**4.1 データセット**

**4.1.1　データセットについて**

* 使用データセットは、広角天井カメラで撮影した人物画像にアノテーション情報を付加した動画像データセットである Wide Surveillance Images from Different Heights 。
* 和歌山大学のVision and Robotics Labが提供しているデータセットで、机、椅子などの家具や荷物があるシーン中で、14名の人物がフレームインとフレームアウトを繰り返す様子が写されている。
* 広角カメラの設置高さを変更するためのクレーン付きのトラックも画面内に写っている。
* このデータセットの特徴は、3m、4m、5mという３つの異なる高さから撮影したデータが存在すること。
* 広角天井カメラからの人物検出タスクにおいて、天井カメラの高さの変更が必要な理由は、検出器をトレーニングした画像データと、推論に用いる画像データが異なると検出精度が大幅に変化するため。
* アノテーション情報は傾きのあるbounding-box情報と、人物の姿勢情報（sit/stand）。
  1. ダイアグラム

     自動的に生成された説明各bounding-boxは, [cx, cy, w, h, angle, lux, luy, rux, ruy, rbx, rby, lbx, lby, class]の14個の値で表される。

| **属性** | **説明** |
| --- | --- |
| cx, cy | 画像の左上隅を(0,0)としたときのbounding-boxの中心座標 |
| w, h | bounding-boxの幅と高さ |
| angle | 上向きの垂直軸から時計回りの回転角度(単位は度)で、範囲は-180～180 |
| lux, luy | bounding-boxの左上隅座標 |
| rux, ruy | bounding-boxの右上隅座標 |
| rbx, rby | bounding-boxの右下隅座標 |
| lbx, lby | bounding-boxの左下隅座標 |
| class | "stand person", "sit person"の2種類 |

* 動画は全部で6本。
  1. シーンは2種類あり、ショッピングセンター内店舗を想定したものとフードコート内を想定したもの。
  2. 屋内, 座る, 大きい, 鏡 が含まれている画像

     自動的に生成された説明それぞれで3m、4m、5mという３つの異なる高さから撮影。

屋内, 座る, 大きい, ブルー が含まれている画像

自動的に生成された説明座る, 大きい, オレンジ, テーブル が含まれている画像

自動的に生成された説明フードコート：3m フードコート：4m

車, 座る, 表示, 大きい が含まれている画像

自動的に生成された説明座る, 表示, 大きい, 眺め が含まれている画像

自動的に生成された説明フードコート：5m ショッピングモール店舗：3m

ショッピングモール店舗：4m ショッピングモール店舗：5m

* 動画は1920×1080で、30fpsで記録。
* 画像データは、各動画を10もしくは15フレームおきにサンプリングし、サイズを640x640に変更したjpegファイル。
* 各データの詳細は以下の通り。

| **ビデオ名** | **総フレーム数** | **アノテーション付きデータ総数(フレーム刻み幅)** | **ビデオの解像度(FPS)** | **画像の解像度** |
| --- | --- | --- | --- | --- |
| fc\_3000 | 18097 | 1200(15) | 1920×1080(30) | 640×640 |
| fc\_4000 | 15262 | 1500(10) | 1920×1080(30) | 640×640 |
| fc\_5000 | 18420 | 1700(10) | 1920×1080(30) | 640×640 |
| sc\_3000 | 15101 | 1000(10000フレームまで20、以降10) | 1920×1080(30) | 640×640 |
| sc\_4000 | 13770 | 1000(10) | 1920×1080(30) | 640×640 |
| sc\_5000 | 14369 | 1000(10) | 1920×1080(30) | 640×640 |

* 本研究は、fc\_3000の動画に焦点を当てて実験を行う。
  1. ただし、アノテーション付きデータのみを使用しているため、データ総数は1200枚の2fps動画。
  2. 加えて、後述する前処理によって600枚の1fpsとしたデータを実験に使用。

**4.1.2　データセットの前処理**

* Wide Surveillance Images from Different Heights は人物検出を想定したデータセットであり、人物追跡については想定されていない。
* そのため、人物に対する追跡IDの真値情報が付与されておらず、追跡手法として評価ができない。
* そこで、データセットの前処理としてアノテーション付きデータに対して追跡IDの真値を手作業で付与する。
* 真値付与のコストを考えて1200枚全てではなく、8枚おきの計150枚のみ行った。
* 加えて、より現実的なシチェーションを考えて、使用動画は1200枚の2fpsではなく、600枚の1fpsの動画とした。
* 14名の人物に対する追跡IDの付与には2種類の考え方がある。
  1. フレームアウトした後に再度フレームインした同一人物に同じIDを付与。
  2. フレームアウトした後に再度フレームインした同一人物に別のIDを付与。
* 屋内, 車, おもちゃ, テーブル が含まれている画像

  自動的に生成された説明ここでは真値付与のコストを考えて、同一人物に同じIDを付与する方法を採用。
* したがって、付与されたIDは1～14の14種類。

付与した追跡IDを描画した例。

* ID追加後の各bounding-boxは, [cx, cy, w, h, angle, lux, luy, rux, ruy, rbx, rby, lbx, lby, class, ID]の15個の値で構成される。
* これを真値として追跡の評価を行う。

**4.2　実験条件**

* 実験で使用する動画はフードコートを想定したシーンの3m上から撮影したfc\_3000のデータ。
* 提案した全方位に対応したDeepSORTの高精度化も行っているため、以下の4種類の実験を行い、その結果を比較する。
  1. 高精度化前の全方位DeepSORT
  2. カルマンフィルタのパラメータを調整した全方位DeepSORT
  3. 閾値付きコスト行列の閾値のパラメータを調整した全方位DeepSORT
  4. 2, 3両方を組み合わせた全方位DeepSORT（最終的な提案手法）
* 実験1について、カルマンフィルタと閾値付きコスト行列のパラメータは以下の通り。
  + カルマンフィルタ：位置の不確実性は1、速度の不確実性は0.1。
  + 閾値付きコスト行列：閾値は9.4877。
* 実験2について、カルマンフィルタと閾値付きコスト行列のパラメータは以下の通り。
  + カルマンフィルタ：位置の不確実性についてはガンマ値を「0.05, 0.3, 0.5, 0.7, 1, 3, 5, 10」とし、値の下限値を「0.1, 0.25, 0.5, 0.75, 1」とした。 また、速度の不確実性についてはガンマ値を「0.05, 0.3, 0.5, 0.7, 1, 3, 5, 10」とし、値の上限値を「0.1, 0.25, 0.5, 0.75, 1」とした。
  + 閾値付きコスト行列：閾値は9.4877。
  + 各パラメータ同士のすべての組み合わせについて実験を行う。
* 実験3について、カルマンフィルタと閾値付きコスト行列のパラメータは以下の通り。
  + カルマンフィルタ：位置の不確実性は1.0、速度の不確実性は0.1。
  + 閾値付きコスト行列：閾値はガンマ値を「0.05, 0.3, 0.5, 0.7, 1, 3, 5, 10」とし、ガンマカーブ適応後の距離に対して中心から「9.4877, 11.070, 12.592, 14.067, 15.507」と変化。
  + すべてガンマ値について実験を行う。
* 実験4について、カルマンフィルタと閾値付きコスト行列のパラメータは以下の通り。
  + カルマンフィルタ：実験2で得られた結果の内、高い精度を示したパラメータの組み合わせ。
  + 閾値付きコスト行列：実験3で得られた結果の内、高い精度を示した閾値。

**4.3　評価指標**

* MOTには複数の評価指標があり、指標同士を組み合わせて新たな指標としていることもある。
* そのため、目的に応じて適切な評価指標を用いることが求められる。
* 本実験ではMOTの性能と正しく長期間追跡できるかの観点から研究の実験を行うことにし、MOTA、IDF1、FP、FN、IDsw、FMを使用。

| **指標** | **最良** | **解説** |
| --- | --- | --- |
| MOTA（Multiple Object Tracking Accuracy） | 100% | 誤検出(FP)と見逃し(FN)、IDの変更の観点からの精度指標 |
| IDF1(ID F1 Score) | 100% | トラッキングの精度と誤検出率を同時に考慮した指標。画面から出たり入ったりする場合、その人物が同じIDを保持しているかどうかを評価する際に役立つ。 |
| FP(False Positive) | 0 | モデルから出力されたトラックがどの正解データにもマッチしなかった回数。 |
| FN(False Negative) | 0 | 正解データに存在するトラックをモデルが予測できなかった回数 |
| IDsw(ID Switch) | 0 | 正解データの2つのトラックが一時交錯する際、モデルが追跡対象の物体を誤ってしまった回数 |
| FM(Fragmentation Metric) | 0 | 1つの対象が複数のトラックに分割（対象が一時的に失われたり、別のIDに切り替わって追跡されたり）された回数とも。 |

* ネックレス, アクセサリー, ロケット が含まれている画像

  自動的に生成された説明FP(False Positive)：非追跡対称を追跡した数。
  + モデルから出力されたトラックがどの正解データにもマッチしなかった回数。最良は0。
* グラフ, 折れ線グラフ

  自動的に生成された説明FN(False Negative)：追跡対象を非追跡した数。
  + 正解データに存在するトラックをモデルが予測できなかった回数。最良は0。
* MOTA： MOTとしての総合的な性能・精度指標。
  + 誤検出(FP)と見逃し(FN)、IDの変更の観点からの精度指標で、「真値に対する追跡ミスの少なさ」を指している。最良は100%。
  + この割合が高い = FP, FN, IDsw が少ない = 正解率が高い とわかる。
* アクセサリー, ネックレス が含まれている画像

  自動的に生成された説明ID Sw（ID Switch）：2つのトラックが一時交錯する際、追跡を誤った回数。
  + 正解データの2つのトラックが一時交錯する際、モデルが追跡対象の物体を誤ってしまった回数。最良は0。
* グラフ, 折れ線グラフ

  自動的に生成された説明FM（Fragmentation Metric）：1つの対象が複数のトラックに分割された回数。
  + ここでの分割とは、対象が一時的に失われたり、別のIDに切り替わって追跡されたりすることを指す。最良は0。
  + 例えば、上記の図では2度分割されて3つのトラックが発生しているのでFM=2となる。
* 上記のようにモデルの出力結果と正解データのトラックの一致度合を数値化したものを指標としているが、それだけではMOTとしての性能をうまく表現できないケースがある。
* 例えば、開始直後にある追跡でIDが切り替わったModelAと、終了直前でIDが切り替わったModelBがあるとする。   
  一般にMOTでは同じ正解データをより長い期間追跡できている方が性能が良いと考えられるので、ModelBの方が性能の良いモデルとしたいが、MOTAで表すと同じ値となってしまう。
* そこで、物体をどれくらい長い間追跡できているかという性能を評価するため、近年はトラックのIDを基準にした指標が使用されるようになってきた。
* IDF1：追跡IDを基準にした性能・精度指標。
  + トラッキングの精度と誤検出率を同時に考慮した指標。
  + 画面から出たり入ったりする場合、その人物が同じIDを保持しているかどうかを評価する際に役立つ。
* なお、IDP・IDRはトラックのIDを基準にした指標でID Recall、ID Precisionを表す。
  + IDR は「追跡のIDレベルで、正しく追跡を行った割合」を指している。
  + この割合が高い = IDFNが少ない = 対象を正確なIDで追跡できている とわかる。
  + IDPは「追跡のIDレベルで、対象を正しく追跡した割合」を指している。
  + この割合が高い = FPが少ない = 追跡した対象が正確である とわかる。
* 既存のMOTモデルでは使用するデータセットに応じてその指標も変化しているが、近年のものはIDを考慮した指標が取り入れられていることが多い。
* ただし、これらの指標は角度について考慮されていないため、評価時はIoUの計算時にbounding-box同士の領域の積集合を求めるようにし、角度を考慮するように変更を加えている。

**4.4　実験結果**

* 以下の4つの実験結果を述べる。
  1. 高精度化前の全方位DeepSORT
  2. カルマンフィルタのパラメータを動的に調整した全方位DeepSORT
  3. 閾値付きコスト行列の閾値のパラメータを動的に調整した全方位DeepSORT
  4. 2, 3両方を組み合わせた全方位DeepSORT（最終的な提案手法）
* それぞれの実験で得られた追跡の評価結果は以下の通り。

|  | **KFパラメータ調整** | **閾値パラメータ調整** | **MOTA↑** | **IDF1↑** | **FP↓** | **FN↓** | **ID Sw↓** | **FM↓** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 実験1 | ✕ | ✕ | **55.24%** | **45.00%** | **396** | **391** | **123** | **170** |
| 実験2 | ◯ | ✕ | **63.11%** | 56.75% | **314** | **320** | 116 | **151** |
| 実験3 | ✕ | ◯ | 57.40% | **67.88%** | 389 | 395 | **82** | 167 |
| 実験4 | ◯ | ◯ | **66.26%** | **63.55%** | **298** | **304** | **84** | **155** |

* その時のパラメータの組み合わせは以下の通り。
  + 実験2：PG=0.3, VG=5, mip=0.1, mav=0.1
  + 実験3：GG=0.7
  + 実験4：GG=0.7, PG=0.3, VG=5, mip=0.1, mav=0.1
* なお、カルマンフィルタのパラメータについて、位置の不確実性のガンマ値をPG、速度の不確実性のガンマ値をVG、位置の不確実性の最小値をmip、速度の不確実性の最大値をmavとし、閾値のパラメータをGGと表記している。
* 実験1は他全ての実験に比べすべての数値が最低値となっている。
* 実験2と実験3を比較した際、MOTAやFP、FNなど追跡そのものの精度を測る指標では、実験2がより優れた結果を、IDF1やIDSwといったIDを考慮した追跡精度を測る指標では、実験3がより優れた結果を示している。
* これはつまり、カルマンフィルタのパラメータを動的に調整することはMOTの精度向上に有効であり、閾値付きコスト行列の閾値のパラメータを動的に調整することはIDに着目したMOTの精度向上に有効であると言える。
* 実験4は実験2と実験3の結果を組み合わせたものだが、全ての指標でより性能が良い結果を示している。
* 調整前の実験1の結果と比較すると、MOTAが11.02%、IDF1が18.55%向上している。
* また、FPは98（24.75%）、FNは93（22.25%）、IDSwは39（31.71%）、FMは15（8.82%）減少している。
* このことから、カルマンフィルタのパラメータと閾値付きコスト行列の閾値パラメータは、いずれも画像中の人物の位置に応じて動的に変化させることがMOTの高精度化に有効だとわかる。

**4.5　考察**

* カルマンフィルタのパラメータがMOTの精度向上に有効であり、閾値のパラメータがIDに着目したMOTの精度向上に有効である理由について、各パラメータに対する指標値の傾向から考察を行う。
* まずはカルマンフィルタのパラメータについて、各パラメータと評価指標がどれほどの関連性があるかの相関係数を算出する。グラフ, ツリーマップ図

  自動的に生成された説明

各パラメータにた評価指標の相関係数ヒートマップ

* 正の相関（パラメータ値大：指標大）が強いほどより暖色、 負の相関（パラメータ値大：指標小）が強いほどより寒色で表現している。
* 全体を通して言えることとして、PGやmavは小さいほど性能は向上し、 VGやmipについては大きいほど性能は向上していることがわかる。
* 各相関係数から、PGやVGはmipやmavに比べると各指標への影響が大きい。
* 特にVGについては他のパラメータに比べて各指標への影響が大きい。
* また、各パラメータについては以下のことが言える。
* PG：
  + MOTA、特にIDF1などはやや強いの負の相関が見られる。 つまりPGが小さいほどMOTAやIDF1などは向上する傾向にある。
  + 一方、IDスイッチやFP・FNなどでは弱い正の相関が見られる。 つまりPGが小さいほどIDスイッチやFP・FNは減少する傾向にある。
  + それぞれの相関の程度から、 PGはIDに関する指標に影響を与えることがわかる。
  + PGの値を小さくすることでMOTの精度が向上する傾向にある。
  + これはつまり、位置の不確実性を中央付近から変化させ始めることが、 追跡の精度向上に寄与することを意味する。
* VG：
  + MOTA、特にRecall・Precitionでは強い正の相関が見られる。 つまりVGが大きいほどMOTAやIDF1などは向上する傾向にある。
  + また、IDスイッチやFP・FNなどでは強い負の相関が見られる。 つまりVGが大きいほどIDスイッチやFP・FNはやや減少する傾向にある。
  + それぞれの相関の程度から、 VGは単純な追跡に関する指標に大きな影響を与えることがわかる。
  + VGの値を大きくすることで、MOTの精度が向上する傾向にある。
  + これはつまり、位置の不確実性を端の方から変化させ始めることが、 追跡の精度向上に寄与することを意味する。
* mip：
  + 全体的にVGの傾向と似ているが、各指標との相関は小さめ。
  + それぞれの相関の程度から、 mipはIDに関する指標に少し影響を与えることがわかる。
  + mipが1.0に近づくほど、MOTの精度が向上する傾向にある。
  + これはつまり、位置の不確実性は1.0に近い値から1.0に動的に増やすことが、追跡の精度向上に寄与することを意味する。
* mav：
  + 全体的にPGの傾向と似ているが、各指標との相関は小さめ。
  + それぞれの相関の程度から、 mavは単純な追跡に関する指標に少し影響を与えることがわかる。
  + mavが0.1に近づくほど、MOTの精度が向上する傾向にある。
  + これはつまり、速度の不確実性は0.1に近い値から0.1に動的に減らすことが、追跡の精度向上に寄与することを意味する。
* ただしこれらは各パラメータの傾向であり、パラメータ同士が及ぼす影響については考慮されていない。（回帰分析の件を入れることで補強できるが、解釈難易度が上がる。）
* すべてのパラメータの組み合わせで特に良い結果を示したものが、MOTAが最も高い「PG=5, VG=3, mip=0.75, mav=0.25」と、IDF1が最も高い「PG=0.3, VG=5, mip=0.1, mav=0.1」。
* 各パラメータの傾向を踏まえて、上記の結果を考察する。
* まず、単純な追跡に関する指標を向上させるにはVGとmavを重視する必要があるが、MOTAに限ると各パラメータの相関から、重視すべきはVGとPGだとわかる。
* グラフ

  自動的に生成された説明X軸にVG、Y軸にPG、Z軸にMOTAをプロットし、MOTAの値によってヒートマップを適応すると以下のようになる。

VG, PG, MOTAの関係を示したグラフ

* そのため、MOTAが最も高い「PG=5, VG=3, mip=0.75, mav=0.25」では、mavは小さめの値だが、VGとPGを重視した結果だと考察できる。 （ただしMOTA上位15個の内VG=3はTopのみで他は5や10なため、図から分かる通り、上振れて傾向から外れている可能性もある。）
* また、IDに関する指標を向上させるにはPGとmipを重視する必要があるが、IDF1に限ると各パラメータの相関から、重視すべきはPGとVGだとわかる。
* X軸にPG、Y軸にVG、Z軸にIDF1をプロットし、IDF1の値によってヒートマップを適応すると以下のようになる。グラフ

  自動的に生成された説明

PG, VG, IDF1の関係を示したグラフ

* PGは小さく、VGは大きくすればIDF1が向上する傾向があることがわかる。
* そのため、IDF1が最も高い「PG=0.3, VG=5, mip=0.1, mav=0.1」では、mipは小さめの値だが、PGとVGを重視した結果だと考察できる。