数値はその方向クラスを表すラベル)	18
図 16	入力画像(左)と得られた HOG 特徴量の可視化(右)	19
図 17	HOG 特徴量による歩行者検出器の学習[9]	19
図 18	解像度の違いによる歩行者誤検出率(miss rate)の変化[41]	21
図 19	8方向に量子化した頭部学習データの分布[27]	22
図 20	向きクラス境界付近に存在するデータ(頭部)	22
図 21	向きクラス境界付近に存在するデータ(身体)	23
図 22	学習データとして用いる実際の映像データセット(上段)と仮想データセット (下段)[49]	24
図 23	細分化した学習データの向きクラス:強ラベル(赤)と弱ラベル(緑)	28
図 24	実験における単眼カメラの設置位置	30
図 25	学習データに含まれる強ラベルデータ画像	32
図 26	学習データに含まれる弱ラベルデータ画像	32
図 27	テストデータセットに含まれる画像	33
図 28	頭部向き推定における向き推定精度:人手によるアノテーション(左)と半教	

師あり学習によるアノテーション (右)	34
図 29 身体向き推定における向き推定精度：人手によるアノテーション (左) と半教師あり学習によるアノテーション (右)	34
図 30 実際の車載カメラから得られる歩行者画像列の例 1	36
図 31 実際の車載カメラから得られる歩行者画像列の例 2	37
図 32 推定する 5 つの歩行者パラメータ	38
図 33 複数のパーティクルからの歩行者パラメータ推定	39
図 34 頭部向きと身体向きの制約がない場合の推定	41
図 35 学習データに追加した非頭部画像	43
図 36 正解領域と推定領域の重なり	44
図 37 テストデータにおける頭部向き推定精度：頭部と身体向きの制約なし (左) と制約あり (右)	45
図 38 テストデータにおける身体向き推定精度：頭部と身体向きの制約なし (左) と制約あり (右)	45
図 39 テストデータにおける歩行者の向き推定結果 1	47
図 40 テストデータにおける歩行者の向き推定結果 2	48
図 41 テストデータにおける歩行者の向き推定結果 3	49
図 42 テストデータにおける歩行者の向き推定結果 4	50
図 43 テストデータにおける歩行者の向き推定結果 5	51

表目次

表 1	実験に使用した単眼カメラの詳細.....	30
表 2	学習データに含まれる強ラベルデータ数.....	30
表 3	学習データに含まれる弱ラベルデータ数.....	31
表 4	テストデータに含まれるデータ数.....	31
表 5	テストデータセットにおける隣接ラベルまでを正解とした時の推定精度.....	34
表 6	テストデータにおける頭部重なり率.....	44
表 7	テストデータにおける隣接ラベルまでを正解とした時の向き推定精度.....	45

第1章 序論

本研究では、車載単眼カメラを用いた歩行者の向き推定手法と、向き推定の高精度化に関する手法を扱う。

本章ではまず、本研究の研究背景と目的を明らかにし、その後本論文の構成について述べる。

1.1 研究背景

交通事故には事故を起こした者、被った者双方において大きな社会的損失が存在する。事故を起こした者は社会における信用の消失や物品の破損による物理的・経済的損失がある。また、被害者は場合によっては更に肉体的・精神的苦痛を味わうことになる。このため、交通事故削減のために様々な努力が行われている。しかしながら、日本において平成 24 年に発生した交通事故の件数はおよそ 66 万件であり、うち重傷事故は 4 万件、死亡事故は 4 千件にものぼり、決して少ない数字ではない[1]。

これら重大事故を起こさないためにも、自動車にはいくつかの安全のためのシステムが搭載されている。例えば、衝突時に身体が車外へ投げ出されないように固定するシートベルトや、車内での衝突を和らげるエアバッグなどが挙げられる。これら事故後の被害を最小限にする技術はパッシブセーフティ (Passive Safety) と呼ばれる。一方で、交通事故を未然に防ぐ技術はアクティブセーフティ (Active Safety) と呼ばれる。例えば、急ブレーキによるステアリング操作不能を起こさないためにポンピングブレーキをコンピュータ制御による自動で行うアンチロック・ブレーキシステム (Antilock Brake System: ABS) などが挙げられる。中でもカメラやレーダなどのセンサを用いて歩行者や自動車を検出し、衝突する危険性がある場合は自動的にブレーキ操作を行うことで衝突を回避するプリクラッシュセーフティシステムは、積極的な事故削減につながる技術の中でも有望であろう。プリクラッシュセーフティシステムでは、ドライバーが注意していないあるいは認識していないような障害物に対しても有効になりうるため、対歩行者以外にも有用なシステムである。

プリクラッシュセーフティシステムにおいて最も重要なのはいかにして素早く、正確に障害物を検出するかである。この障害物検出技術、特にその被害の大きさから歩行者を対象を絞った歩行者検出技術では多くの研究が行われている。用いられるセンサとしてはカメラセンサや赤外線センサ、レーザレーダ (LIDAR: Light Detection and Ranging, Laser Image Detection and Ranging)、(無線) レーダなどが挙げられる。カメラセンサは車両の周囲を撮影するように設置され、撮影された映像に対して画像処理を適用することで映像中の歩行者を検出する技術に関する研究が盛んである。ステレオカメラを用いた歩行者検出においては、画像の色情報やテクスチャ情報に加え各画素におけるカメラからの距離 (深度) も得られるため、色情報とテクスチャ情報、深度情報の 3 つを用いた歩行者検出を行うことができる。赤外線センサは赤外領域の光 (赤外線) を捉えることができるセンサで、可視光を用いないため通常のカメラに比べ暗所や夜間でも用いることができるセンサである。しかし、得られる情報は通常単階調の画像であり、また通常のカメラに比べ対象物体の詳細な模様やテクスチャを得ることができないため、赤外線センサを用いた歩行者検出では物体の大まかな形状を用いることが多い。LIDAR はパルス状に発光するレーザを照射し、周囲の物体に反射した際の散乱光を測定するセンサである。発光から反射光を測定するまで

の時間を計測することで周囲の物体までの距離を正確に求めることが可能であり、またレーザに用いられる光束の密度は一般的に非常に高いため、雨滴や煙などを透過して測定することができる。レーダでは電波を周囲に向けて発射し、周囲の物体の反射波を測定する。障害物検知においてはミリ波レーダが用いられることが多い。

これらのセンサの中で最もよく使われるセンサは単眼カメラである。カメラは豊富なテクスチャ情報を含み、どのような歩行者がどの位置に、どういった姿勢でいるかといった詳細な情報まで得ることができる。更に、単眼カメラはステレオカメラに比べ小型化されているため、車内の様々なスペースに取り付けることができる。また、単眼カメラは量産されているため他のセンサよりも安価に手に入れることができる。

カメラから歩行者を検出・認識する研究はコンピュータビジョンの分野の中でも古くから行われてきたテーマの一つである。前述の通り、カメラから得られるテクスチャ情報には本質的に非常に多くの情報が含まれているためである。現実的な車載システムを考えた時、単純に歩行者を検出し事故回避のため停止する以外にも、カメラによる歩行者の姿勢認識などを行うことで歩行者がどの程度危険であるかを推定し、ドライバーに警告を行う、進行方向を変更するといったより状況に応じた適切な決定を下すことが可能となる。特に、都心部における交差点においては運転車両の周囲に歩行者が多数存在し、常に歩行者が検出されるため歩行者検出技術だけでは車載システムとして適切な対応をとることができない。このような場合にも、歩行者の姿勢推定手法を用いれば、検出された歩行者のうち特にどの歩行者が運転車両にとって危険であるかを判定することが可能となるだろう。

しかし、カメラの設置環境や撮影時に加わるノイズ、背景の複雑さ、他の物体による遮蔽（オクルージョン）、対象となる歩行者の外見や姿勢の多様さがあるため、多様な環境下で高速かつ高精度に歩行者を検出・認識することは非常に難しい。カメラを用いた歩行者検出・認識技術に関しては今まで多くの研究がなされてきたが、未だに解決しきれていない問題がある。

1.2 本研究の目的

前節で述べたように、車載カメラを用いた研究としては主に事故防止（アクティブセーフティ）のための歩行者検出・認識技術に関するものが多く行われてきた。本論文で扱う内容はより詳細な歩行者情報であり、都心部など歩行者が非常に多い環境においても実用的な事故防止システムの実現に寄与すると考えられるものとして次の情報を推定する。

- ・歩行者の身体向き
- ・歩行者の頭部位置
- ・歩行者の頭部向き

これら三つの情報は特に歩行者の危険度を推定するために重要な情報である。例えば、推定した歩行者の身体向きからは、歩行者が移動しうる方向を推定することができる。推定された歩行者の移動方向が自動車の前を横切るような場合は、特にその歩行者に注意するべきであるとドライバに警告を促すことができるだろう。また、推定した歩行者の頭部位置・頭部向きからは、歩行者が注意を向けている方向を推定することができる。これにより、特にドライバが運転する自動車に注意を向けていないような歩行者は危険度が高く、逆に気付いている場合は危険度が低い状況であると判断できるだろう。

提案手法では、歩行者の向きを連続量として扱うには非常に多くのデータが必要となり学習が困難であることから、8方向に量子化された向きを考え、歩行者の向きがこのうちのどの向きに属するかを推定する手法について提案する。また、身体向きと頭部向きとに制約を設けることでより高精度に向き推定を行う手法についても扱う。更に、学習データから半自動的に向きラベルを決定することにより、推定精度を向上させる手法についても提案する。最後に、これらの向き推定を各歩行者に対して時間連続で行うことにより、システムとして安定した高精度な推定を行う手法について提案する。提案した手法について、実際の車載映像を用いて評価を行う。

1.3 本論文の構成

本論文は次のように構成される。

本章では、本研究の背景および目的について述べた。

第二章では、本研究に関連する研究として歩行者の姿勢推定について述べる。

第三章では、歩行者画像を入力した際の歩行者の向き推定手法について述べる。

第四章では、歩行者の向き推定手法における問題点について述べた後、より高精度な向き推定器の学習手法について述べる。また、提案する高精度な向き推定器の学習手法をテストデータセットを用いて評価する。

第五章では、歩行者の向き推定手法を実際の歩行者画像列に対して適用する際にパーティクルフィルタを用いた手法について述べる。更にパーティクルフィルタに歩行者の頭部向きと身体向きとの拘束を考慮することによって、より安定した高精度な向き推定が行える手法についても述べる。

第六章では本研究の結論を述べる。

第2章 関連研究

人物の姿勢推定に関する研究は、撮影するカメラや背景などの環境、対象となる人物の多様な服装や姿勢などの要因があるため、非常に難しいタスクの一つである。しかし、それらの姿勢情報を応用したアプリケーションとしてはヒューマンインタラクションや人物の行動分析、より高精度な人物検出などが考えられるため、解決すべき重要な研究分野の一つでもある。人物の姿勢推定は人物検出と同様にコンピュータビジョンの分野では多くの研究が行われてきた[2,3,4,5,6,7,8,11,12,13,14,15]。一口に姿勢推定と言っても、それぞれの研究が対象とする問題によって実際に推定する内容については様々である。本章では、これらの研究について述べる。

2.1 2次元姿勢推定

2次元姿勢推定は、入力画像中における人物の各部位の位置や角度などを推定する手法である。人物の各部位の表現方法としては、人物の各部位とそれらの接続をモデル化した **Pictorial Structures**[7,8]が用いられることが多い。人物の各部位としては主に頭部や腕、胴体、脚部などが挙げられ、部位間の接続は頭部から胴体、胴体から腕など、実際の人体構造に基づいた設計がなされる。**Pictorial Structures** を表現するデータ構造として無向グラフや木構造が用いられることが一般的である。それらのデータ構造を使った **Pictorial Structures** では各部位はノードとして、各部位の接続はノード間のエッジとして表現される。更に、エッジ毎に重みを変えることで、各ノードが変形した際に課されるペナルティを調整することができるため、より自然な人物の姿勢推定が行える。人体の各部位は矩形を用いて表現されるのが一般的である（図 1）。

Fischler ら[7]が述べた **Pictorial Structures** モデルのパラメータ推定では、背景差分により抽出した前景部分とモデルとの重複率が用いられていたが、近年では HOG 特徴量[9]などを用いてアピランスを基にしたパラメータ推定[11]や、より部位間の関係に注目したパラメータ推定手法[14,15]、より細かな部位の変形も考慮したパラメータ推定手法[12,13]などが提案されている。

Pictorial Structures を用いた研究としては、前述した姿勢推定自体の高精度化を図る研究の他、歩行者検出での検出精度を向上させる研究などが挙げられる。Ronford ら[10]は非常に単純な画像特徴量による部位検出器と **Pictorial Structures** を用いた歩行者検出手法を提案した。Andriluka ら[2,3]は **Pictorial Structures** を基にした人体の各部位に基づいたモデルを歩行者検出に用いることでより高精度に歩行者を検出・追跡する手法を提案した。また、**Pictorial Structures** では通常頭部や胴体といった推定する部位は予め人により定め

られるが、Felzenszwalb ら[5,6]は学習データセットから自動的に識別しやすい部位を決定することで高精度に歩行者検出を行う手法を提案した。しかし、この手法では検出に用いる部位を自動的に決定してしまうため、検出された部位の特定や、人物の意味的な姿勢を推定することなどはできない。

このように、2次元姿勢推定に関する研究では非常に多くの手法が提案されている。

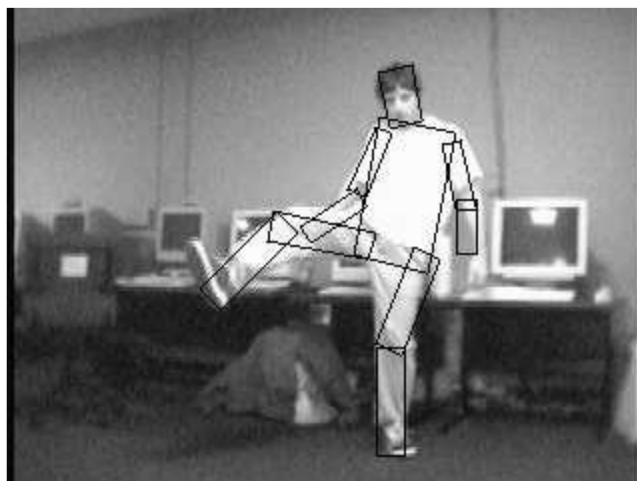


図 1 Pictorial Structures による人体の姿勢表現[8]

2.2 3次元姿勢推定

前節では人物の姿勢推定の一つとして、画像内における人物の部位の位置や角度といった情報抽出の手法について述べた。しかし、それらは画像内における姿勢情報であり、実空間における人物の姿勢とは一致しない。本節で述べる 3次元姿勢推定は、実空間における人物の姿勢を推定する研究であり、人物の姿勢推定の最終目標であると言える。しかし、2次元情報である画像から 3次元の姿勢情報を推定することは本質的に解が曖昧にならざるをえないため、2次元姿勢推定よりも非常に困難なタスクである。また、人物自身のオクルージョンに観測が不可能であるような部位も起こりうるという問題もある。そのため、入力画像から人物の 3次元姿勢を推定する際は画像内の詳細なテクスチャ情報を用いた制約、複数カメラを用いた制約、時間軸での解析による制約、撮影環境による制約（カメラの画角や設置位置、撮影方向）など多くの制約条件を課すことにより求める。それら制約条件を定めるために、従来では研究室や実験室など限られた環境での 3次元姿勢推定に関する研究が多く行われている。Gall ら[17]は複数のカメラを用いることで人物の 3次元姿勢を推定する手法を提案している。Vondrak ら[18]は人体のモデルに対して物理シミュレーションを用いた 3次元姿勢推定手法を提案している。Balan ら[16]は撮影環境を限定した上で背景差分を用いて人物の正確なシルエットを計算し、予め用意しておいた詳細な 3次元モデルを

シルエットにフィッティングさせることで人物の 3 次元姿勢を推定する手法を提案している。

一方、近年限られた環境以外での人物の 3 次元姿勢推定手法も提案されはじめている。Andriluka ら[4]はまず入力画像中の歩行者を検出し、歩行者の 2 次元姿勢推定と向き推定を行った後、それらを制約条件とし更に 3 次元姿勢の部位間の制約条件を加えることで 3 次元姿勢推定を行う手法を提案している (図 2)。

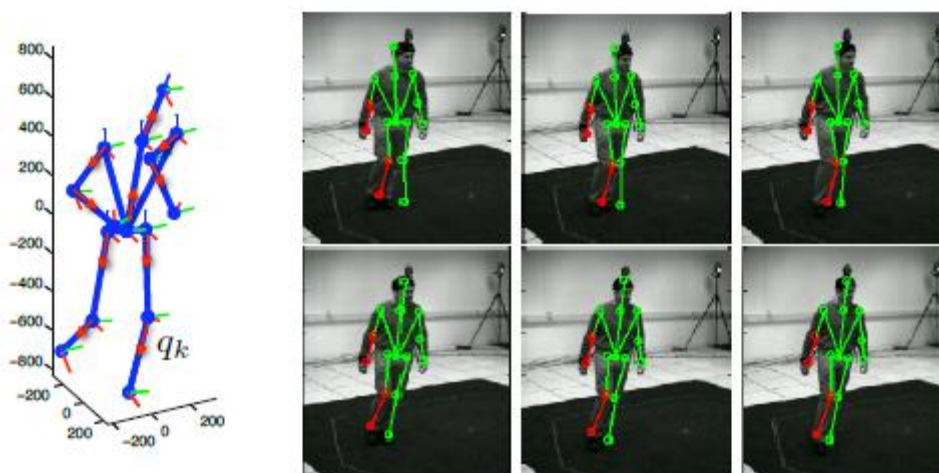


図 2 3 次元姿勢推定 (左 : 3 次元姿勢モデル 右 : 3 次元姿勢モデルフィッティング) [4]

2.3 人体の各部位の向き推定

前節までは人物の全身の姿勢推定手法について述べた。本節では、人体の各部位の検出及び向き推定について述べる。ここで主に検出・推定の対象となるのは頭部や胴体部分などである。

2.3.1 頭部の検出及び向き推定

人体の各部位を検出・推定する際、特に頭部に関しては人物の表情などの情報が含まれることから、ヒューマンインタラクション分野をはじめ応用先が考えられる。Viola ら[19,20]が提案した手法では、Haar-like 特徴量とカスケードされた識別器 (AdaBoost) を用いることで高速に頭部を検出することができる。この手法はその検出精度の高さと高速性からその後の頭部検出の研究に大きく影響を及ぼし、現在ではこの技術が組み込まれた製品も一般に普及し始めている。

一方、検出した後より詳細な頭部の情報を抽出する研究としては、頭部の視線推定や向き

推定などが挙げられる。頭部の視線推定では主に顔のテクスチャ情報や瞳孔の情報を用いるため、現時点では頭部画像が高解像度に取得できる場合に限られる。Smith ら[22]は車両運転者の視線推定のため、運転者の頭部を撮影するカメラから運転者の顔領域を検出した後に瞳孔部分を検出する手法を提案している。Qiang ら[21]は、ダッシュボードに設置したカメラから運転者の頭部を撮影し、運転者の瞳を画像内から探索・検出する手法を提案している（図 3）。この手法では運転者の頭部の 3 次元的な角度や鼻や口といった顔の目以外の部位も推定可能である。



図 3 入力画像からの視線推定[21]

一方頭部の向き推定では、頭部の向いている大まかな方向を推定する研究が多く、瞳孔などのテクスチャが満足に得られないような低解像度の画像に適用されることが多い。Murphy ら[23]は車両運転者の補助システムのため、運転者の頭部を撮影するカメラから得られた映像から顔検出をした後、LGO 特徴量を用いて運転者の顔の左右・上下の角度を推定する手法を提案している。Schulz ら[24,27]は運転者ではなく、車両付近の歩行者の頭部向きとその位置を頭部姿勢推定器によって同時に推定し、更にパーティクルフィルタを用いることで時間連続的に安定して推定する手法を提案している（図 4）。頭部位置の探索は歩行者画像の上部に限定することで、不自然な位置に頭部が推定されることはない。頭部の向き推定において、頭部のテクスチャが十分な解像度で得られない場合、テクスチャのみを頼った向き推定は困難である。そこで、人物の移動方向を用いて推定を行う手法なども提案されている。Robertson ら[26]は映像中の人物の頭部肌部分と人物の速度情報を考慮することで頭部の向き推定を行う手法を提案している。

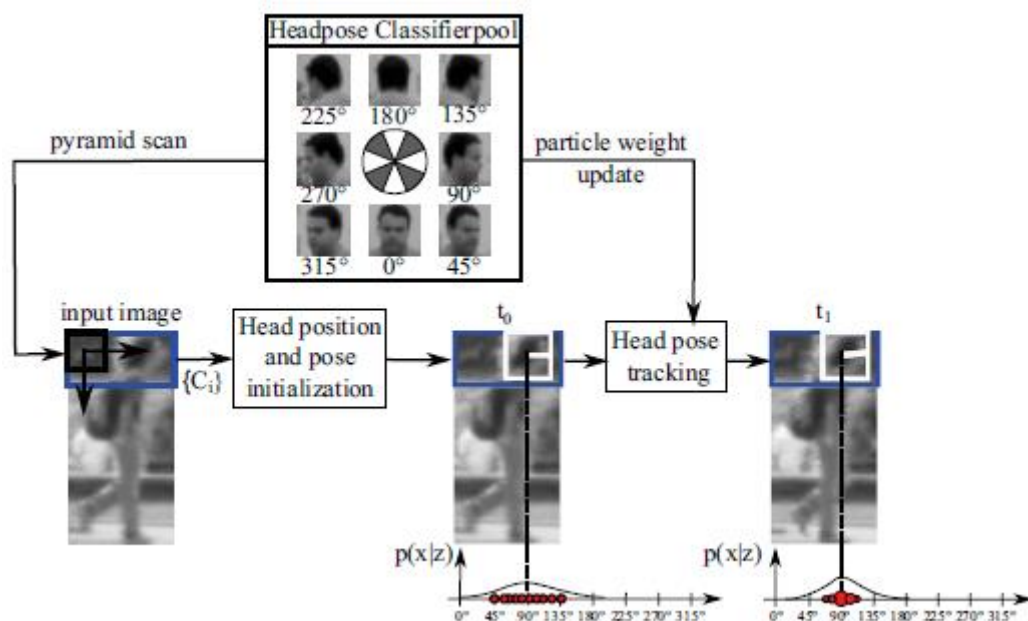


図 4 頭部検出と向き推定のフレームワーク [24]

2.3.2 身体の向き推定

頭部以外の部位の向き推定としては、主に胴体、ひいては身体全体が向き推定の対象となる。身体全体の向き推定では対象人物の進行方向を推定することが可能である。また、身体全体の検出と向き推定とを同時に行うことで、複数の向きも考慮しつつ人物を検出することができるため、人物検出を目的とした研究も行われている。Enzweiler ら[25]は歩行者の向きを前後左右の4つにカテゴリ分けした学習データからそれぞれ識別器を学習させることで、より高精度に歩行者検出を行う手法を提案している。Gandhi ら[28]は歩行者の向きを8方向にカテゴリ分けし、HOG 特徴量と SVM を用いて歩行者の向きを推定し、更に隠れマルコフモデルを用いることにより時系列に沿って歩行者の進行方向を推定する手法を提案している (図 5)。



図 5 車載カメラ映像からの歩行者向き推定[18]

2.4 本研究の位置づけ

前節まで述べてきたように，人物の姿勢推定は画像内の人物の各部位の位置と角度を推定する 2 次元姿勢推定と，実世界における人物の 3 次元姿勢推定，更に人物自体の向き推定といった研究内容に分けることができる．本研究ではこれら人物の姿勢推定技術のうち，主に交通シーンを対象とした向き推定手法について扱う．

従来は交通シーンを対象とした歩行者検出技術の研究が多く行われてきた [2,3,4,9,25,31,37,38,39,51,52,53]．これは，交通事故の削減を目的として車両の周囲の歩行者を検出し，車両の至近距離に歩行者が存在する場合は自動的に減速するなど，純粋なプリクラッシュセーフティシステムのための技術である．しかし，特に都心部における交差点など，車両の周囲に多数の歩行者が存在するような環境においては，常に歩行者が検出されてしまうため停止状態が続いてしまうなど，システムとして適切な対応がとれないという問題が考えられる．

交通シーンにおける歩行者の身体向き情報からはその歩行者の進路方向を推定することが可能である．また，歩行者の頭部向き情報では，その歩行者がどの方向へ注意を向けているかなどといった推定を行うことが可能である．そのため，歩行者の身体向きと頭部向きはいずれも重要な情報である．そしてこのような情報は，車両の周囲に多数の歩行者が存在するような環境におけるアクティブセーフティやプリクラッシュセーフティなどのシステムの観点から非常に重要であると考えられる．なぜなら，これら歩行者の身体・頭部向き情報が扱える場合，アクティブセーフティのシステムは各歩行者が自車両にとってどの程度危険であるかが計算できるため，歩行者を検出した場合に停止以外のより実践的な対応がとれるシステムを開発できると考えられるからである．そこで本研究では，こうした先進的なアクティブセーフティのシステムに寄与する技術として，交通シーンを対象とした歩行者の身体及び頭部向き推定手法について扱う．

第3章 歩行者向き推定

本章では一枚の歩行者画像を入力とした歩行者の向き推定手法について述べる。推定する向きとしては歩行者の頭部向きと身体向きの二つである。しかし、入力には歩行者の画像のみであるため、歩行者の頭部向きを推定するにあたって歩行者の頭部位置も同時に推定する必要がある。したがって、提案手法が求めるのは一枚の歩行者画像を入力として、歩行者の頭部位置、頭部向き、身体向きの三つである（図 6）。

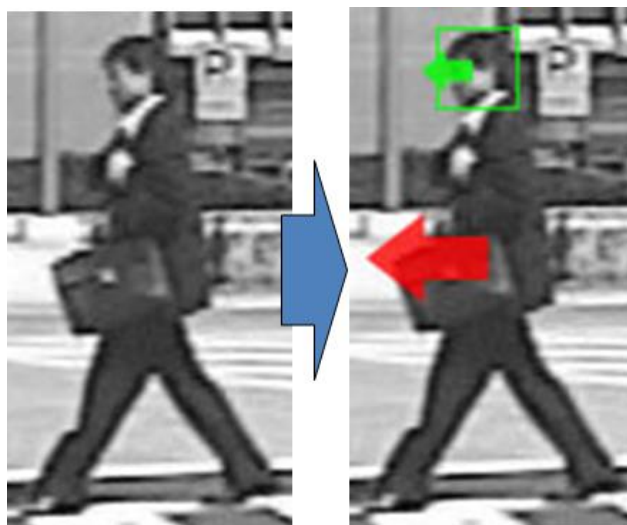


図 6 入力画像（左）と推定する頭部位置・向き（緑矩形・緑矢印）、身体向き（赤矢印）

3.1 頭部位置・向き推定

本節では、頭部位置及び向き推定の手法について述べる。

3.1.1 推定するパラメータ

頭部位置・向き推定で推定するパラメータは、入力画像中における頭部の位置 (x, y) 及び頭部の大きさ s 、また頭部の向き d_h の 4 つである。頭部の領域は画像中で正方形であると仮定しているため、頭部の大きさを表すパラメータは 1 つである。また、頭部向き d_h は図 7 のように 360 度を 8 方向に量子化されており、頭部向き推定においては入力画像がこの 8 方向のクラスのうちのどの方向に属するかを推定する。

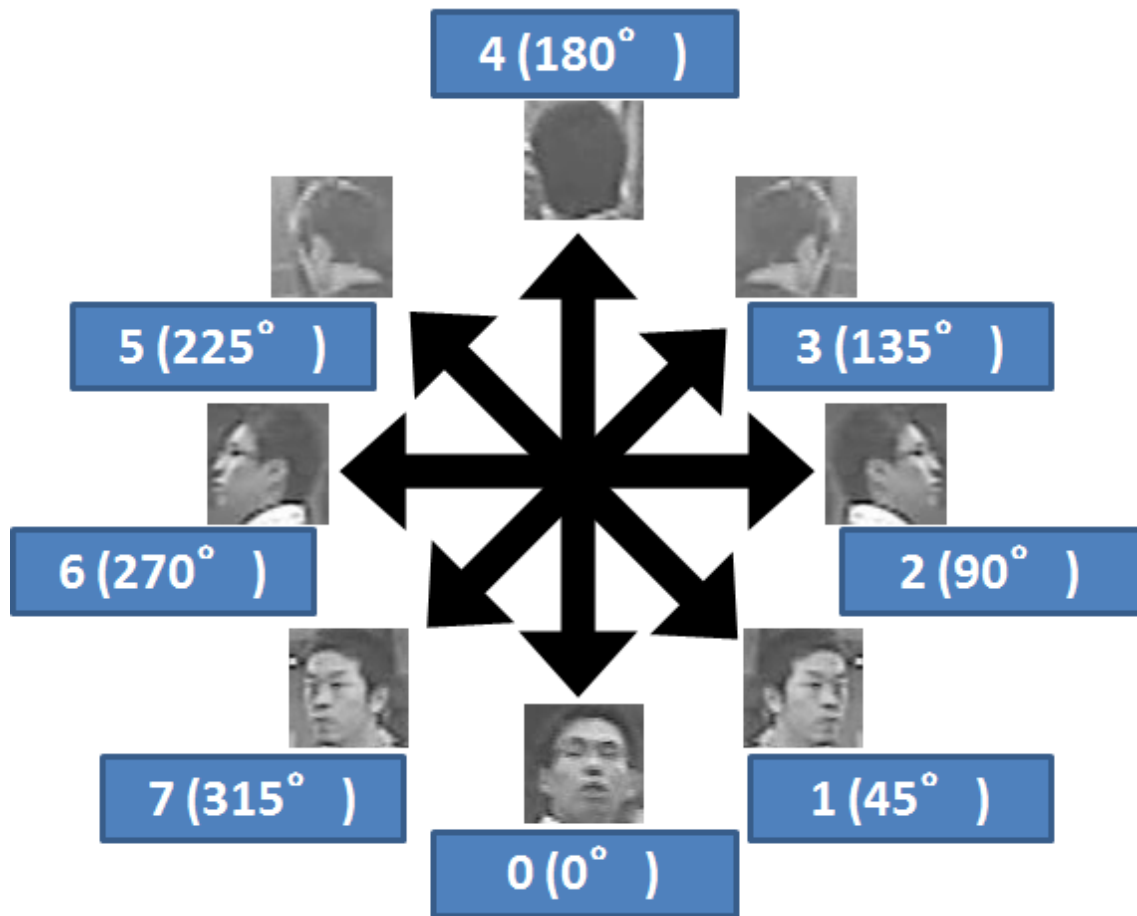


図 7 8方向に量子化された頭部向き（数値はその方向クラスを表すラベル）

3.1.2 サポートベクトルマシンを用いた頭部向き推定

頭部の向きを推定するにあたっては、サポートベクトルマシン[29]による多クラス分類器を用いた。サポートベクトルマシンはベクトルを入力とした分類器であるため、入力画像からサポートベクトルマシンが扱えるようにベクトルとして抽出する必要がある。これらは画像特徴量と呼ばれるが、本提案手法では頭部領域の画像特徴量として HOG 特徴量をベースとした CoHOG 特徴量[31]を用いた。CoHOG 特徴量については 3.1.4 節で述べる。

入力画像から画像特徴量を抽出した後、得られた画像特徴量が予め学習した多クラス分類器を用いてどの方向クラスに属するかを求める。より具体的には、学習した多クラス分類器は入力ベクトルに対して各クラスの尤度を与える。頭部の領域を表すパラメータは頭部位置 (x,y) 及び頭部の大きさ s であるため、入力画像を I としたとき各方向クラス d_h の尤度は次の式で表すことができる。

$$P(d_h|I, x, y, s) \quad (3-1)$$

ある1つの頭部パラメータに対して，その頭部領域の向き d_h^* は最も高い尤度を持つ方向クラスと推定することができる．すなわち，

$$d_h^* = \arg \max_{d_h} P(d_h | I, x, y, s) \quad (3-2)$$

を満たすような d_h^* が，ある1つの頭部領域の向きであると推定できる．

3.1.3 サポートベクトルマシンを用いた頭部位置・向き推定

前節ではある1つの頭部領域に対してその向きを推定する手法について述べた．しかし，与えられた入力画像から頭部位置 (x, y) 及び頭部の大きさ s を推定する際には，入力画像中の頭部領域となる候補から探索する必要がある．最も単純には，図8に示すように，画像中の各位置 (x, y) と可能な大きさ s から得られる全ての領域について多クラス分類器を用いて分類を行う必要がある．



図8 入力画像中の頭部領域の探索

求める頭部パラメータ H を次のように定義する．

$$H = [x, y, s, d_h]^T \quad (3-3)$$

このとき，推定された頭部パラメータ H^* は，画像中の部分領域のうち頭部として最大の尤度を持つ領域とした場合，頭部向き推定の多クラス分類器を用いて次のように表せる．

$$H^* = \arg \max_H P(d_h | l, x, y, s) \quad (3-4)$$

一般的なオブジェクト検出の分野では、画像中の対象物体の検出する際は検出された全ての領域に対して統合処理を行い最終的な物体検出領域とする手法がとられる[9,20,32]。しかしここでは、検出される頭部領域の解像度が低いため正解頭部領域のまわりに頭部領域が検出されるとは限らないこと、また、頭部領域の検出と同時に向き推定も行っていることから、最も高い尤度を持つ領域を最終的な推定された頭部領域とする。頭部位置を推定する際、正解頭部領域以外では背景や歩行者の胴体部分など、非頭部領域も探索対称として含まれるため、実際にサポートベクトルマシンを用いた多クラス分類器を学習させる際は図7に示した8方向クラスの頭部画像の他、非頭部クラスの画像も含まれる。

また、頭部領域は歩行者画像の上部に存在すると考えられるため、頭部領域の探索範囲は歩行者画像の上部のみに限定する(図9)。さらに、現実的な歩行者の頭部大きさを考慮して、歩行者をおよそ5頭身から8頭身と仮定している(例えば歩行者画像の高さが128ピクセルであった場合、頭部の大きさは16~24ピクセルとなる)。



図 9 歩行者の頭部探索範囲

3.1.4 CoHOG 特徴量

入力画像からサポートベクトルマシンが扱えるベクトルとして画像特徴量を抽出する際、用いる画像特徴量を適切に選択することは向き推定精度の観点から重要である。従来手法における顔検出では Haar-like 特徴量[19,20]や LBP 特徴量[33,34,35,36]といった画像特徴量が用いられてきた。しかし、これらの画像特徴量は主としてカメラ方向に頭部を向けてい

る頭部に限って検出を行うのに適した画像特徴量である．そのため，8方向の頭部向きを扱う本提案手法に適用するのは難しいと考えられる．そこで，本提案手法では頭部の前後向きを判別しやすいテクスチャ情報の他，頭部の左右向きを判別しやすい輪郭情報も取り入れるため HOG 特徴量をベースとした特徴量として CoHOG 特徴量を用いる．

CoHOG 特徴量は Watanabe ら[31]が提案した HOG 特徴量をベースにした特徴量で，ピクセル単位の勾配情報と周囲のピクセルの勾配との共起を記述することができ，特に低解像度における歩行者検出などに有効である．テクスチャ情報と輪郭情報を取り入れたピクセル単位の勾配情報を用いていることと，さらにそれらの共起を考慮することで低解像度においても非常に詳細な情報を記述することができる．そのため，本提案手法が対象とするような低解像度の頭部画像においても高い分類精度が期待できるため，CoHOG 特徴量を選択した．本節では CoHOG 特徴量について述べる．

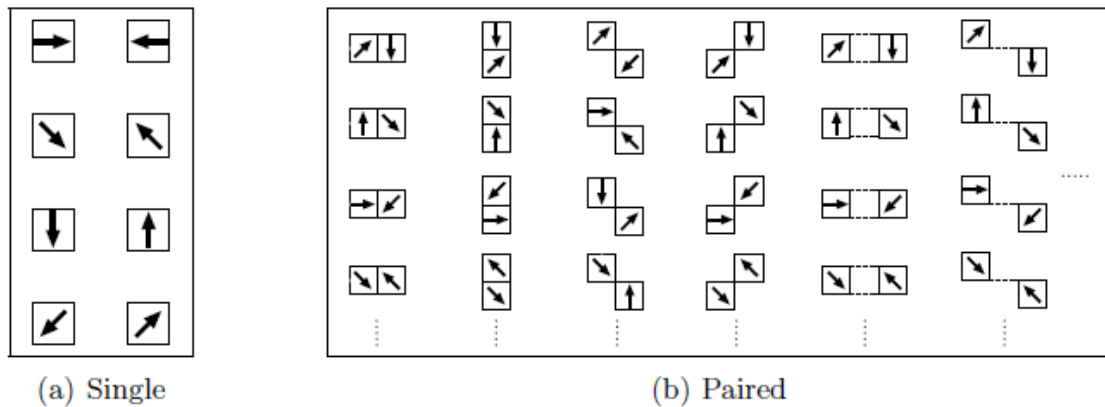


図 10 CoHOG 特徴量における勾配量子化 (a) とその共起ペア (b) [31]

CoHOG 特徴量では，まずピクセル毎に輝度値の勾配を計算し，それを 8 方向に量子化する (図 10(a))．さらに得られたピクセルごとの勾配方向を周囲のピクセルの勾配方向とペアとして扱うことで，勾配方向の共起を記述することができる (図 10(b))．ペアとなるピクセルは相対位置 (オフセット) が異なる場合は異なるペアとして扱う．

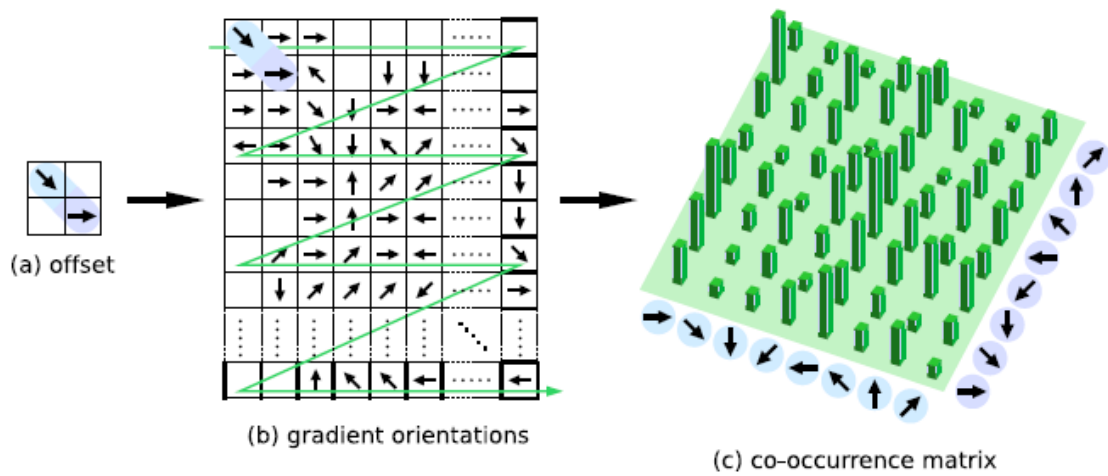


図 11 CoHOG 特徴量におけるピクセルのペア(a)と勾配方向の共起(b)およびその共起行列(c) [31]

あるオフセットのピクセルペアに対して、その勾配方向を投票する 8×8 の行列（共起行列）を考える。この共起行列に対して、各ピクセルペアとそのピクセルペアの勾配方向から、対応する行列要素に投票（1を加算）する。そのため、得られる共起行列では元画像中において勾配方向の共起が多かった要素の値が大きくなる（図 11）。

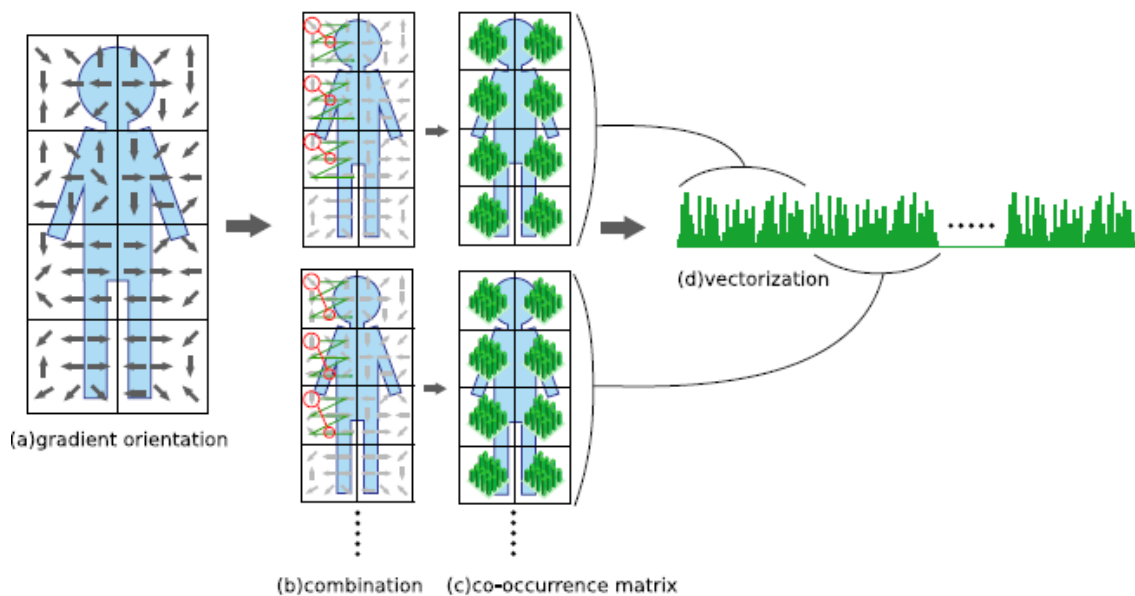


図 12 CoHOG 特徴量の抽出[31]

元画像を複数のブロックに分割した後、それぞれのブロック毎に共起行列を計算し、共起行列の要素をベクトルとして並べたものが CoHOG 特徴量である（図 12）。ここで、単純

にあるピクセルと画像内の他のピクセルとをペアとして共起を計算しようとする膨大な計算時間が必要であり、また得られる特徴量の次元数も非常に大きなものになってしまう。そこで、CoHOG 特徴量ではペアとなるピクセルのオフセットは図 13 のように制限されている。

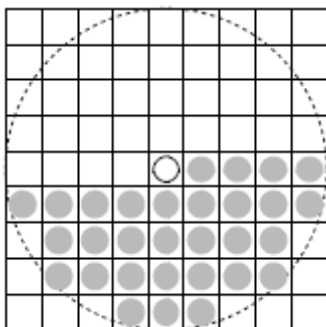


図 13 CoHOG 特徴量におけるピクセルのペアのオフセット[31]

図 13 において円の中心の白い丸は対象としているピクセルを、円内の他の黒い丸はそのペアとなりうるピクセルを表している。CoHOG 特徴量ではピクセルのペアとして 30 個のピクセルを用いている。図 13 から確認できるように、CoHOG 特徴量ではピクセルのペアは対象となるピクセルより下の領域にのみ限定している。しかし、CoHOG 特徴量のピクセルペアによる投票は画像中の各ピクセルに対して行われるため、対象となるピクセルより上の領域に存在するピクセルとの共起は、上の領域に存在するピクセルが対象ピクセルとなっているときの共起として共起行列に投票される (図 14)。

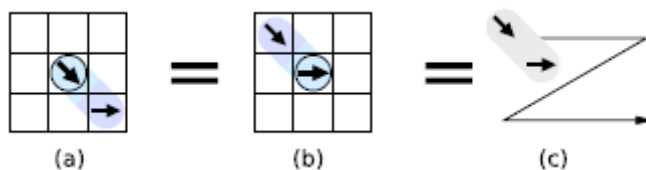


図 14 CoHOG 特徴量におけるペアのオフセットの対称性[31]

3.2 身体向き推定

本節では入力画像中の歩行者の身体向き推定手法について述べる。歩行者の入力画像が与えられた時、頭部向き推定の時と同様にサポートベクトルマシンを使った多クラス分類器を用いて身体向きを推定する。推定する身体向きは 360 度を 8 方向に量子化した 8 つの方向クラスとする (図 15)。

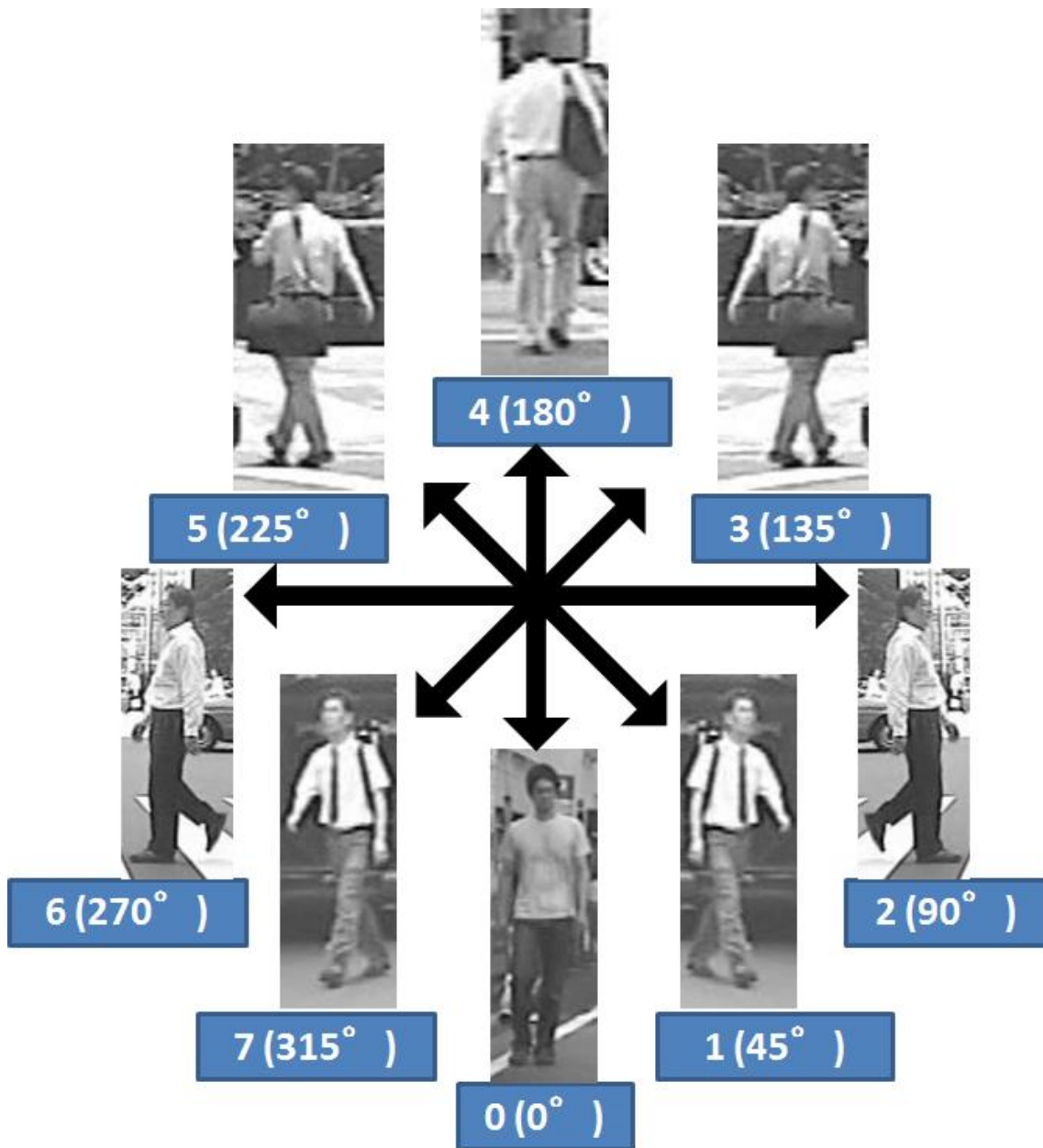


図 15 8 方向に量子化された身体向き（数値はその方向クラスを表すラベル）

身体向き推定では，頭部向き推定の時とは異なり入力として歩行者画像が与えられているため位置推定を行う必要がない．そのため，身体向き推定で推定するパラメータは身体向き d_b のみである．推定された身体向き d_b^* は，サポートベクトルマシンを使った多クラス分類器と入力画像 I を用いて次のように表せる．

$$d_b^* = \arg \max_{d_b} P(d_b|I) \quad (3-5)$$

身体向き推定を行う際は画像特徴量として HOG 特徴量[9]を用いた。HOG 特徴量は Dalal ら[9]によって提案された画像特徴量の 1 つで、画像の大まかな勾配情報を記述することができる特徴量である (図 16)。対象となる様々な姿勢や服装が考えられる歩行者の大まかな勾配情報を記述することで、特に歩行者検出に対して有効であることが確認されている (図 17)。

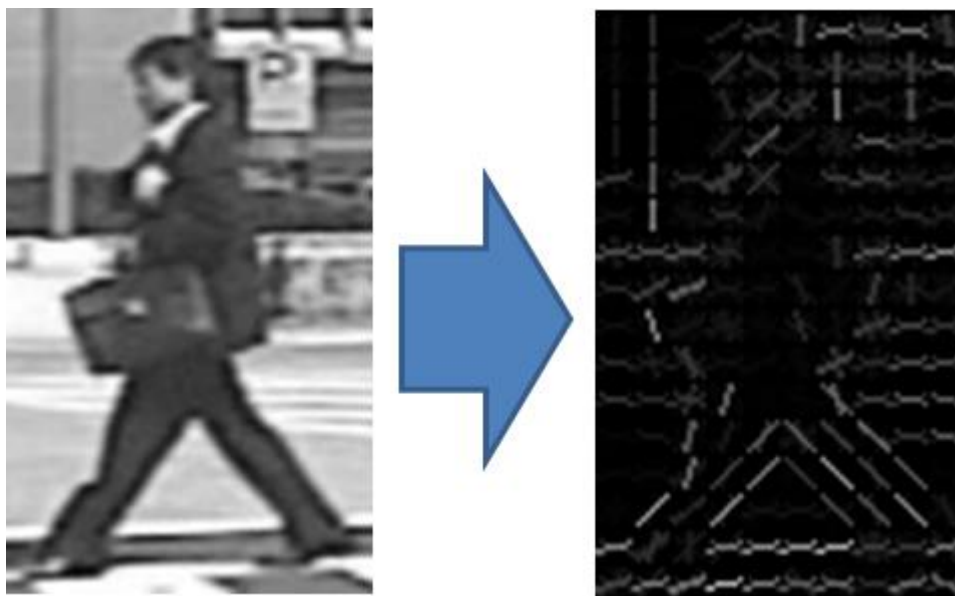


図 16 入力画像 (左) と得られた HOG 特徴量の可視化 (右)



図 17 HOG 特徴量による歩行者検出器の学習[9]

第4章 半教師あり学習を用いた向き推定

前章では歩行者画像を入力とした歩行者の頭部位置・向き推定手法と身体向き推定手法について述べた。本章では、特に歩行者の頭部・身体向き推定における問題について述べた後、それに対する解決手法として半教師あり学習を用いた向き推定器を構築する手法について述べる。

4.1 向き推定における問題点

向き推定においては、いくつか問題点が挙げられる。本節ではそれら考えられる問題点について述べる。

4.1.1 解像度による問題

向き推定の対象となる画像は様々な解像度が考えられるが、第2章で述べたように高解像度の場合は向きだけでなく対象人物の2次元姿勢推定や3次元姿勢推定といった手法がとられることが多い。こうした高解像度な画像を得るためには、対象人物の至近距離にカメラを設置する、あるいは高解像度な画像を撮影可能な特別なカメラを用いる必要がある。しかし、本研究が対象とする交通シーンにおいては対象人物の至近距離にカメラを設置することは公共空間のため難しく、また高解像度な画像を撮影可能なカメラは非常に高価であるため実際のインフラや車載デバイスを考えた際のコストとして無視することはできない。そのため、交通シーンにおいて得られる歩行者画像は低解像度であることが多い。

入力画像から向きを推定する際は、十分な解像度がない場合正確な推定は非常に難しい。Dollarら[40,41]は歩行者検出において、従来の歩行者検出手法は解像度の違いによって検出精度が大きく異なる点について指摘している(図18)。入力が歩行者画像のため、向き推定においても同様の問題が存在すると考えられる。特に頭部の解像度は歩行者全体の解像度から見ても非常に小さいため、高精度に向き推定を行うのは難しいと考えられる。

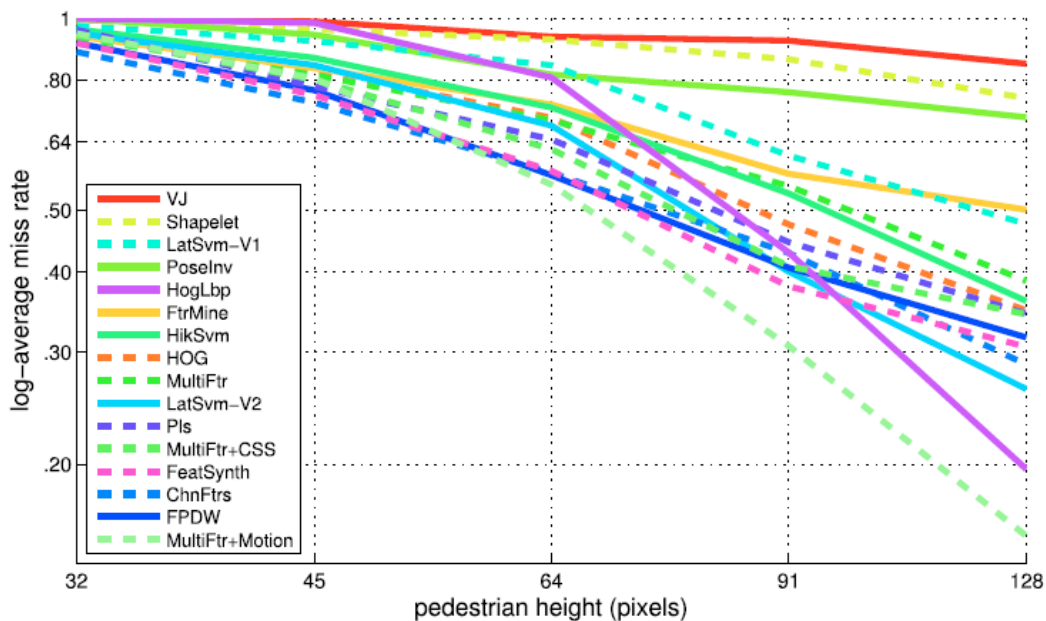


図 18 解像度の違いによる歩行者誤検出率 (miss rate) の変化[41]

4.1.2 学習データによる問題

向き推定に関するもう一つの問題点は、学習データ自体を生成することが難しいことである。本提案手法が扱う向き推定では、図 7 と図 15 に示したように、入力画像が 8 方向に量子化した方向クラスのいずれに属するかを推定する処理である。そのため、学習データとしては多数の歩行者画像を、各方向クラスのいずれに属するかを手によりアノテーションする必要がある。ここで、歩行者の向きをアノテーションする際、人の目による認識においても歩行者の向きを正しくアノテーションすることが難しいという問題が存在する。Schulz ら[27]は 8 方向に量子化した頭部向きを推定する分類器を生成する際、手によりアノテーションした学習データの分布を可視化した結果、学習データが曖昧にならざるを得ない問題点について指摘している (図 19)。特に隣接する方向クラスに分布するような学習データも存在するため、ブースティングを用いた外れ値に対して頑強な識別器を生成してしまうと、向き推定においては高精度な推定が行えないという問題についても指摘している。これらアノテーションが難しいデータは、特に方向クラスの境界付近に存在するデータであることが多い (図 20, 図 21)。

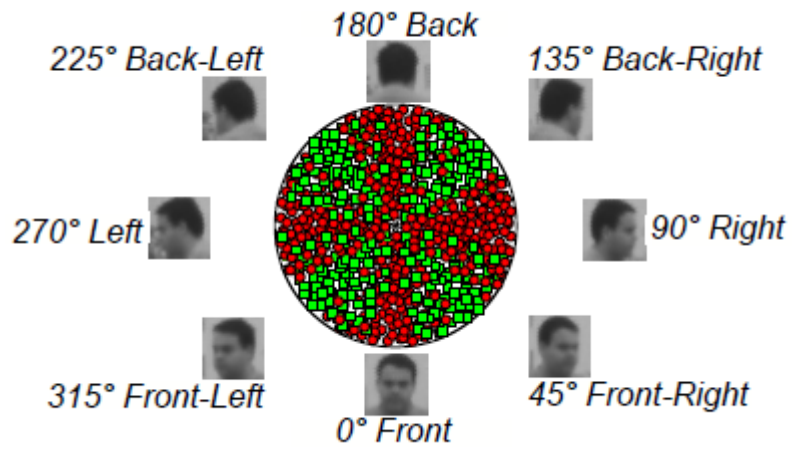


図 19 8方向に量子化した頭部学習データの分布[27]



図 20 向きクラス境界付近に存在するデータ (頭部)



図 21 向きクラス境界付近に存在するデータ (身体)

4.2 解決手法の検討

前節では歩行者の向き推定をするにあたって、考えられる問題点を述べた。このうち、解像度による問題は、解像度が高い画像に比べ本質的に含まれる情報が限られていることから、非常に困難な課題であると考えられる。一方、学習データ自体が持つ曖昧性に関する問題は、いくつかの解決手法が考えられる。

4.2.1 複数人によるアノテーション

1つは、学習データのアノテーションを複数人によって行う手法である。これは複数人で学習データのアノテーションすることで、各学習データに対してより一般的な、より多くの人が推定する向きラベルが得られるため、個人で学習データのアノテーションした場合よりも安定した学習データが得られると考えられる。しかし、この手法においては各学習データに対してつけられた複数の向きアノテーションから、最終的な向きアノテーションを決定する必要があるため、本質的な曖昧性を完全に除去することはできていないと考えられる。

4.2.2 撮影時に他のセンサを用いたアノテーションの生成

他には、学習データとなる画像を撮影する際、対象となる歩行者の正確な情報を他のセンサやマーカを用いて計測しておき、その情報を元に向きアノテーションを付加する手法である。向き情報は他のセンサから高精度に得られるため、向きアノテーションとして非常に正確なものが得られ、かつ曖昧性は存在しない。しかし、こうした学習データを生成するには特殊な環境が必要であり、多くの学習データを収集することは困難であると考えられる。

4.2.3 仮想的な歩行者モデルを用いた学習データ生成

人体のモデリングソフトや仮想的な歩行者を用いて、任意の視点からの描画結果を画像とすることで自動的に学習データを生成するような手法も考えられる。Marin ら[49]やVazquez ら[50]は仮想的な人物画像を学習データとして HOG 特徴量やサポートベクトルマシンなどを用いた歩行者検出手法を提案している (図 22)。しかしこの手法では、人体のモデリングを用いているため実際の歩行者との外見や姿勢の違いがあること、実際の歩行者に見た目を近づけるために照明環境や服装といったテクスチャなど多数のパラメータを変化させる必要がある。また、専用の人体モデルなども多数用意する必要があるため、本研究では扱わなかった。

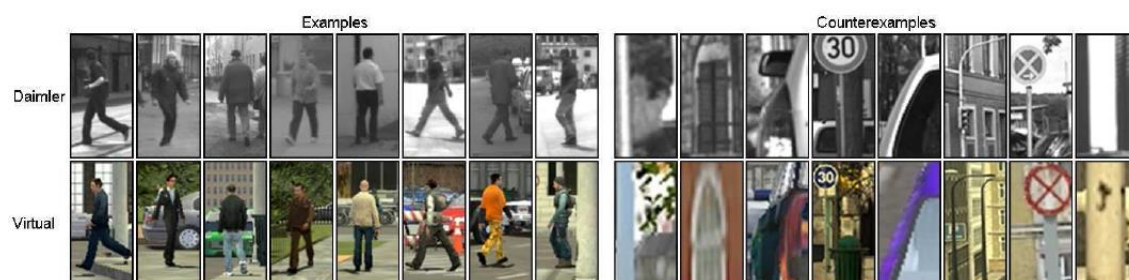


図 22 学習データとして用いる実際の映像データセット (上段) と仮想データセット (下段) [49]

4.2.4 統計的手法を用いたアノテーションの修正

他の考えられるアプローチとしては、人手によってアノテーションした学習データに対して統計的手法を用いてアノテーションを修正する手法がある。用いる統計的手法としてはまずクラスタリングが考えられる。しかし、クラスタリングは学習データの分布を効率よく表現するのに適した手法であるが、得られた分布がどの向きに対応するかといった情

報を抽出することはできない。他に考えられる統計手法としては半教師あり学習が挙げられる[42,43,44,45,46,47,48]。半教師あり学習ではラベル有りデータとラベル無しデータの両方を用いて最終的な識別器を生成する手法である。

本研究で扱う 8 方向の向きクラス学習データを生成する際、特に問題となるのは 8 方向クラスの境界付近に分布するような学習データである。そこで、本研究では目視により簡単に 8 方向クラスに分類できる学習データはラベル有りデータ、目視により簡単に 8 方向クラスに分類できないような学習データをラベル無しデータとし、半教師あり学習によって向き推定器を生成する手法について提案する。提案手法では、従来問題となってきた目視により簡単にクラス分類できないような学習データについて、半教師あり学習が自動的にラベルをアノテーションし学習することにより、学習データにおけるアノテーションの曖昧性を排除すると同時に、より学習データに適したラベルをアノテーションできると考えられる。

4.3 半教師あり学習

半教師あり学習はラベル有りデータとラベル無しデータの両方を用いた学習手法である。半教師あり学習は主にラベル有りデータを大量に集めるためにコストがかかりすぎる問題や、ラベル有りデータのアノテーション自体が困難な問題を解決するために用いられる。本研究で対象としている 8 方向の向き推定では、各学習データに適切なラベルをアノテーションすることが困難な問題と考えられ、そのため不完全なラベルデータから適切な向き推定器を得るために半教師あり学習が有効であると考えられる。半教師あり学習では、ラベル有りデータを用いてラベル無しデータのラベルを得る手法や、ラベル有りデータを用いて得た識別超平面に対してラベル無しデータを用いることでより識別精度が高いと考えられる識別超平面を調整するような手法などがある。

半教師あり学習の学習手法としては、Self-Training[47,48]、Co-Training[42]、Semi-Supervised Support Vector Machine (S3VM) [46]、生成モデルを用いた手法[43,44]やグラフを用いた手法[43,44]などが挙げられる。本提案手法では、半教師あり学習の手法として Self-Training を用いた。Self-Training はラベル有りデータを用いてラベル無しデータのラベルを得る手法であり、単純ながら非常に広範囲の問題に適用可能な手法である。本節では半教師あり学習手法の 1 つである Self-Training について述べる。

4.3.1 Self-Training

Self-Training はラベル有りデータとラベル無しデータの両方を用いて識別器を学習する半教師あり学習手法の 1 つである。Self-Training はラベル有りデータと他の教師あり学習手法を組み合わせるラベル無しデータのラベルを得る手法である。

アルゴリズム 4-1 Self-Training による学習

- 入力としてラベル有りデータの集合 S_L , ラベル無しデータの集合 S_U とする。
- 1: ラベル有りデータ S_L から教師あり学習手法を用いて識別器 f を得る
 - 2: 1 で得た識別器 f を用いてラベル無しデータ $x \in S_U$ のラベルを予測する
 - 3: ラベル無しデータ $x \in S_U$ と 2 で得た予測されたラベル $f(x)$ をラベル有りデータの集合 S_L に追加する
 - 4: 1~3 をラベル無しデータの集合 S_U が空集合になるまで繰り返す
 - 5: 得られたラベル有りデータの集合 S_L と教師あり学習手法を用いて最終的な識別器を学習する
-

Self-Training による学習手順をアルゴリズム 4-1 に示した。アルゴリズム 4-1 から、Self-Training は他の教師あり学習手法を必要とする学習手法であることが確認できる。しかし、Self-Training に用いることができる教師あり学習手法に制限はなく、既存のどのような教師あり学習手法を適用できるため、単純ながらも非常に広範囲な応用が可能である。アルゴリズム 4-1 のステップ 3 において予測されたラベル無しデータのラベルを、ラベル有りデータの集合に加える際の基準はいくつか存在する。それらの基準としては次のようなものが存在する。

1. ラベル無し学習データを全てラベル有りデータに加える
2. ラベル無し学習データのうち、ランダムに複数個をラベル有りデータに加える
3. ラベル無し学習データのうち、最も識別しやすい複数個をラベル有りデータに加える
4. ラベル無し学習データのうち、尤度が一定以上のものを全てラベル有りデータに加える

このうち、1 の基準を用いた場合繰り返しは起こらず、一回学習を行うだけで収束する。しかし、これは最初に与えられたラベル有り学習データによって得られた識別超平面を用いているため、最初に与えられたラベル有り学習データに強く依存した識別超平面が得られてしまう。2 と 3 の基準を用いた場合、ラベル無しデータの数を N_U 、一回にラベル有りデータに追加する学習データの数を A とすると、 N_U/A 回の繰り返しが発生する。これは、ラベル有り学習データにラベル無し学習データを少しずつ追加していき、徐々に識別超平面を調整していくことに対応する。そのため、1 の基準ほど局所解、すなわち最初の

ラベル有り学習データに依存した識別超平面にはならない。しかし、この方法では複数回の学習が必要となり、一回の学習に時間がかかるような教師あり学習手法を用いてしまうと、全ての学習データのラベルを得るのに膨大な時間がかかってしまうといった欠点がある。3と4の基準を用いる場合は、教師あり学習手法によって得られた識別器が、各クラスの尤度を出力できる必要がある。しかし3と4の基準の場合、より確からしい学習データを優先的にラベル有りデータに加えていくため、1や2の基準を用いた時よりも正確で安定した結果が得られる。本提案手法では、8方向の向きクラスに対して尤度を出力する多クラス分類器を用いるため、3の基準を用いる Self-Training を用いた。

4.4 半教師あり学習を用いた向き学習データのアノテーション

前節では一般的な半教師あり学習の手法について述べた。本節では、それら半教師あり学習を本研究に具体的にどのように適用するかについて述べる。

4.4.1 学習データのアノテーションの細分化

4.1 節で述べたように、向き推定手法の問題点の1つとして、正確にアノテーションされた学習データを生成することが挙げられる。特に、向きクラスの境界付近に存在する学習データをいずれの向きクラスとしてアノテーションするべきか、目視によっても難しいという問題がある。

そこで本研究では、そうした向きクラスの境界付近に存在する学習データも正確にアノテーションされた学習データを生成することを目的として、半教師あり学習を用いる。具体的には、目視により簡単にアノテーションが可能であるような画像は人手によりアノテーションし、半教師あり学習におけるラベル有りデータとする。そして、目視により簡単にアノテーションができない、向きクラスの境界付近のデータに関してはそのまま半教師あり学習におけるラベル無しデータとして扱う。このようにすることで、半教師あり学習によってラベル無しデータのラベルを自動的に得ることが可能となる。しかし、向き推定では境界付近に存在する学習データのラベルとしてありうるのは隣接するラベルのいずれかである。そこで、学習データに対して図 23 のように 8 方向より更に 16 方向に細分化したアノテーションを行う。このうち、目視により簡単にアノテーション可能な向きクラスを強ラベル (図 23 (赤))、向きクラスの境界付近に存在するため目視により簡単にアノテーションすることができない向きクラスを弱ラベル (図 23 (緑)) とする。これらを半教師あり学習を用いて学習する際、強ラベルは従来の半教師あり学習におけるラベル有りデータ、弱ラベルは従来の半教師あり学習におけるラベル無しデータとして扱う。従来の半教師あり学習と違う点は、弱ラベルは半教師あり学習によって自動的にアノテシ

ンされるが、アノテーションされるラベルは隣接する強ラベルのどちらかであるという制約がある点である。これにより、半教師あり学習によって学習データが自動的にアノテーションされるが、半教師あり学習によって明らかに異常なラベルがアノテーションされることはない。

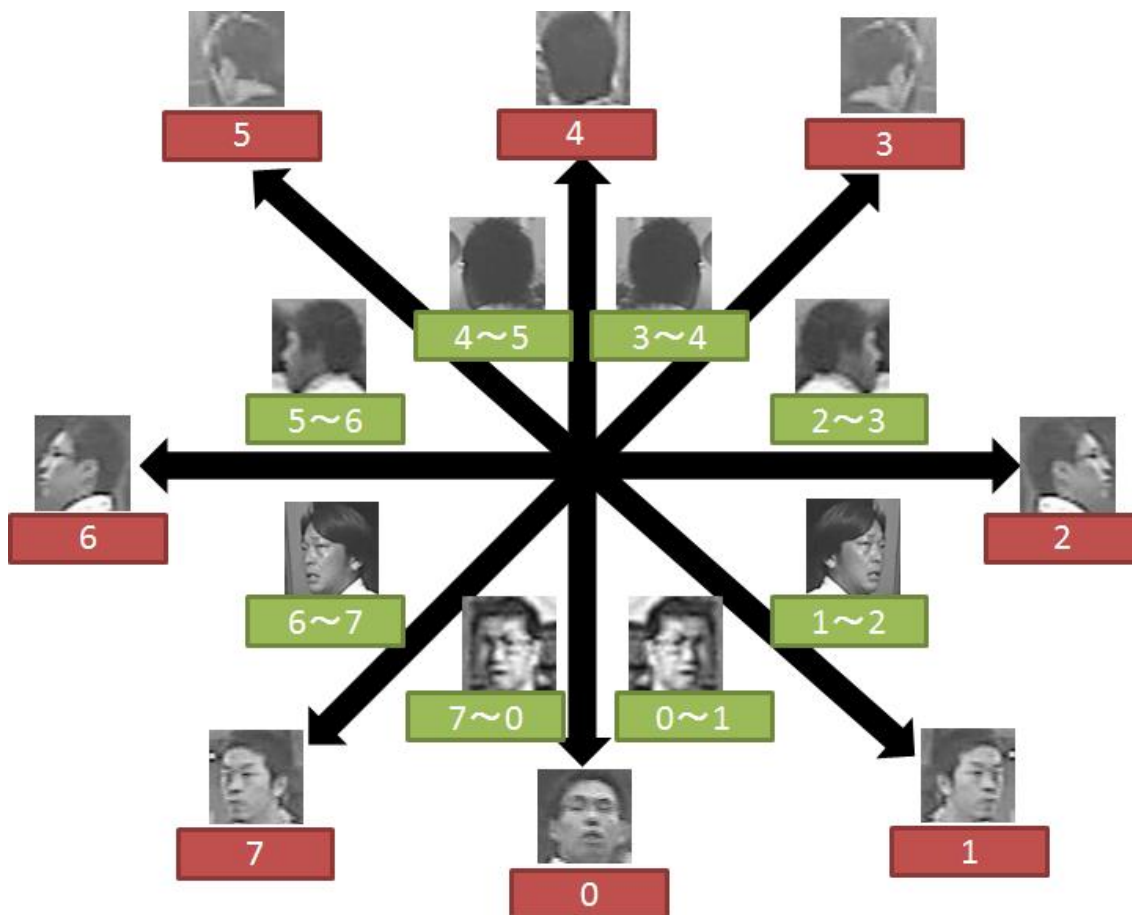


図 23 細分化した学習データの向きクラス：強ラベル（赤）と弱ラベル（緑）

4.4.2 Self-Training による自動アノテーション

本節では、前節までに述べた半教師あり学習手法の1つである Self-Training を、細分化された向きアノテーションを持つ学習データに対して適用した場合の手法について述べる。ただし、用いる教師あり学習手法から得られる識別器は、入力されたデータに対して各クラスの尤度を出力可能な多クラス分類器である。

アルゴリズム 4-2 細分化されたラベルに対する Self-Training

入力として強ラベル付きの学習データ S_S 、弱ラベル付きの学習データ S_W が与えられる。また、一度の操作で強ラベル付きの学習データに追加する弱ラベル付きの学習データの個数は 1 クラスあたり A とする。

- 1 : 強ラベル付きの学習データ S_S と教師あり学習手法を用いて多クラス分類器を得る
 - 2 : 1 で得た識別器を用いて弱ラベル付きの学習データ $x \in S_W$ の各ラベルの尤度を得る
 - 3 : 2 で得た弱ラベル付き学習データとその最も高いクラスの尤度のうち、最も尤度が高い A 個のデータをそのクラスのラベル有りデータの集合 S_S に追加する
 - 4 : 1~3 を弱ラベル付きデータの集合 S_W が空集合になるまで繰り返す
-

アルゴリズム 4-2 は強ラベルと弱ラベルを用いた Self-Training による学習データの自動アノテーションの処理について述べたものである。通常の半教師あり学習の処理であるアルゴリズム 4-1 と異なる点は、本研究では対象としているクラスが 8 つあることから、一回の操作で各クラスへ追加される学習データに制限を設けていることが挙げられる。もし一回の操作で追加される学習データ全体の個数にだけ制限を設けている場合、例えば弱ラベル付き学習データが全てある 1 つのクラスの尤度が最も高かった場合、その操作によってある 1 つのクラスの学習データだけが大きく増加してしまう。この手法を用いると、4.3.1 節で述べたように、特定の局所解に陥ってしまう可能性が高い。各クラス毎に追加されるデータの個数制限を設けた場合、次の学習によって弱ラベルの予測が変化する可能性があるため、より正確な識別超平面が得られると考えられる。

4.5 評価実験

本節では、前節まで述べた半教師あり学習を用いた向き推定手法について、その有効性を確認するためテストデータセットを用いて評価を行う。比較となる対象は、従来の人手によるアノテーションと教師あり学習のみを用いた手法である。本提案手法では強ラベルと弱ラベルの 2 つの異なるラベル付けがあるが、従来の手法ではラベル付けされたデータのみを用いる。そのため、評価実験に用いる学習データ中の弱ラベル付けされたデータは、従来の手法に適用するため更に人手による 8 方向の向きアノテーションがされている。なお、本研究が扱うのは歩行者の頭部向きと身体向きの推定であるため、評価実験では頭部向きと身体向きのいずれに対してもデータセットを用意し、提案手法が有効であるかどうかを評価する。

4.5.1 実験に用いたデータセット

実験に用いたデータセットは当研究室が独自に収集した画像で、全て車載単眼カメラから得られたものである。用いた単眼カメラの情報を表 1 に示す。また、実験映像の撮影時における単眼カメラの設置位置についても図 24 に示す。カメラの設置位置は、様々な歩行者を想定するために、車両の前方方向以外にも車両の側面方向を捉えるような位置にも設置している。

表 1 実験に使用した単眼カメラの詳細

階調	256 段階 (8 ビット) 単階調
解像度	VGA (640 × 480)
フレームレート	10FPS
画角	20° ~100°

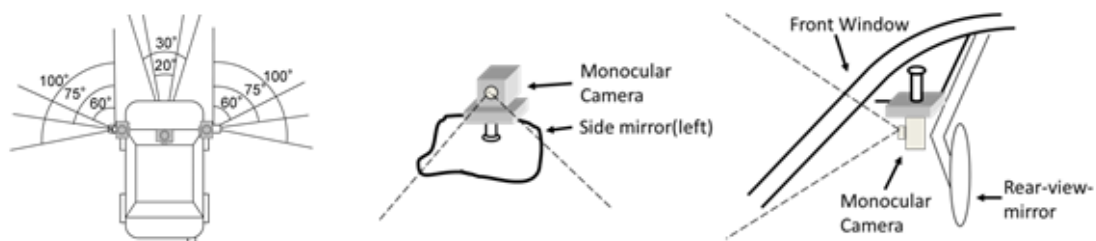


図 24 実験における単眼カメラの設置位置

車載単眼カメラを用いて撮影した映像から、歩行者および歩行者の頭部領域を人手によりアノテーションした後、それらを切り取った画像を学習、テストデータとして用いる。学習データに含まれる画像数を表 2 と表 3 に、テストデータに含まれる画像数を表 4 に示す。

表 2 学習データに含まれる強ラベルデータ数

ラベル	頭部	身体
0	420	202
1	127	233
2	993	877
3	380	281
4	98	396
5	380	281
6	993	877
7	127	233
合計	3,518	3,380

表 3 学習データに含まれる弱ラベルデータ数

ラベル	頭部	身体
0-1	281	458
1-2	440	487
2-3	867	828
3-4	159	364
4-5	159	364
5-6	867	828
6-7	440	487
7-0	281	458
合計	3,494	4,274

表 4 テストデータに含まれるデータ数

ラベル	頭部	身体
0	62	82
1	35	51
2	212	266
3	67	103
4	52	220
5	67	103
6	212	266
7	35	51
合計	742	1,142

学習データに含まれる弱ラベルデータについて、本提案手法では弱ラベルデータはそのまま扱うが、既存の人手によるアノテーションと教師あり学習のみを用いた手法と比較するため、弱ラベルデータについては人手による 8 方向の向きアノテーションもされている。それぞれの学習データセット、テストデータセットに含まれる画像を図 25, 図 26, 図 27 に示す。



図 25 学習データに含まれる強ラベルデータ画像



図 26 学習データに含まれる弱ラベルデータ画像



図 27 テストデータセットに含まれる画像

また，実験に用いる際，全ての頭画像は 18×18 ピクセルに，身体画像は 64×128 ピクセルにリサイズしている．

4.5.2 実験結果

従来手法と提案手法それぞれの向き推定精度を混同行列にして次に示す（図 28，図 29）．ただし，ここでの向き推定精度は，8方向の多クラス分類における分類精度と同じである．また，示した混同行列では，行が正解ラベル，列が多クラス分類器により推定された向きラベルである．

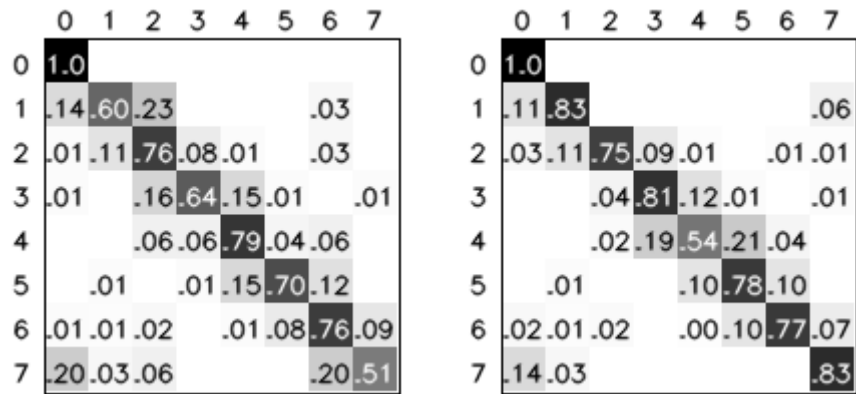


図 28 頭部向き推定における向き推定精度：人手によるアノテーション（左）と半教師あり学習によるアノテーション（右）

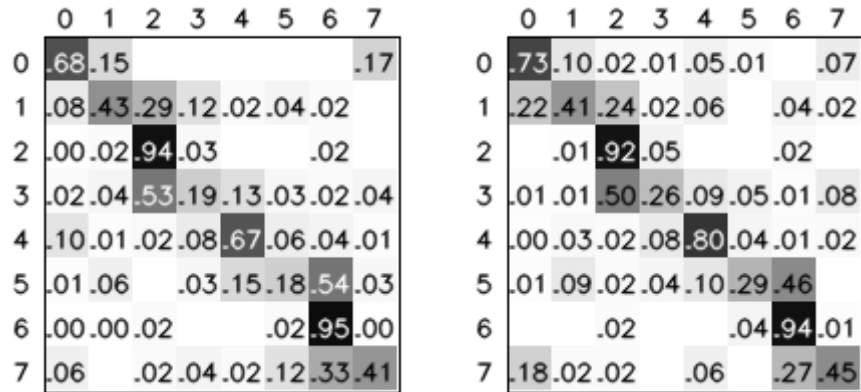


図 29 身体向き推定における向き推定精度：人手によるアノテーション（左）と半教師あり学習によるアノテーション（右）

次に、全体の推定精度を表 5 に示す。表 5 では全体の推定精度を算出する際、向き推定自体の困難さや曖昧さがあるため、推定されたラベルが正解ラベルと同じか、正解ラベルに対して隣接するラベルまでを正しい推定としてカウントした。

表 5 テストデータセットにおける隣接ラベルまでを正解とした時の推定精度

	頭部向き推定精度[%]	身体向き推定精度 [%]
従来手法 (人手によるアノテーションを用いた場合)	94.6	90.6
提案手法 (半教師あり学習を用いた場合)	95.5	92.9

表 5 から，隣接ラベルまでを正解としてカウントした場合，提案手法である半教師あり学習を用いた場合の方が，いずれの場合においてもより高い推定精度を達成していることが確認できる．従来手法に比べ提案手法は，頭部向き推定では 0.9%，身体向き推定においては 2.8%の推定精度の向上が確認された．

第5章 パーティクルフィルタを用いたトラッキング

グ

前章までは、与えられた入力画像から歩行者の頭部向きと身体向きとを推定する手法について述べた。しかし、実際のシステムとして構築する際は、歩行者検出システムから時系列に沿った歩行者画像列が得られる。そこで、本章では歩行者の画像列を入力として、時系列に沿って安定かつ高精度に向き推定を行う手法について述べる。また、頭部・身体向き推定の際、頭部向きと身体向きとの制約を尤度関数に取り込むことで、より高精度に向き推定を行う手法についても本章で述べる。

5.1 歩行者画像列に対する向き推定

本研究が対象とする手法は歩行者画像を入力とした向き推定である。本手法を実際のシステムとして構築する際、最も単純にはまず歩行者検出システムがあり、それらの出力として得られた歩行者画像に対して向き推定を行うことになる。すなわち、時系列に沿った歩行者画像列に対して向き推定手法を用いることになる。しかし、時系列に沿った歩行者の向きは、カメラの撮影間隔より短時間で大きく変化することはないと考えられる。

また、第4章で述べた手法では、歩行者画像列に含まれる各画像に対して毎回頭部位置を探索する必要がある。ここで、頭部領域を表す3つのパラメータ頭部位置 (x, y) と頭部の大きさ s について探索する必要があるため、頭部領域の探索範囲を歩行者の上部領域に絞ったとしても探索する空間が非常に大きいという問題が考えられる。しかし、時系列に沿った歩行者画像列においては、歩行者の頭部位置についても短時間で大きく変化することはないと考えられる。



図 30 実際の車載カメラから得られる歩行者画像列の例 1



図 31 実際の車載カメラから得られる歩行者画像列の例 2

実際の車載カメラから得られた映像中の歩行者を図 30 および図 31 に示した。図 30 に示した歩行者は、頭部及び身体向き、更に頭部位置に関しても大きな変化は見られない。図 31 に示した歩行者は、頭部及び身体向きは時系列に沿って緩やかに変化しているが、隣接フレームにおいては、向きが 90° 以上変化するなどの大きな変化は見られない。

こうした時系列に沿った解析では、前のフレームにおける情報を用いて効率よく推定を行うことが可能である。そのような手法としてはカルマンフィルタやパーティクルフィルタ[54,55]が挙げられる。パーティクルフィルタを用いると、時系列に沿った向きの推定結果の安定化が期待できる。また、頭部位置の探索においても前のフレームの情報を用いることで探索範囲を限定し、効率的に処理を行うことが可能であると考えられる。そこで本研究では、時系列に沿った歩行者画像列に向き推定手法を適用する際、パーティクルフィルタを用いた。

5.2 パーティクルフィルタ

パーティクルフィルタは逐次モンテカルロ法とも呼ばれる、フィルタリングにも用いることが可能な手法の 1 つである。状態の時間的変化をモデル化し更に確率論的に解くことで、対象となる離散的な観測が得られるシステムの内部状態を効率的に推定する手法である。本節では、パーティクルフィルタの処理について述べる。

5.2.1 パーティクルの設計

パーティクルフィルタでは、状態を持つ複数のパーティクルを用いてシステムの内部状態の確率密度分布を近似する。本研究においては、対象とするのは歩行者であり、推定したいパラメータは頭部位置 (x,y) 、頭部の大きさ s 、頭部向き d_h 、身体向き d_b の 5 つのパラメータである (図 32)。

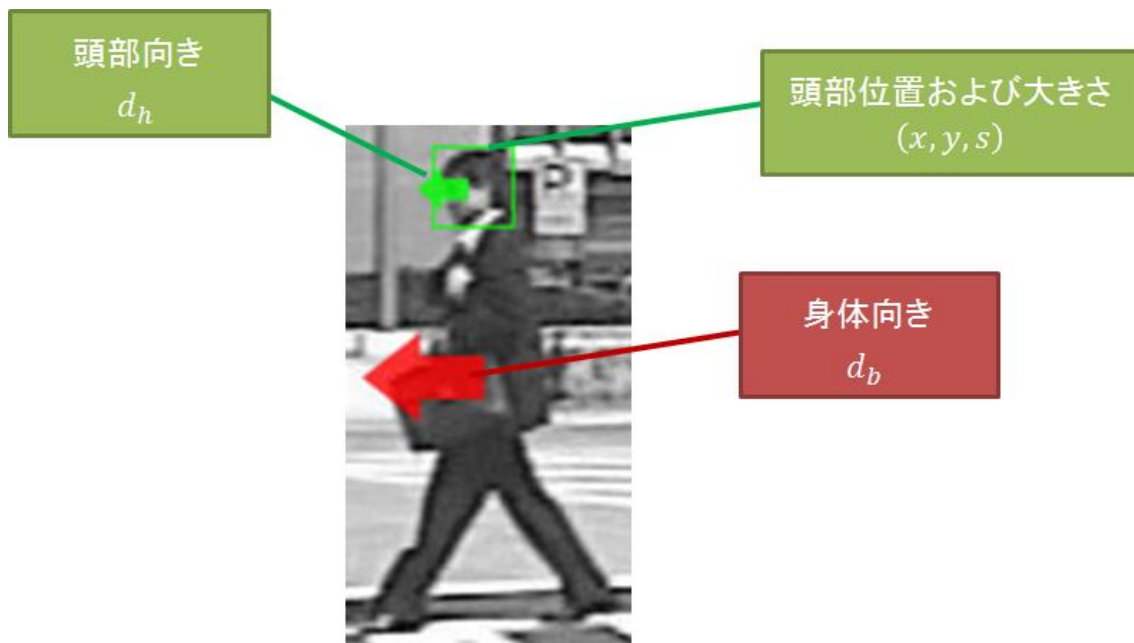


図 32 推定する 5 つの歩行者パラメータ

そのため、1つのパーティクルが持つ状態も先述した 5 つのパラメータであり、1つのパーティクル \mathbf{q} は次のように書き表せる。

$$\mathbf{q} = [x, y, s, d_h, d_b]^T \quad (5-1)$$

このとき、内部状態の確率密度分布を近似するために必要なパーティクルの数は、パーティクルが持つ状態の数に依存する。特に本研究が想定するパーティクルの状態は 5 変数を持ち、そのうち頭部位置 (x, y) および頭部の大きさ s は空間が大きいので、十分な推定精度を保つために必要となるパーティクル数は大きくなる。例えば、頭部位置の探索範囲を 10×10 ピクセルの範囲に、頭部の大きさを ± 2 ピクセルの範囲に限定したとしても、探索する必要がある空間は $10 \times 10 \times 8 = 500$ となる。ここに、頭部向きおよび身体向きの両隣を探索範囲とした場合、探索空間は $500 \times 3 \times 3 = 4500$ となる。本研究では、将来的にリアルタイムなシステムに組み込めることも想定し、2000個のパーティクルを用いている。

5.2.2 尤度関数の設計

前節では 1 つのパーティクルが持つ状態の設計について述べた。パーティクルフィルタにおいては、複数個のパーティクルから最終的な内部状態を推定する (図 33)。本節では、その際に必要となる各パーティクルの重みを計算するための尤度関数について述べる。

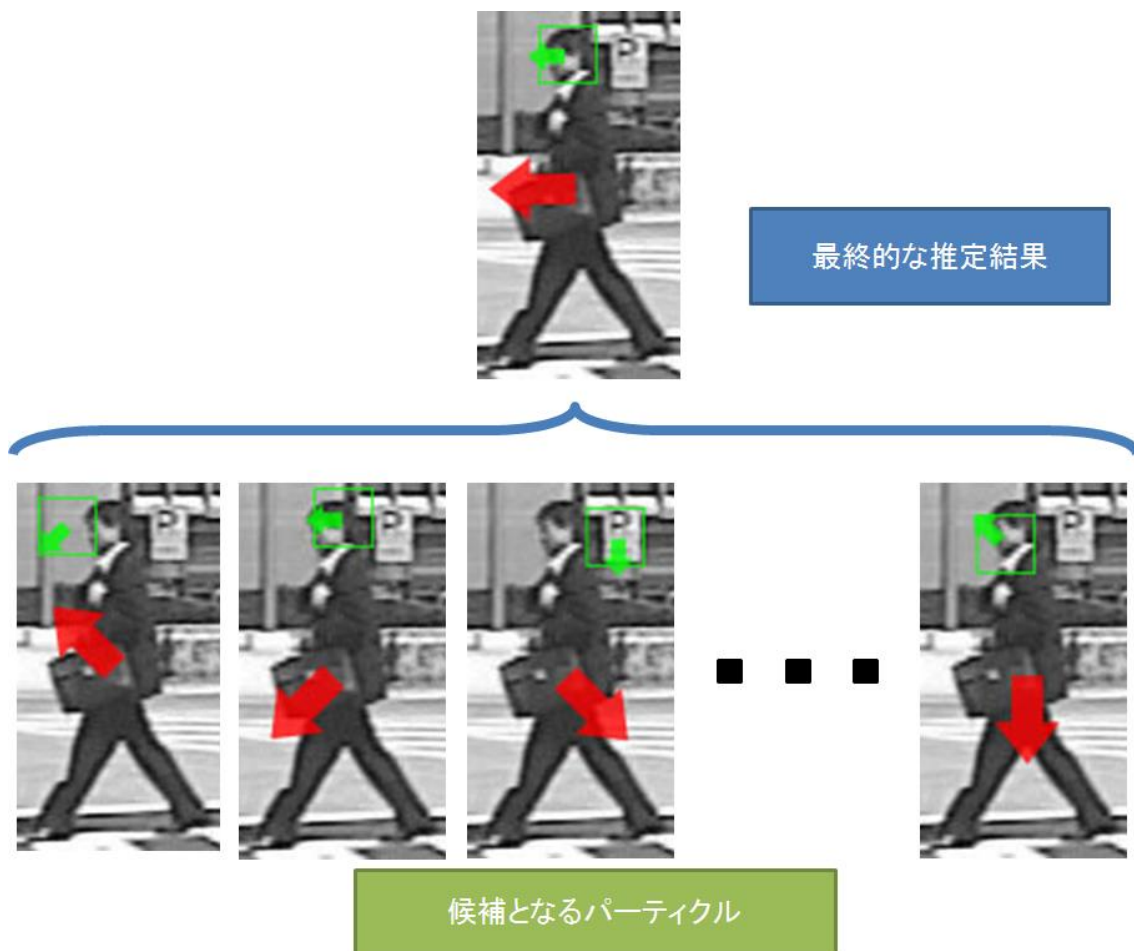


図 33 複数のパーティクルからの歩行者パラメータ推定

パーティクルフィルタにおける尤度関数においては、そのパーティクルの状態が観測にどれほどあてはまっているかを計算する。最も単純には、1つパーティクルは頭部位置・大きさ・向きと身体向きを保持しているため、あるパーティクルの尤度 $L(\mathbf{q})$ は頭部向き推定器と身体向き推定器が出力する尤度を用いて次のように表すことができる。

$$L(\mathbf{q}) = P(d_h|I, x, y, s) \times P(d_b|I) \quad (5-2)$$

このようにすることで、頭部向き推定器と身体向き推定器が出力する尤度が高いほどパーティクルの尤度が高くなる。すなわち、より入力画像にあてはまるようなパーティクルの尤度が高くなることが分かる。

5.2.3 パーティクルフィルタの更新

前節までで、本研究にパーティクルフィルタを適用するために必要な要素について述べた。本節では、それらを用いたパーティクルフィルタの処理について述べる。

アルゴリズム 5-1 パーティクルフィルタ

入力として T 枚の歩行者の画像列 $I_1, I_2, I_3, \dots, I_T$ と初期推定値 \hat{Q}_0 をとる。時刻 t における推定された歩行者のパラメータを \hat{Q}_t 、用いるパーティクルの数を M 、 i 番目のパーティクルの重みを $w_i, i = 0, 1, 2, \dots, M-1$ 、パーティクル \mathbf{q} の尤度を $L(\mathbf{q})$ とする。

1 : M 個のパーティクルを初期推定 \hat{Q}_0 の周囲に生成する。

2 : パーティクルの重みを初期化する, $w_i \leftarrow \frac{1}{M}, i = 0, 1, 2, \dots, M-1$

3 : 時刻 $t = 1$ から $t = T$ まで, ステップ 4 からステップ 7 までを順番に繰り返す。

4 : 各パーティクルの重みを更新する, $w_i \leftarrow L(\mathbf{q}_i), i = 0, 1, 2, \dots, M-1$

5 : パーティクルの重みを正規化する, $w_i \leftarrow \frac{w_i}{\sum_j^M w_j}$

6 : 時刻 t における推定結果を計算する, $\hat{Q}_t = \sum_i^M w_i \mathbf{q}_i$

7 : パーティクルの重みに従いパーティクルを再生成する, i 番目のパーティクル \mathbf{q}_i の周囲に Mw_i 個のパーティクルを生成する, ただし, 再生成されたパーティクルの総数は M 個となるように調整する

8 : 7 で生成されたパーティクルを次の繰り返して用いるパーティクルとする

アルゴリズム 5-1 にパーティクルフィルタを用いた入力画像列の処理を示した。逐次的に入力画像を処理することにより、各時刻における歩行者パラメータが推定される。

5.2.4 向き推定器の尤度のみを用いた場合の問題点

パーティクルの尤度を計算する際、観測として得られる歩行者画像列とパーティクルの状態から、頭部向き推定器と身体向き推定器が出力する尤度をもとにした。しかし、この尤度関数では頭部向きと身体向きとの間に制約が全く存在しないため、図 34 に示すような推定結果が得られる可能性もある。



図 34 頭部向きと身体向きの制約がない場合の推定

図 34 では、対象歩行者の頭部領域と頭部向きは正しく推定されているが、身体向きが真反対の方向だと推定されている例である。本研究が用いた HOG 特徴量では、おおまかな歩行者の輪郭情報を用いているため、向きによっては判別が難しい。しかし、人体の構造からこのような頭部向きと身体向きが正反対になるようなことは非常に困難であると考えられる。

5.3 頭部向きと身体向きを考慮した尤度

前節では、歩行者画像列に対して向き推定手法を用いる際にパーティクルフィルタを適用した手法について述べた。しかし、パーティクルの尤度計算の際に向き推定器のみを用いると、パーティクルの状態によっては頭部向きと身体向きが正反対の方向として推定されてしまうという問題が考えられる。しかし、現実の歩行者ではこのような姿勢は人体の

構造から非常に困難であり、推定結果として頭部向きか身体向きのいずれかもしくは両方が間違っていると考えられる。

また、頭部向き推定器と身体向き推定器とでは、判別が行いやすい向きクラスが異なる。例えば、頭部向き推定において、対象歩行者の頭部がカメラに向かっている場合、鼻や目といった顔のテクスチャが現れるため向き推定は比較的容易であると考えられる。しかし、カメラとは正反対の方向に頭部を向けている場合、頭部の殆どの領域が頭髪であったり帽子であったりするため、ノイズが多く得られた画像のみで判別することが難しい。身体向きにおいては、特徴量として HOG 特徴量を用いているため、大まかな歩行者の輪郭情報が使われる。そのため、カメラに向かっている歩行者とカメラと正反対に向かっている歩行者は輪郭情報が似ているため、判別することが難しい。

更に、一般に歩行者においては頭部向きと身体向きは一致していることが多いと考えられる。以上のような検討から、本提案手法では頭部向きと身体向きとの間に存在する制約条件を明示的にパーティクルの尤度計算に取り入れた。具体的には、頭部向きと身体向きとの相対角度に von Mises 分布を用いる。von Mises 分布は角度における正規分布と非常によく似た性質を持つ。von Mises 分布では相対角度が小さいほど確率が高く、相対角度が大きいほど確率が低いという性質を持つ。von Mises 分布を用いた時、パーティクルの尤度を計算する尤度関数 $L(\mathbf{q})$ は次のように表される。

$$L(\mathbf{q}) = P(d_h|I, x, y, s) \times P(d_b|I) \times p_M(\theta_{hb}, 0, \beta_{hb}) \quad (5-3)$$

ただし、 p_M は von Mises 分布、 θ_{hb} は頭部向きと身体向きの相対角度、 β_{hb} は von Mises 分布のパラメータを表す。

頭部向きと身体向きとの間の制約条件を考慮することで、頭部向きと身体向きが正反対であるような間違った推定を削減でき、更に頭部向きと身体向きは一致しやすいという特徴を取り入れることが可能であると考えられる。

5.4 評価実験

提案手法として歩行者画像列に対して頭部向きと身体向きの拘束を尤度関数に取り入れたパーティクルフィルタによる向き推定手法の有効性を確認するため、実際の車載カメラを用いて撮影した映像を使って評価実験を行った。特に頭部向きと身体向きの拘束を評価するため、尤度関数に頭部向きと身体向きの拘束を取り入れたパーティクルフィルタと、取り入れていないパーティクルフィルタとで比較を行う。本節では、評価実験に用いたデータセットおよび実験結果について述べる。

5.4.1 実験に用いたデータセット

実験に用いるデータセットには実際の車載単眼カメラから得られた映像を用いる。車載単眼カメラの詳細や設置位置などについては4.5.1節で述べた内容と同じである。ただし、頭部学習データセットには歩行者画像から頭部を探索するために必要となる非頭部画像として16,533枚が追加されている。



図 35 学習データに追加した非頭部画像

学習データセットに追加した非頭部画像を図35に示した。図35上段では、車載単眼カメラから撮影した映像のうち、車両やビルなどの建物、道路上の標示など、歩行者が全く含まれないような画像を示している。図35下段では、歩行者の頭部周辺の画像を示している。これは、歩行者の頭部周辺の画像も非頭部画像として学習データに加えることで、歩行者の頭部領域を推定する際、より正確に歩行者の頭部領域を推定することができると考えられるからである。

テストデータとして用いた歩行者のシーケンス数は92で、それらのシーケンスに含まれる総フレーム数は2,129である。

5.4.2 実験結果

歩行者画像列に対する向き推定の結果として、頭部領域の重なり率と頭部・身体向き推定精度によって評価する。頭部領域の重なり率は、歩行者の頭部領域の探索がどの程度正解領域に近いかを測るための指標となる。頭部・身体向き推定精度は8方向の向きクラスのカテゴリ精度によって評価する。

頭部の重なり率 R は、頭部の正解領域 $r_{GroundTruth}$ と推定された頭部領域 $r_{Predicted}$ から、次のように計算される。

$$R = \frac{r_{GroundTruth} \cap r_{Predicted}}{r_{GroundTruth} \cup r_{Predicted}} \quad (5-4)$$

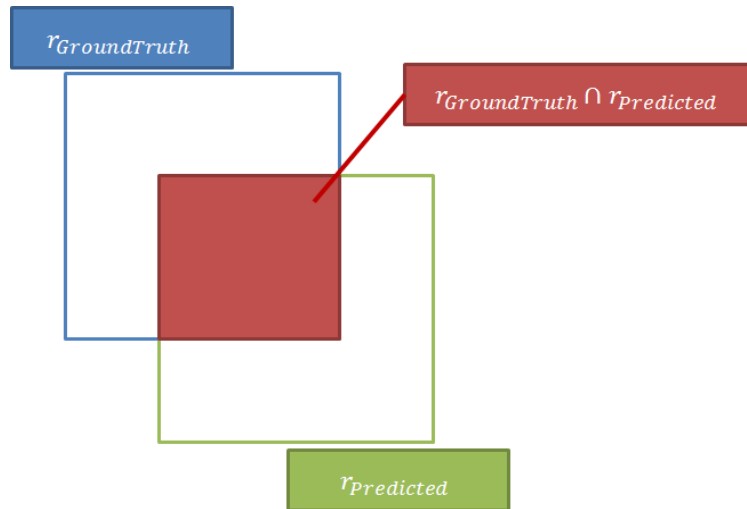


図 36 正解領域と推定領域の重なり

図 36 は正解領域と推定領域の重なりを図示したものである。正解領域と推定領域のいずれにも重なっている領域 $r_{GroundTruth} \cap r_{Predicted}$ は、正解領域と推定領域のいずれかに重なっている領域 $r_{GroundTruth} \cup r_{Predicted}$ より面積が等しいか小さいため、重なり率は 1 より大きくなる。また、 $r_{GroundTruth} \cap r_{Predicted}$ の面積は 0 より小さくならないため、重なり率も 0 より小さくならない。重なり率が大きければ大きいほど、正解領域と推定領域が重なっていると考えられる。

表 6 テストデータにおける頭部重なり率

	重なり率の平均
頭部と身体向きの制約を用いない場合	0.634
頭部と身体向きの制約を用いた場合	0.641

表 6 に実際のテストデータに対して推定を行った際の頭部重なり率を示した。頭部と身体向きの制約を用いていない従来手法に比べ、本提案手法の方が僅かながら重なり率が向上していることが確認できる。

次に、頭部向きと身体向きの推定精度について評価する。テストデータにおける頭部と身体向き推定精度を混同行列として図 37 と図 38 に示す。

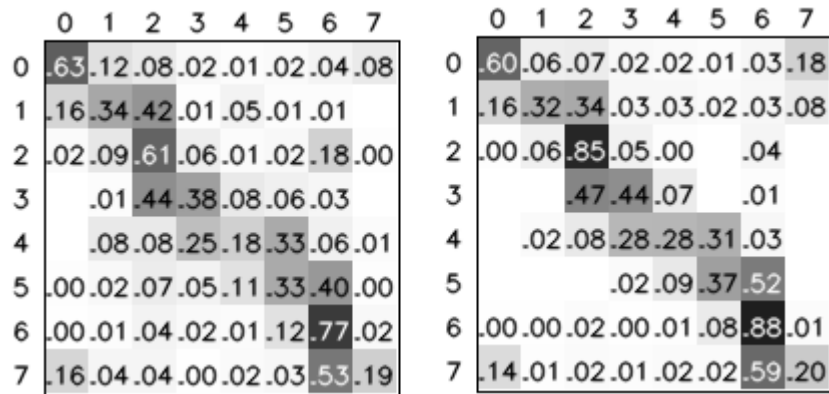


図 37 テストデータにおける頭部向き推定精度：頭部と身体向きの制約なし（左）と制約あり（右）

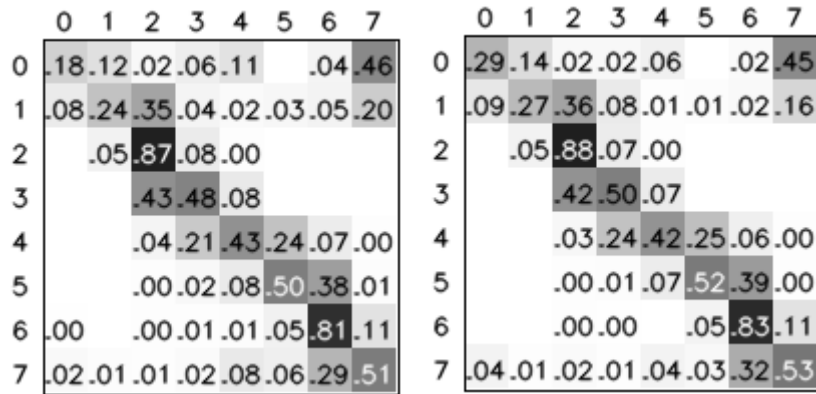


図 38 テストデータにおける身体向き推定精度：頭部と身体向きの制約なし（左）と制約あり（右）

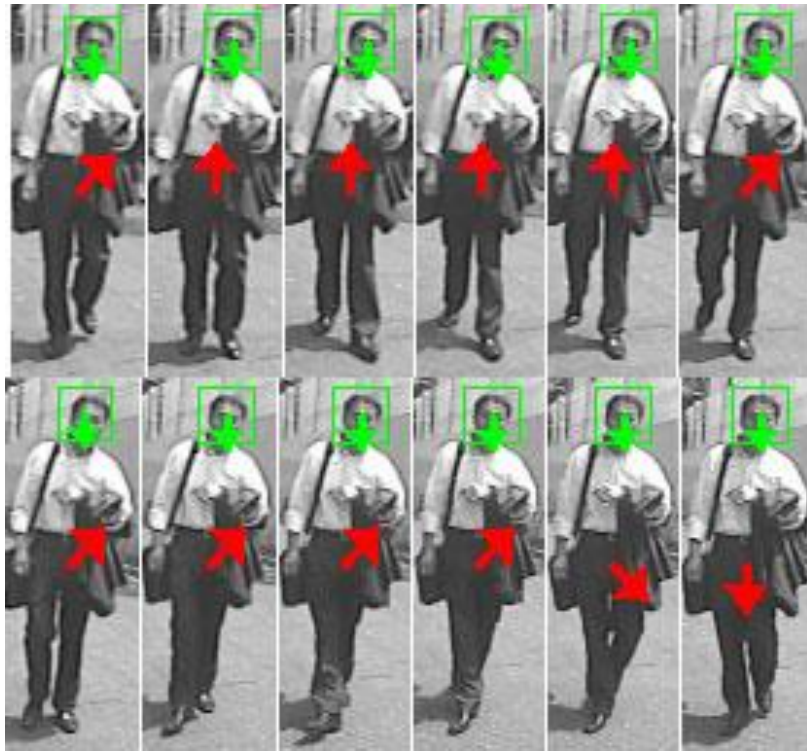
また、全フレームにおける向きの推定精度として、隣接する向きクラスも正解としてカウントした場合の推定精度を表 7 に示す。

表 7 テストデータにおける隣接ラベルまでを正解とした時の向き推定精度

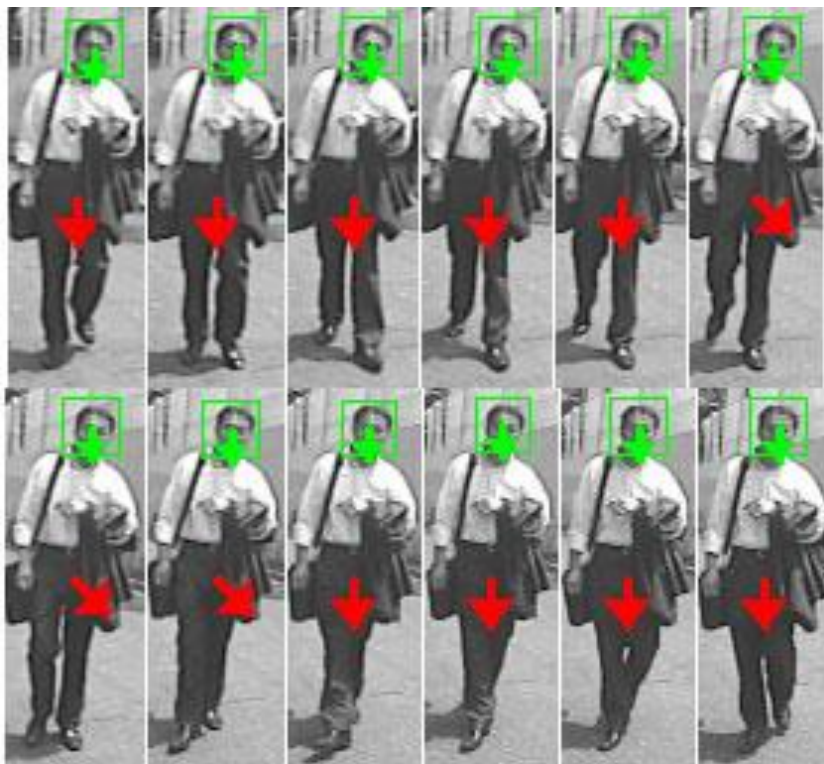
	頭部向き推定精度[%]	身体向き推定精度[%]
頭部と身体向きの制約を用いない場合	86.1	90.7
頭部と身体向きの制約を用いた場合	93.5	94.0

表 7 から、本提案手法が扱う頭部と身体向きの間での制約を考慮した尤度関数が向き推定精度の高精度化に有効であることが確認できる。頭部向き推定においては 7.4%，身体向き

推定においては3.3%の向上が見られる。テストデータに含まれる歩行者画像列に対して、頭部と身体向きを考慮しない推定手法を用いた場合の推定結果と、制約を考慮した場合の推定結果を可視化したものをいくつか次に示す。



(a) テストデータにおける向き推定結果 1 : 頭部と身体向きにの制約なし

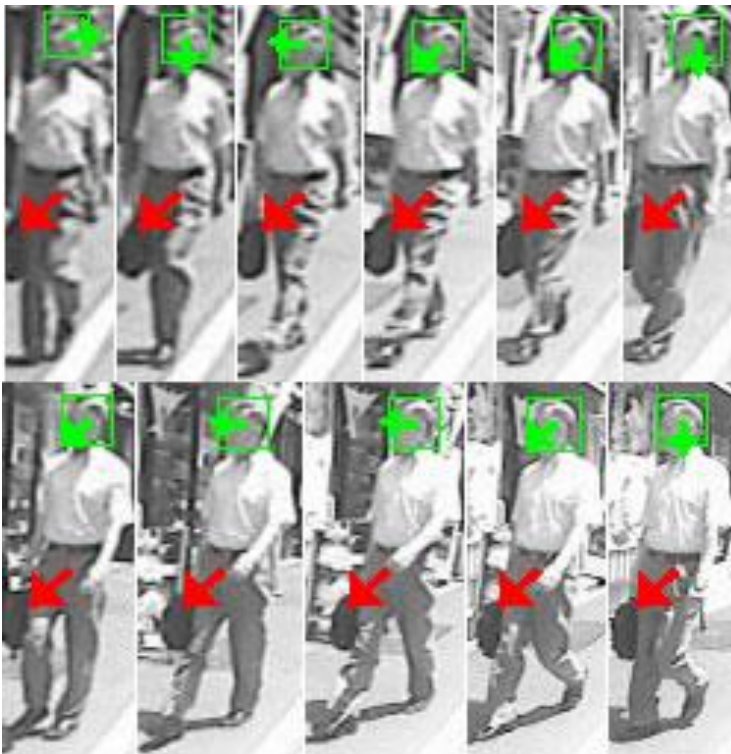


(b) テストデータにおける向き推定結果 1 : 頭部と身体向きにの制約あり

図 39 テストデータにおける歩行者の向き推定結果 1

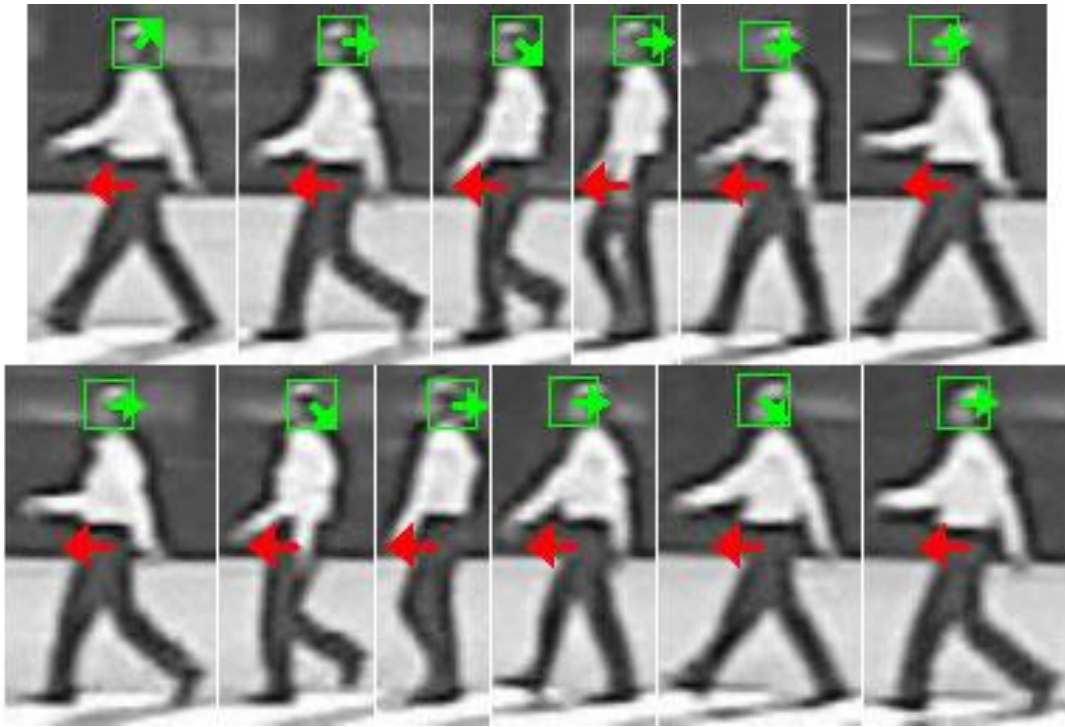


(a) テストデータにおける向き推定結果 2 : 頭部と身体向き of 制約なし

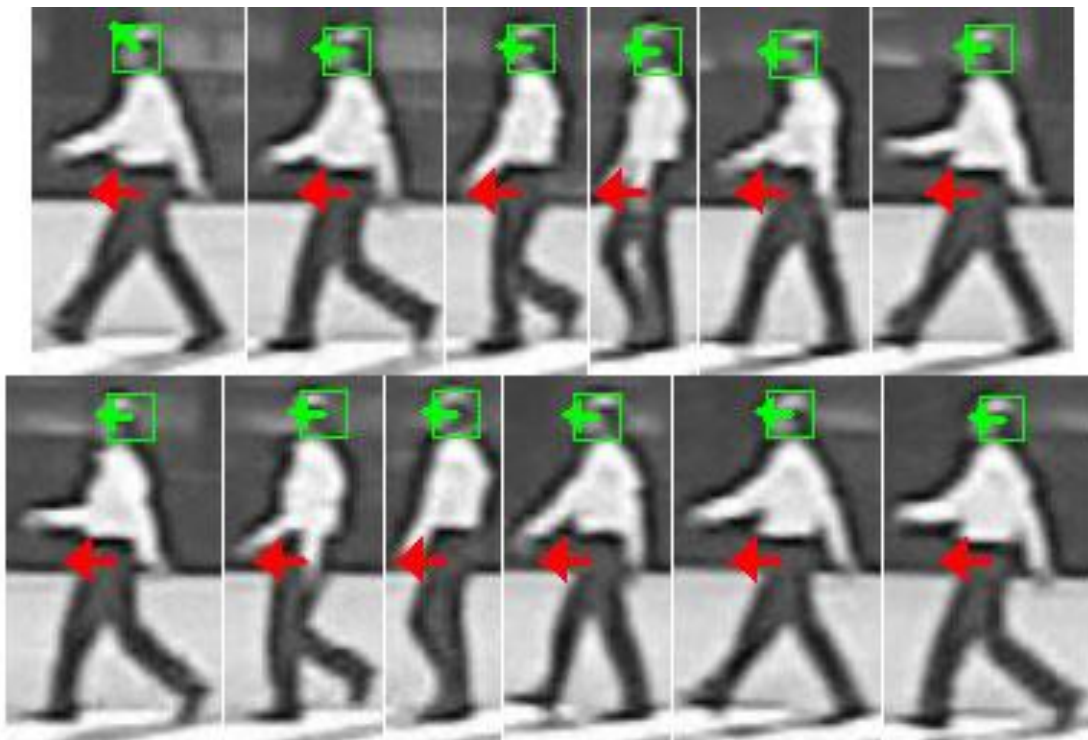


(b) テストデータにおける向き推定結果 2 : 頭部と身体向き of 制約あり

図 40 テストデータにおける歩行者の向き推定結果 2

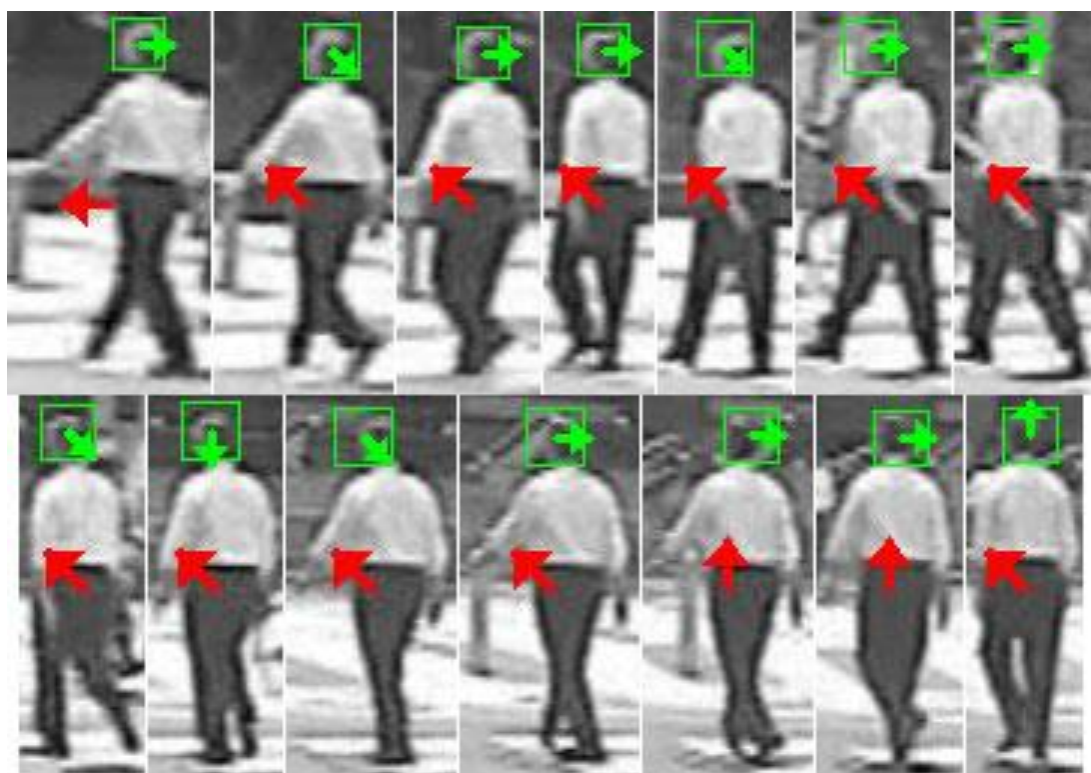


(a) テストデータにおける向き推定結果 3 : 頭部と身体向き of 制約なし

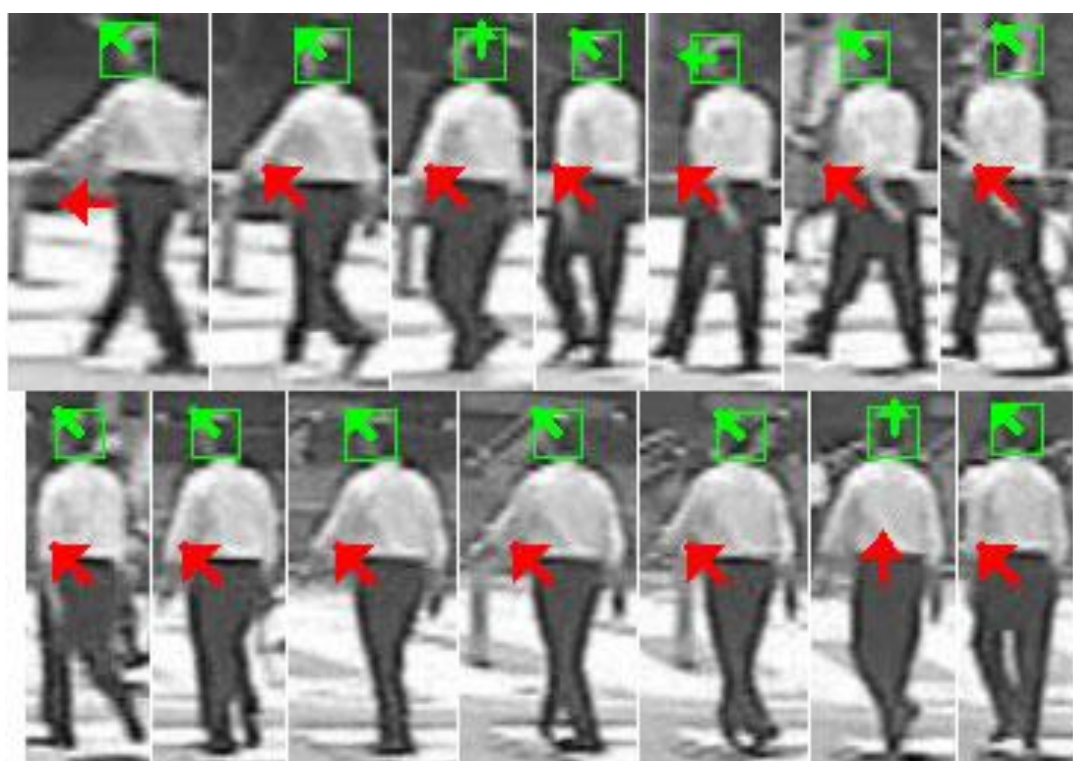


(b) テストデータにおける向き推定結果 3 : 頭部と身体向き of 制約あり

図 41 テストデータにおける歩行者の向き推定結果 3

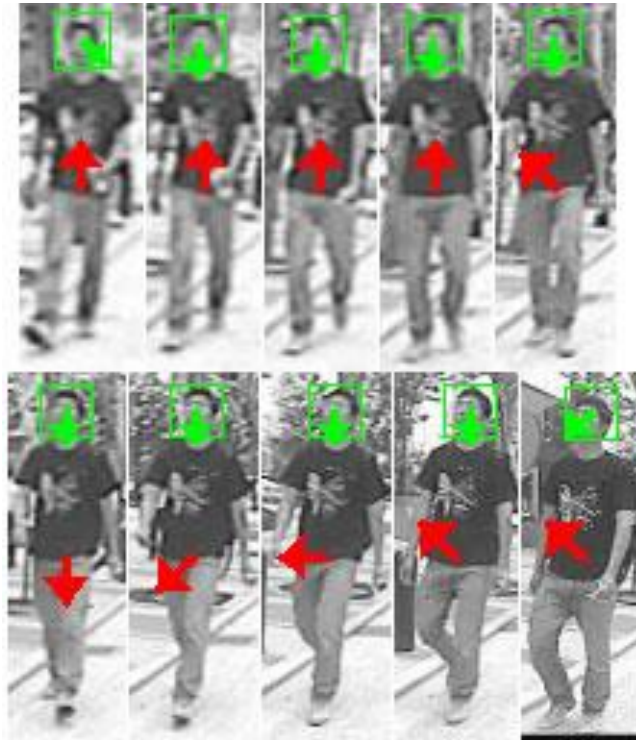


(a) テストデータにおける向き推定結果 4 : 頭部と身体向きへの制約なし



(b) テストデータにおける向き推定結果 4 : 頭部と身体向きへの制約あり

図 42 テストデータにおける歩行者の向き推定結果 4



(a) テストデータにおける向き推定結果 5 : 頭部と身体向きにの制約なし



(b) テストデータにおける向き推定結果 5 : 頭部と身体向きにの制約あり

図 43 テストデータにおける歩行者の向き推定結果 5

第6章 結論

6.1 本研究のまとめ

本研究では、交通シーンにおける先進的なシステムの実現への寄与を目的とし、歩行者の頭部位置・向き推定及び身体向き推定手法について述べた。歩行者の向き推定手法を高精度化するにあたって、次の2つの手法を提案した。

6.1.1 半教師あり学習を用いた向き推定の高精度化

向き推定手法自体には、解像度の問題や学習データのアノテーションに関する問題があることを述べた。このうち、学習データのアノテーションが曖昧である問題に注目し、半教師あり学習を用いることで自動的に学習データのアノテーションを生成する手法を提案した。半教師あり学習を用いた自動的な学習データのアノテーション生成手法では、学習データの向きクラスを従来の8方向から16方向（強ラベル8方向+弱ラベル8方向）へ細分化することで、半教師あり学習に適用可能な形とした。評価実験により、半教師あり学習を用いて生成したアノテーションを用いた学習データによる向き推定手法の方が高い向き推定精度を達成することを確認した。

6.1.2 頭部向きと身体向きを考慮した尤度関数の設計による向き推定の高精度化

本研究が対象とする歩行者向き推定手法は、歩行者検出システムの後に適用される処理である。歩行者検出システムからは一般的に歩行者を時系列に沿ってトラッキングする処理も含まれているため、歩行者向き推定手法には歩行者画像列が入力される。ここで、カメラの撮影間隔より短い時間に、歩行者の向きが大きく変化することや、歩行者の頭部位置が大きく異なることは考えにくい。そこで、本研究では時系列に沿った解析に有効な手法の1つであるパーティクルフィルタを用いて、歩行者画像列に対して安定かつ高精度に向き推定を行う手法について提案した。特に、パーティクルフィルタの尤度関数として、頭部向きと身体向きとの拘束を考慮した尤度関数を適用することで、高精度な向き推定手法についても提案した。提案手法に対して、実際の車載カメラから得られた映像を用いた評価実験を行い、特に頭部向きと身体向きとの拘束を入れた尤度関数が向き推定の高精度化に寄与していることを確認した。

6.2 今後の課題

前節までは本研究で得られた成果について述べた。しかし、本提案手法を実際のシステムとして用いるまで、まだ解決すべき問題点は残っている。本節では、それら解決すべき問題点として今後取り組むべき課題について述べる。

6.2.1 歩行者の向き推定精度

本研究では歩行者画像列に対する向き推定精度として、頭部及び身体向きの制約を考慮することで、頭部向きではおよそ 93%、身体向きではおよそ 94%という結果を得た。しかし、交通安全の観点から見ると、これらの推定精度は決して高いものではない。人命がかかったシステムの開発においては、用いられる手法に要求される信頼性は非常に高い。そのような要求を達成するためにも、今回の研究より更に向き推定精度を向上させるような手法について今後取り組む必要があるだろう。

歩行者の向き推定そのものでは、画像から抽出する画像特徴量の吟味や、より向き推定に特化した学習・識別手法の発展が期待される。特に頭部向き推定は、身体向きに比べ非常に解像度が低いこと、更に頭部領域の探索という処理が入るため、非常に推定精度を高めることが困難である。本提案手法では低解像度においても高い記述力を持つ CoHOG 特徴量を用いたが、更に向き推定に特化した画像特徴量についても今後検討していく必要があるだろう。

6.2.2 歩行者の状態推定

本提案手法では歩行者の頭部向きと身体向きの推定手法について述べた。しかし、実際の交通シーンにおいては車両の周囲の歩行者が立ち止まっているか、歩いているか、あるいは走っているかなど、対象歩行者の状態は様々である。こうした対象歩行者の状態は、対象歩行者の移動速度に関連する情報であり、車両前方への飛び出しなどを考慮した場合は推定すべき情報であると考えられる。より実践的なシステムの開発においては、本研究が対象としている歩行者向きと、更に歩行者の状態を考慮に入れた歩行者の危険度推定などが必要となってくるであろう。

6.2.3 子供の検出、姿勢推定

実際の交通シーンにおいて、特に子供は保護すべき対象として重要である。本提案手法における評価実験では子供のデータが含まれていないため、子供に対して本提案手法がど

の程度有効であるのか、今後確認していく必要がある。また、子供は歩行者検出システムにおいても未だに残されている課題の 1 つである。歩行者検出システムが子供を検出した場合は、その情報を用いた向き推定を行うなど、より子供に特化した向き推定手法も考えることで、より安全な交通の実現を目指していくことが期待される。

6.2.4 歩行者検出手法との統合

本研究では歩行者画像からの向き推定手法のみを取り上げ、歩行者検出システムとの統合的なシステムについては評価していない。本研究における評価実験では、歩行者の向き推定手法の精度のみに焦点を絞るため、人手により切り取った歩行者画像列をテストデータとして用いた。しかし、実際の歩行者検出システムにおいては、必ずしも正確な歩行者領域が抽出されるわけではないことや、非歩行者領域が歩行者として抽出される可能性もあることが考えられる。今後これらの問題点への対応手法について検討する必要があるだろう。

また、歩行者検出システムにおける問題点への対処の他、歩行者検出システムが取得可能であるような情報との積極的な応用についても考えるべきであろう。例えば、歩行者検出システムは検出された各歩行者のトラッキングも行っている。トラッキングしている歩行者の画像内での位置と、更に自車両の車速や舵角情報を元にすれば、車両からの歩行者の相対位置や歩行者の絶対速度を計算することが可能である。特に歩行者の絶対速度は歩行者の身体向きと大きく関連があると考えられるため、そうした歩行者検出システムからの情報を用いた歩行者向き推定の高精度化も可能である。

6.2.5 歩行者のオクルージョン

対象となる歩行者が他の歩行者や車両により部分的あるいは全て遮蔽されているような状況についても今後考慮する必要がある。本提案手法では歩行者の向き推定に主に画像の勾配情報を用いている。そのため、歩行者の遮蔽された領域のテクスチャ情報がノイズとなり、向き推定に悪影響を及ぼすことが考えられる。本提案手法では対象歩行者のオクルージョンに関しては考慮しておらず、身体向きの推定では常に全身が映っていることが前提となっている。しかし、歩行者が多数存在するような環境においては歩行者同士のオクルージョンが多発すること、雨天時などは傘をさしているため身体の一部が遮蔽されるなど、日常的な状況であってもオクルージョンは発生する。

歩行者のオクルージョンにロバストな向き推定を行うためには、歩行者のより詳細な姿勢情報を用いる、歩行者の絶対速度情報のみを用いるなどの手法をとっていく必要があるだろう。歩行者の全身がオクルージョンされている場合は、カメラによる姿勢推定などは不可能であるが、無線センサなど他のセンサを用いた歩行者検出技術を用いて歩行者を検出した後、歩行者の軌跡を元に姿勢を推定するなど、センサフュージョンによる解決手法の発

展が期待される。

謝辞

本研究を進めるにあたって、上條俊介准教授を始め非常に多くの方にお世話になりました。特に上條俊介准教授には、基本的な研究の仕方から始まりプレゼンテーションの行い方、また研究者としてどうあるべきかについても丁寧に指導して頂き、非常に感謝しております。

秘書の三輪裕子さんには、外部での発表や研究室として必要な物資を購入する際など、研究生を送るにあたって必要となってくる事務作業において非常に助けられました。しかしそんな事務作業を頼む時でも、いつも親切に対応してくださいました。この場を借りて感謝致します。

研究室に入った後、先輩として親身にアドバイスをしてくださった日坂翔馬さん、新見洋正さん、そして同じ画像系の研究分野であったため時に研究上重要なアドバイスをしてくださった高橋俊充さんの先輩方にはとてもお世話になりました。また、研究において悩んでいる時もそれを前進させるアドバイスを多く下さった研究員の古艶磊さんにはとても感謝しています。そして、同期として一生懸命研究に集中する姿勢を見せてくれた青木尚久君、外部発表などの際は非常に丁寧に準備を進め慎重かつ丁寧に研究を進める姿勢で刺激を与えてくれた垣内紀明君、研究室の些細なことから分野の違う私の研究など様々なところで相談にのってくれた三浦俊祐君の三人にもとても感謝しています。二年間という長いようで短い研究生生活の間でしたが、本当にお世話になりました。また、後輩の砂川健一君と陳飛宇には、あまり先輩としてやるべき事をやってあげられなかったような気がしています、申し訳ありません。

最後に今一度、研究生生活においてお世話になった全ての方々に感謝致します。

平成 26 年 2 月 6 日

発表文献

国際論文

[1] S.Yano and S.Kamijo: “Estimation of Pedestrian Pose and Orientation Using On-Board Camera With Histograms of Oriented Gradients Features”, International Journal of Intelligent Transportation Systems Research (条件付き採録)

国内発表

[2] S.Yano and S.Kamijo: “HOG 特徴量とパーティクルフィルタを用いた車載カメラからの歩行者向き推定”, 自動車技術会 2013 年秋季大会, Oct. 2013.

参考文献

- [1] “交通事故発生状況（警察庁）”, URL: <http://www.npa.go.jp/toukei/koutuu48/home.htm> accessed on 2014/01/15
- [2] M.Andriluka, S.Roth and B.Schiele: “People-tracking-by-detection and people-detection-by-tracking”, IEEE Conference on Computer Vision and Pattern Recognition, pp.1-8, June. 2008.
- [3] M.Andriluka, S.Roth and B.Schiele: “Pictorial structures revisited: People detection and articulated pose estimation”, IEEE Conference on Computer Vision and Pattern Recognition, pp.1014-1021, June. 2009.
- [4] M.Andriluka, S.Roth and B.Schiele: “Monocular 3D pose estimation and tracking by detection”, IEEE Conference on Computer Vision and Pattern Recognition, pp.623-630, June. 2010.
- [5] P. Felzenszwalb, D. McAllester, D. Ramanan: “A Discriminatively Trained, Multiscale, Deformable Part Model”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1-8, June. 2008.
- [6] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan: “Object Detection with Discriminatively Trained Part Based Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, Issue. 9, pp.1627-1645, Sep. 2010.
- [7] M.A.Fischler and R.A.Elschlager: “The representation and matching of pictorial structures”, IEEE Transactions on Computer, Vol.22, Issue.1, pp.67-92, Jan. 1973.
- [8] P.Felzenszwalb and P.Huttenlocher: “Pictorial Structures for Object Recognition”, International Journal of Computer Vision, Vol.61, Issue.1, pp.55-79, Jan. 2005.
- [9] N.Dalal and B.Triggs: “Histograms of oriented gradients for human detection”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.1, pp.886-893, June. 2005.
- [10] R.Ronford, C.Schmid and B.Triggs: “Learning to parse pictures of people”, European Conference on Computer Vision, 2002.
- [11] S.Johnson and M.Everingham: “Combining discriminative appearance and segmentation cues for articulated human pose estimation”, IEEE International Conference on Computer Vision Workshops, pp405-412, Sept-Oct. 2009.
- [12] Y.Yang and D.Ramanan: “Articulated pose estimation with flexible mixtures-of-parts”, IEEE Conference on Computer Vision and Pattern Recognition, pp.1385-1392, June. 2011.
- [13] Y.Yang and D.Ramanan: “Articulated Human Detection with Flexible Mixtures of Parts”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.35, Issue.12, pp.2878-2890, Dec. 2013.
- [14] H.Jiang: “Human Pose Estimation Using Consistent Max Covering”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.33, Issue.9, pp.1911-1918, Sept. 2011.

- [15] B.Sapp, C.Jordan and B.Taskar: "Adaptive pose priors for pictorial structures", IEEE Conference on Computer Vision and Pattern Recognition, pp.422-429, June. 2010.
- [16] A.O.Balan, L.Sigal, M.J.Black and J.E.Davis: "Detailed Human Shape and Pose from Images", IEEE Conference on Computer Vision and Pattern Recognition, pp.1-8, June. 2007.
- [17] J.Gall, B.Rosenhahn, T.Brox and H.P.Seidel: "Optimization and Filtering for Human Motion Capture", International Journal of Computer Vision, Vol.87, Issue.1-2, pp.75-92, March. 2010.
- [18] M.Vondrak, L.Sigal and O.C.Jenkins: "Physical Simulation for Probabilistic Motion Tracking", IEEE Conference on Computer Vision and Pattern Recognition, pp.1-8, June. 2008.
- [19] P.Viola and M.Jones: "Rapid object detection using a boosted cascade of simple features", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.1, pp.511-518, 2001
- [20] P.Viola and M.Jones: "Robust real-time face detection", IEEE International Conference on Computer Vision, pp.747, 2001
- [21] J.Qiang, Z.Zhiwei and P.Lan: "Real-time nonintrusive monitoring and prediction of driver fatigue", IEEE Transactions on Vehicular Technology, Vol.53, Issue.4, pp.1052-1068, July. 2004.
- [22] P.Smith, M.Shah and N.da Vitoria Lobo: "Determining driver visual attention with one camera", IEEE Transactions on Intelligent Transportation Systems, Vol.4, Issue.4, pp.205-218, Dec. 2003.
- [23] E.Murphy-Chutorian, A.Doshi and M.M.Trivedi: "Head Pose Estimation for Driver Assistance Systems: A Robust Algorithm and Experimental Evaluation", IEEE Intelligent Transportation Systems Conference, pp.709-714, Sept-Oct. 2007.
- [24] A.Schulz and R.Stiefelhagen: "Video-based Pedestrian Head Pose Estimation for Risk Assessment", IEEE Conference on Intelligent Transportation Systems, pp.1771-1776, Sept. 2012.
- [25] M.Enzweiler and D.M.Gavrila: "Integrated pedestrian classification and orientation estimation", IEEE Conference on Computer Vision and Pattern Recognition, pp.982-989, June. 2010.
- [26] N.Robertson and L.Reid: "Estimating Gaze Direction from Low-Resolution Faces in Video", Proceeding of European Conference on Computer Vision, Vol.3952, pp.402-415, 2006.
- [27] A.Schulz, N.Damer, M.Fischer and R.Stiefelhagen: "Combined Head Localization and Head Pose Estimation for Video-Base Advance Driver Assistance Systems", Pattern Recognition, Vol.6935, pp.51-60, 2011.
- [28] T.Gandhi and M.M.Trivedi: "Image based estimation of pedestrian orientation for improving path prediction", IEEE Intelligent Vehicles Symposium, pp.506-5011, June. 2008.
- [29] C.Cortes and V.Vapnik: "Support-vector networks", Machine Learning, Vol.20, Issue.3, pp.273-297, Sept. 1995.
- [30] C.Chih-Chung and L.Chih-Jen: "LIBSVM: A library for support vector machines", ACM

Transactions on Intelligent Systems and Technology, Vol.2, Issue.3, pp.21-27, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2011.

[31] T.Watanabe, S.Ito and K.Yokoi: "Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection", *Advances in Image and Video Technology*, Vol.5414, pp.37-47, 2009.

[32] 山内悠嗣, 山下隆義, 藤吉弘亘: "[サーベイ論文]統計的学習手法による人検出", 電子情報通信学会 パターン認識・メディア研究会 (PRMU) 技術報告, Vol.112, No.197, pp.113-126, 2012.

[33] T.Ojala, M.Pietikainen and T.Maenpaa: "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pp.971-987, July. 2002.

[34] T.Ahonen, A.Hadid and M.Pietikainen: "Face Description with Local Binary Patterns: Application to face recognition", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.28, Issue.12, pp.2037-2041, Dec. 2006.

[35] C.Shan, S.Gong and W.Peter: "Robust facial expression recognition using local binary patterns", *IEEE International Conference on Image Processing*, pp.370-373, Sept. 2005.

[36] D.Maturana, D.Mery and A.Soto: "Face Recognition with Decision Tree-based Local Binary Patterns", *Computer Vision – ACCV 2010*, pp.618-629, 2011.

[37] C.Wojek, W.Walk and B.Schiele: "Multi-cue onboard pedestrian detection", *IEEE Conference on Computer Vision and Pattern Recognition*, pp.794-801, June. 2009.

[38] D.M.Gavrila and S.Munder: "Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle", *International Journal of Computer Vision*, Vol.73, Issue.1, 2007.

[39] P.Viola, M.J.Jones and D.Snow: "Detecting pedestrian using patterns of motion and appearance", *IEEE International Conference of Computer Vision*, Vol.2, pp.734-741, Oct. 2003.

[40] P.Dollar, C.Wojec, B.Schiele and P.Perona: "Pedestrian detection: A benchmark", *IEEE Conference on Computer Vision and Pattern Recognition*, pp.304-311, June. 2009.

[41] P.Dollar, C.Wojec, B.Schiele and P.Perona: "Pedestrian Detection: An Evaluation of the State of the Art", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.34, Issue.4, pp.743-761, April. 2012.

[42] A.Blum and T.Mitchell: "Combining labeled and unlabeled data with co-training", *Proceedings of the eleventh annual conference on Computational learning theory*, pp.92-100, 1998.

[43] X.Zhu: "Semi-Supervised Learning Literature Survey", (1530), Technical report, Computer Sciences, University of Wisconsin-Madison, 2005.

[44] O.Chapelle, B.Scholkopf and A.Zien: "Semi-Supervised Learning", MIT Press, 2006.

[45] A.Kapoor, K.Gauman, R.Urtasun and T.Darrell: "Active Learning with Gaussian Processes for Object Categorization", *IEEE International Conference on Computer Vision*, pp.1-8, Oct. 2007.

- [46] K.Benett and A.Demiriz: "Semi-supervised support vector machines", Advances in Neural Information Processing Systems, 1998.
- [47] D.Yarowsky: "Unsupervised word sense disambiguation rivaling supervised methods", Proceedings of the 33rd annual meeting on Association for Computational Linguistics, pp.189-196, 1995.
- [48] C.Rosenberg, M.Hebert, H.Schneiderman: "Semi-Supervised Self-Training of Object Detection Models", IEEE Workshops on Application of Computer Vision, Vol.1, pp.29-36, Jan. 2005.
- [49] J.Marin, D.Vazquez, D.Geronimo and A.M.Lopez: "Learning Appearance in Virtual Scenarios for Pedestrian Detection", IEEE International Conference on Computer Vision and Pattern Recognition, pp.137-144, June. 2010.
- [50] D.Vazquez, A.Lopez, J.Marin and D.Ponsa: "Virtual and Real World Adaptation for Pedestrian Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Issue.99, August. 2013.
- [51] S.Kamijo, K.Fujimura, Y.Shibayama: "Pedestrian Detection Algorithm for On-board Cameras of Multi View Angles", IEEE Intelligent Vehicles Symposium, June. 2010.
- [52] H.K.im, Y.Shibayama and S.Kamijo: "Precise Segmentation and Estimation of Pedestrian Trajectory using On-board Monocular Camera", IEICE ITS Special Issue, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E95-A, No.1, pp.296-304, Jan. 2012.
- [53] H.K.Kim, Y.Shibayama and S.Kamijo: "Precise Segmentation and Position Estimation of Pedestrians by the combination of the HOG Classifier and S-T MRF Model", The Transportation Research Board (TRB) 91st Annual Meeting, Paper#12-4526, Jan. 2012.
- [54] M.S.Arulampalam, S.Maskell, N.Gordon and T.Clapp: "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking", IEEE Transactions on Signal Processing, Vol.50, Issue.2, pp.174-188, Feb. 2002.
- [55] G.Kitagawa: "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models", Journal of Computational and Graphical Statistics, Vol.5, Issue.1, 1996.