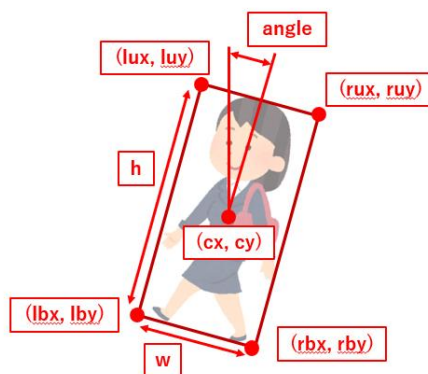


## 4.1 データセット

### 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. 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



lbox, 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. それぞれで 3m、4m、5m という 3 つの異なる高さから撮影。



フードコート：3m



フードコート：4m



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  $1920 \times 1080$  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  $640 \times 640$  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)	$1920 \times 1080(30)$	$640 \times 640$
fc_4000	15262	1500(10)	$1920 \times 1080(30)$	$640 \times 640$
fc_5000	18420	1700(10)	$1920 \times 1080(30)$	$640 \times 640$
sc_3000	15101	1000(10000 フレームまで 20、以降 10)	$1920 \times 1080(30)$	$640 \times 640$
sc_4000	13770	1000(10)	$1920 \times 1080(30)$	$640 \times 640$
sc_5000	14369	1000(10)	$1920 \times 1080(30)$	$640 \times 640$

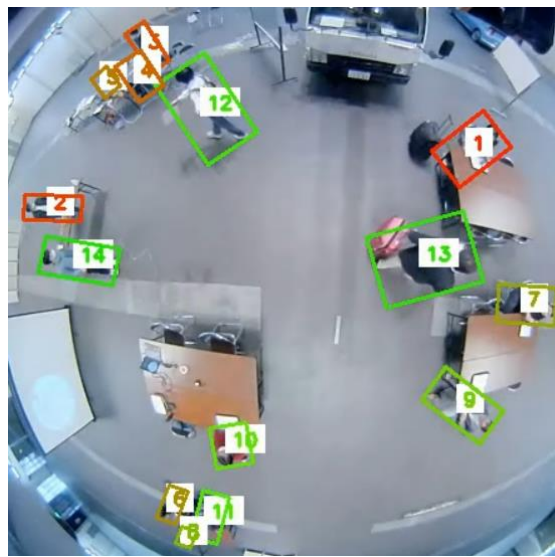
- 本研究は、fc\_3000 の動画に焦点を当てて実験を行う。
  1. ただし、アノテーション付きデータのみを使用しているため、データ総数は 1200 枚の 2fps 動画。
  2. 加えて、後述する前処理によって 600 枚の 1fps としたデータを実験に使用。

#### 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、2fps 分の 1200 枚全てではなく、~~8枚おきの~~**0.25fps 分である計 150 枚のみ行った。**
- 加えて、より現実的なシチュエーションを考えて、使用動画は 1200 枚の 2fps でなく、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。
- ~~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。~~
- **実装した DeepSORT は 70 フレーム間オクルージョンが発生するとその ID は棄却さ**

れる。

- つまり、70 フレーム以上オクルージョンが発生するケースが存在するならば、理想的な追跡はその人物に対して再度同じ ID を割り当てる事が不可能になるため、真値として同一人物に同じ ID を付与する方法は不適切である。
- しかし、今回使用する 1fps の動画には 70 フレームを超えるオクルージョンは存在しないため、理想的な追跡はその人物に対して再度同じ ID を割り当てる事が可能である。
- そのため、ここでは真値付与の方法として同一人物に同じ 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。~~
  - ~~高精度化前の全方位 DeepSORT~~
  - ~~カルマンフィルタのパラメータを調整した全方位 DeepSORT~~
  - ~~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~~
  - ~~2, 3 両方を組み合わせた全方位 DeepSORT (最終的な提案手法)~~

- 従来の DeepSORT を全方位画像に対応できるように拡張したモデルを手法 1 とし、その高精度化を行ったモデルを手法 2 とする。
- また、手法 2 については異なる 2 種類の方法で高精度化を行ったため、以下のように定義する。
  - 手法 2.1: カルマンフィルタのパラメータを調整した全方位 DeepSORT
  - 手法 2.2: 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
- これを踏まえ、各方法のみでの場合と 2 種類を組み合わせた場合とでそれぞれ評価を行い、手法 1 の評価結果と比較を行う。
- したがって、以下の 4 種類の実験を行う。
  - 実験 1: 手法 1 の、全方位に対応した DeepSORT が設計できたか
  - 実験 2: 手法 2.1 は手法 1 と比較して高精度化を達成したか
  - 実験 3: 手法 2.2 は手法 1 と比較して高精度化を達成したか
  - 実験 4: 手法 2 は手法 1 と比較して高精度化を達成したか
- 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。
  - **なお、手法 2.1 は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った後、最適な組み合わせを最終的な結果としている。**
- 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。
  - **なお、手法 2.2 はすべてのガンマ値について実験を行った後、最も優れた値を最終的な結果としている。**
- 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。

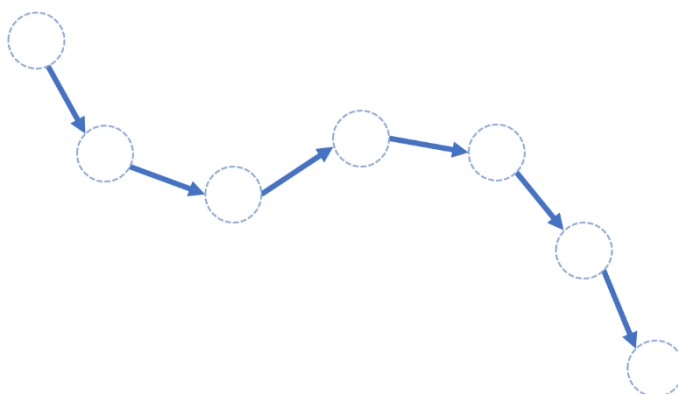


### 4.3 評価指標

- MOT には複数の評価指標があり、指標同士を組み合わせる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。

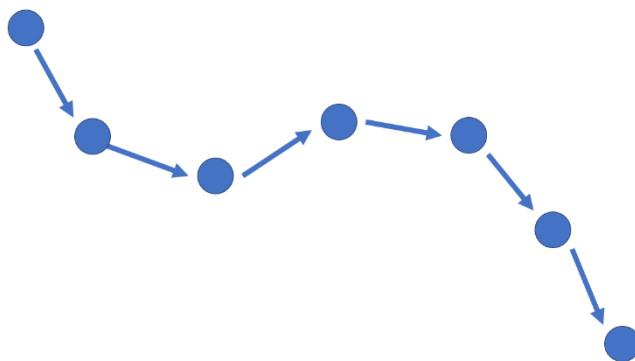
指標	最良	解説
MOTA (Multiple Object Tracking Accuracy)	100%	誤検出(FP)と見逃し(FN)、ID の変更の観点からの精度指標
IDF1(ID F1 Score)	100%	トラッキングの精度と誤検出率を同時に考慮した指標。画面から出たり入ったりする場合、その人物が同じ ID を保持しているかどうかを評価する際に役立つ。
FP(False Positive)	0	モデルから出力されたトラックがどの正解データにもマッチしなかった回数。
FN(False Negative)	0	正解データに存在するトラックをモデルが予測できなかった回数
IDsw(ID 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。

- FP(False Positive)：非追跡対象を追跡した数。



- 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。

- 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。

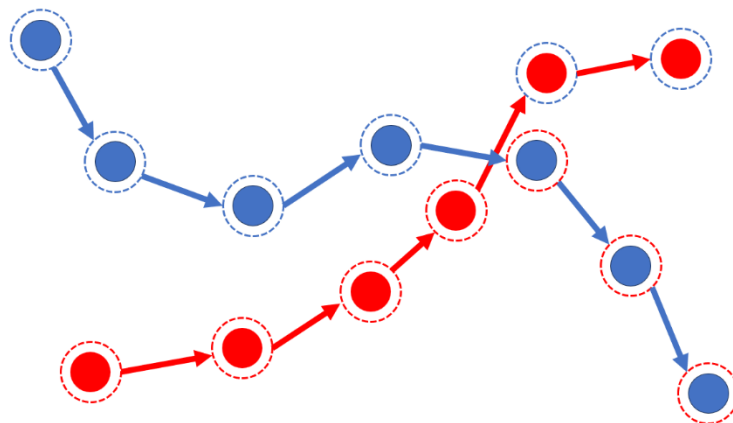
- MOTA： MOT としての総合的な性能・精度指標。

$$MOTA(\%) = \left( 1 - \left( \frac{FP + FN + ID\ sw}{gt} \right) \right) \times 100$$

➤ 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%。

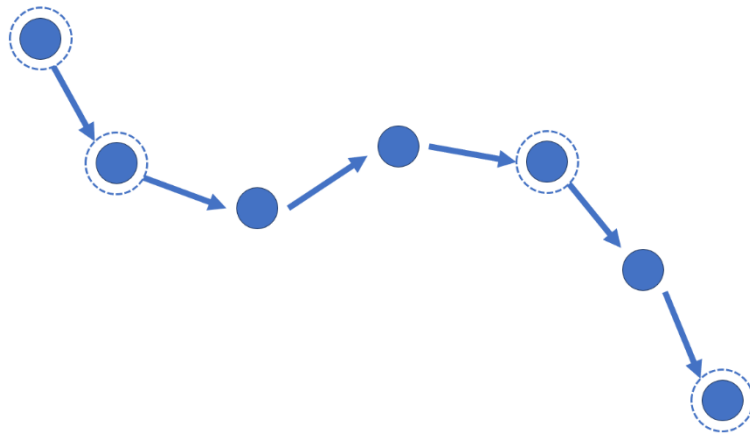
➤ この割合が高い = FP, FN, IDsw が少ない = 正解率が高い とわかる。

- 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。

- 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。

$$IDF1(\%) = \left( \frac{2 \times IDR \times IDP}{IDR + IDP} \right) \times 100$$

- 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。
- 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。

$$IDR(\%) = \left( \frac{IDTP}{IDTP + IDFN} \right) \times 100 \quad IDR(\%) = \left( \frac{IDTP}{IDTP + IDFP} \right) \times 100$$

- 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)
- 5. 実験 1: 手法 1 の、全方位に対応した DeepSORT が設計できたか
- 6. 実験 2: 手法 2.1 は手法 1 と比較して高精度化を達成したか
- 7. 実験 3: 手法 2.2 は手法 1 と比較して高精度化を達成したか
- 8. 実験 4: 手法 2 は手法 1 と比較して高精度化を達成したか
- 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。
- ~~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	⊖	✕	<u>63.11%</u>	56.75%	<u>314</u>	<u>320</u>	116	151
実験 3	✕	⊖	57.40%	<b>67.88%</b>	389	395	<b>82</b>	167

実験 4	⊖	⊖	66.26%	<u>63.55%</u>	298	304	<u>84</u>	<u>155</u>
---------	---	---	--------	---------------	-----	-----	-----------	------------

- 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

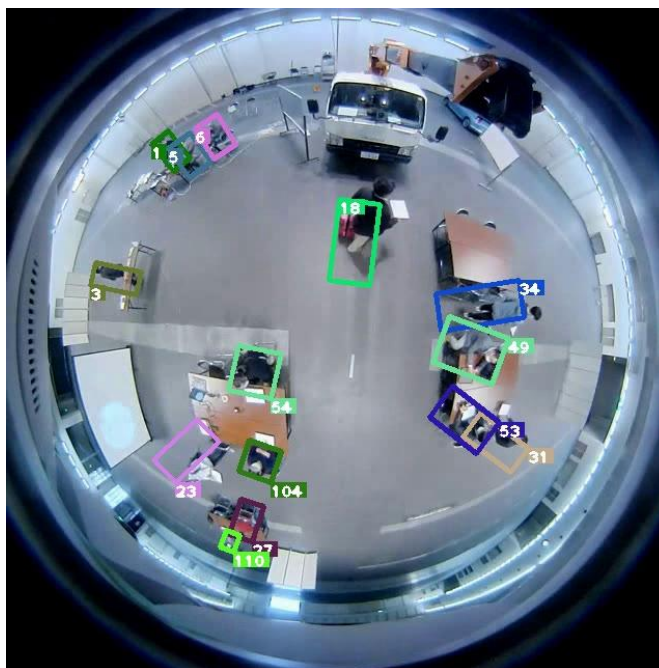
➤—実験 3 : GG=0.7

➤—実験 4 : GG=0.7, PG=0.3, VG=5, mip=0.1, mav=0.1

- 実験 1 について、手法 1 から得られた追跡の評価結果は以下の通りである。

	KF パラメータ調整	閾値パラメータ調整	MOTA ↑	IDF1 ↑	FP ↓	FN ↓	ID Sw ↓	FM ↓
手法 1	×	×	55.24%	45.00%	396	391	123	170

- 実験 1 における追跡の結果例を以下に示す。



- 従来の DeepSORT では対応していない、角度情報が付与された bounding-box に対応した追跡が実現できていることがわかる。
- しかし、課題点として画像の中央付近での追跡が困難である点や、立っている状態から座る状態に変化する際に ID スイッチが発生し、異なる ID が割り振られる点などが挙げられる。
- そのような例を以下に示す。

➤ 画像の中央付近での追跡に失敗する例



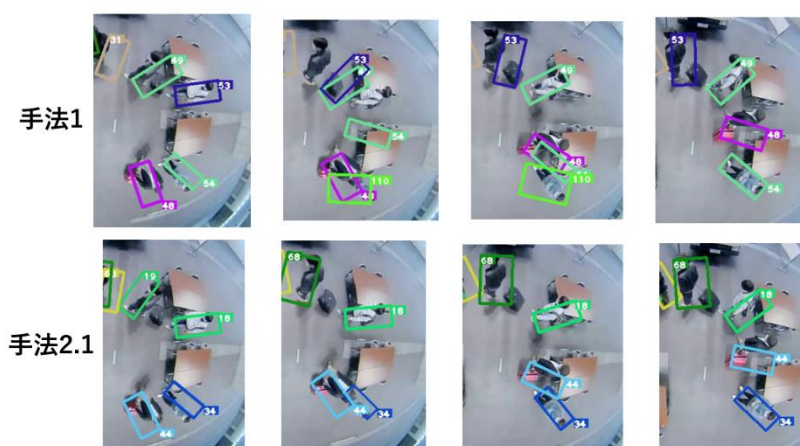
➤ 立っている状態から座る状態に変化する際に異なる ID が割り振られる例



- こうした課題点の解決のために手法 1 の高精度化を目指したものが手法 2 であり、その結果については実験 4 で確かめる。
- 実験 2 について、手法 2.1 から得られた追跡の評価結果は以下の通りである。

	KF パラメータ調整	閾値パラメータ調整	MOTA ↑	IDF1 ↑	FP ↓	FN ↓	ID Sw ↓	FM ↓
手法 2.1	○	×	63.11%	56.75%	314	320	116	151

- このときのパラメータの組み合わせは「PG=0.3, VG=5, mip=0.1, mav=0.1」である。
- 手法 1 の結果と比較すると、手法 2.1 は MOTA が 7.87%、IDF1 が 11.75% 向上している。
- また、FP は 82、FN は 71、IDSw は 7、FM は 19 減少している。
- 実験 2 の結果例として、手法 1 と手法 2.1 の追跡結果を動画から比較したものを以下に示す。



- 手法 1 では上下の 2 人でそれぞれ ID スイッチが発生しているが、手法 2.1 では左上の人物でしか発生しておらず、他 3 人は適切な追跡できていることが確認できる。
- また、左と比較して右の例は ID が小さいが、これは過去フレームでの人物と適切にマッチングが行われているためといえる。
- 指標の表で手法 2.1 はすべての指標で手法 1 を上回っていることと、上記の結果例から、カルマンフィルタのパラメータを動的に調整することで全方位 DeepSORT の高精度化を達成したと言える。
- 実験 3 について、手法 2.2 から得られた追跡の評価結果は以下の通りである。

	KF パラメータ調整	閾値パラメータ調整	MOTA ↑	IDF1 ↑	FP ↓	FN ↓	ID Sw ↓	FM ↓
手法 2.2	×	○	57.40%	67.88%	389	395	82	167

- このときの閾値パラメータ値は「GG=0.7」である。
- 手法 1 の結果と比較すると、手法 2.2 は MOTA が 2.16%、IDF1 が 22.88%向上している。
- また、FP は 7、IDSw は 41、FM は 3 減少している一方、FN は 4 増加している。
- 実験 3 の結果例として、手法 1 と手法 2.2 の追跡結果を動画から比較したものを以下に示す。

### 手法1



### 手法2.2



- 手法 1 では中央付近を経由する際に ID スイッチが発生しているケースが見られるが、手法 2.2 では ID スイッチが起こらず、追跡の質が改善していると言える。
- 指標の表で手法 2.2 はほぼすべての指標で手法 1 を上回っていることと、上記の結果例から、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で全方位 DeepSORT の高精度化を達成したと言える。

- 実験 4 について、手法 2.1 と手法 2.2 を組み合わせた手法 2 から得られた追跡の評価結果は以下の通りである。

	KF パラメータ調整	閾値パラメータ調整	MOTA ↑	IDF1 ↑	FP ↓	FN ↓	ID Sw ↓	FM ↓
手法 2	○	○	66.26%	<u>63.55%</u>	298	304	<u>84</u>	<u>155</u>

- このときのパラメータの組み合わせは「GG=0.7, PG=0.3, VG=5, mip=0.1, mav=0.1」である。
- 調整前の実験 1 の結果と比較すると、~~手法 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、FN は 93、IDSw は 39、FM は 15 減少している。
- 実験 1 での結果として手法 1 では画像の中央付近での追跡が困難である点や、立っている状態から座る状態に変化する際に ID スイッチが発生し、異なる ID が割り振られる点などを課題点として挙げたが、手法 2 について同じフレームの結果を確認すると以下の通りであった。

- 画像の中央付近での追跡に失敗していた例

手法1



手法2



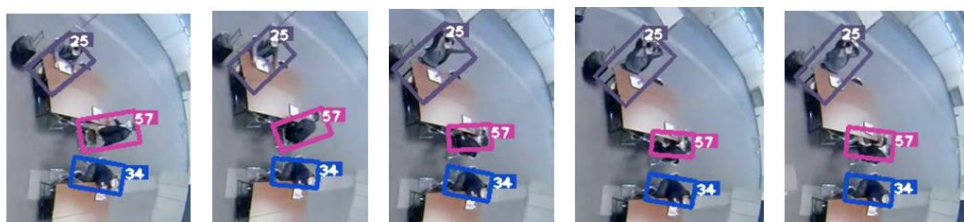


- 立っている状態から座る状態に変化する際に異なる ID が割り振られていた例

### 手法1



### 手法2



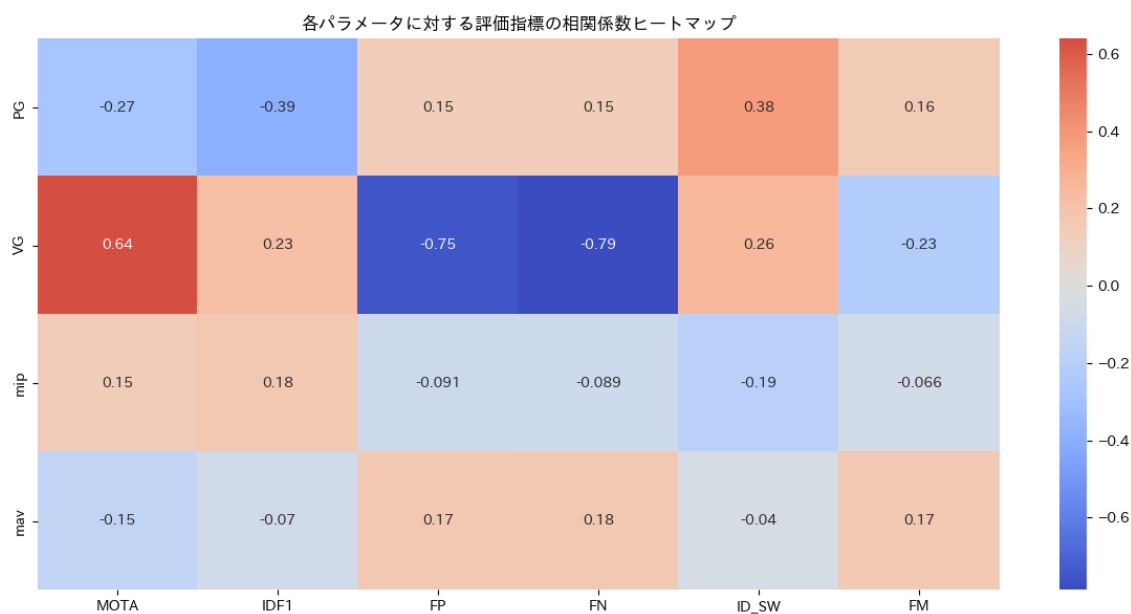
- 指標の表で手法 1 の値を上回っていることと上記の結果例から、手法 2 は手法 1 に比べ高精度化を達成していると言える。
- ところで、実験 2 手法 2.1 と実験 3 手法 2.2 を比較した際、MOTA や FP、FN など追跡そのものの精度を測る指標では、実験 2 手法 2.1 がより優れた結果を、IDF1 や IDSw といった ID を考慮した追跡精度を測る指標では、実験 3 手法 2.2 がより優れた結果を示している。
- **したがって**、カルマンフィルタのパラメータを動的に調整することは MOT の精度向上に有効であり、閾値付きコスト行列の閾値のパラメータを動的に調整することは ID に着目した MOT の精度向上に有効であると言える。
- **つまり**、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。

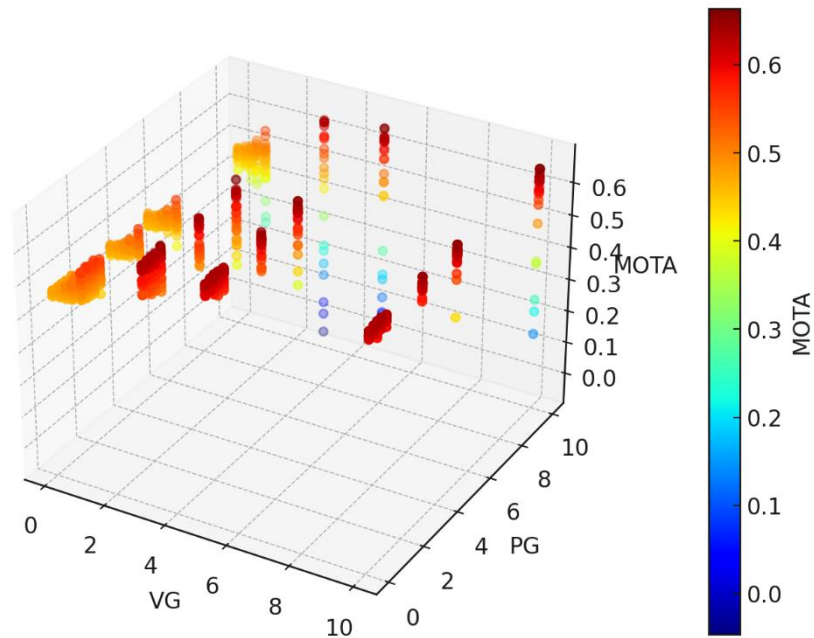


各パラメータに対する評価指標の相関係数ヒートマップ

- 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 などやや強い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・Precision 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。

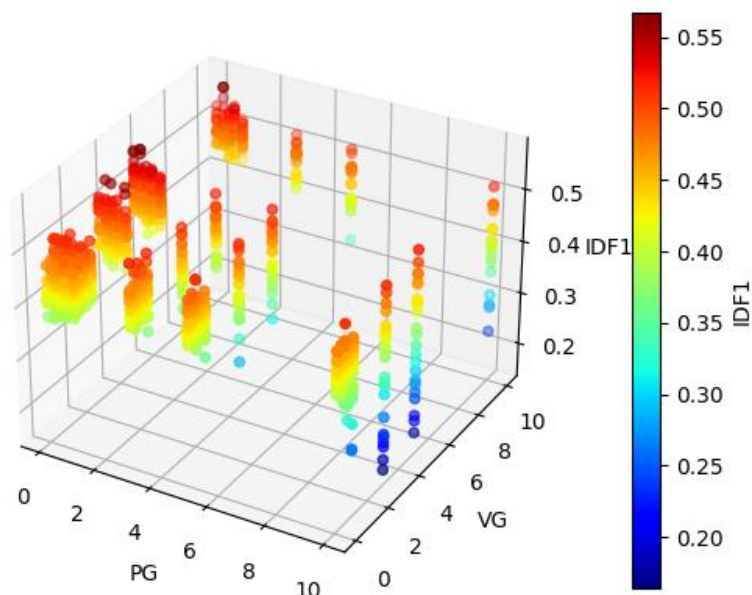
- X軸に VG、Y軸に PG、Z軸に MOTA をプロットし、MOTA の値によってヒートマップを適応すると以下ようになる。



VG, PG, MOTA の関係を示したグラフ

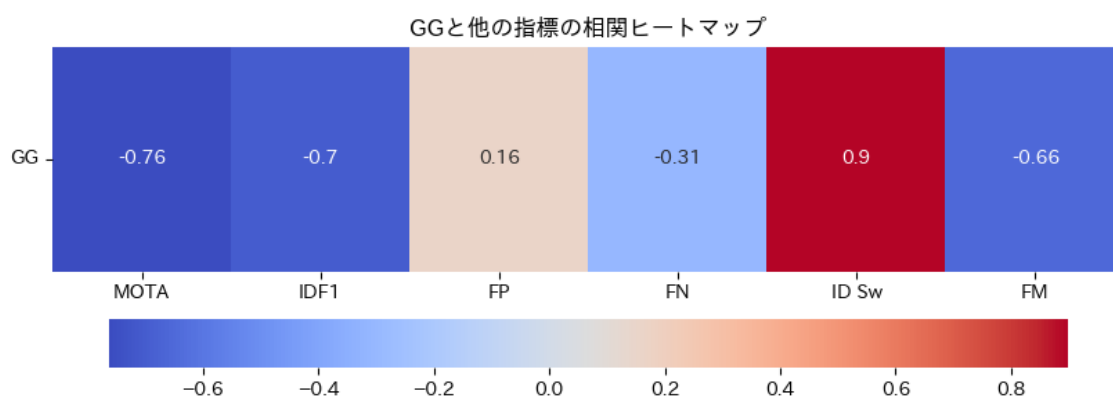
- そのため、MOTA が最も高い「PG=5, VG=3, mip=0.75, mav=0.25」では、mav は小さな値だが、VG と PG を重視した結果だと考察できる。（ただし MOTA 上位 15 個の内 VG=3 は Top のみで他は 5 や 10 なため、図から分かる通り、上振れて傾向から外れている可能性もある。）
- また、ID に関する指標を向上させるには PG と mip を重視する必要があるが、IDF1 に限ると各パラメータの相関から、重視すべきは PG と VG だとわかる。

- 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ようになる。



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 は小さな値だが、PG と VG を重視した結果だと考察できる。
- 次に、閾値のパラメータ GG について、GG 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。



- 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。

- ここから、GG は小さいほど MOTA や IDF1、IDSw が優れた値を示すことがわかる。
- しかし、FN や FM には MOTA や IDF1 と同じ負の相関が存在し、GG を小さくしすぎると長期的な追跡の精度が下がってしまうこともわかる。
- したがって、手法 2.2 の結果から、最良の GG は GG=0.7 であることからわかるように、GG は 0 以上 1 未満の値を選ぶことがバランスよく高精度な結果を得られると言える。
- GG が 0 以上 1 未満ということは、ガンマカーブは上に凸の形を描く。
- これは閾値が小さい値を取る範囲が狭く、大きい値を取る範囲が広いことを指している。
- つまり、追跡に使用する情報は bounding-box の位置情報よりも、人物の外観情報をやや重視するが有効だといえる。
- 固定値に比べ優れた結果を得たことから、中心からの位置に応じて閾値を動的に変化させることは、全方位 DeepSORT の高精度化に有効だとわかった。
- 最後に、カルマンフィルタと閾値のパラメータを動的に調整した手法 2 の結果について考察する。

- 以下に各手法での評価結果を示す。

	KF パラメータ調整	閾値パラメータ調整	MOTA ↑	IDF1 ↑	FP ↓	FN ↓	ID Sw ↓	FM ↓
手法 1	×	×	55.24%	45.00%	396	391	123	170
手法 2.1	○	×	<u>63.11%</u>	56.75%	<u>314</u>	<u>320</u>	116	151
手法 2.2	×	○	57.40%	67.88%	389	395	82	167
手法 2	○	○	66.26%	<u>63.55%</u>	298	304	<u>84</u>	<u>155</u>

- 手法 2 は、手法 2.1 と手法 2.2 の最良の結果を示したパラメータを組み合わせたもので、「GG=0.7, PG=0.3, VG=5, mip=0.1, mav=0.1」である。
- 手法 2.1 の結果からカルマンフィルタのパラメータ調整により追跡としての精度が向上し、手法 2.2 の結果から閾値のパラメータ調整により ID を考慮した追跡の精度が向上したことがわかる。
- それらの結果を組み合わせた手法 2 の結果から、互いの良い結果のみを抽出したような優れた値を得られたことがわかる。

- これはカルマンフィルタと閾値付きコスト行列が互いに独立した処理を行っているためだと考えられる。
- カルマンフィルタは予測結果と実際に検出した bounding-box の中間の値を最終的な予測結果として得るが、予測結果と実際に検出した bounding-box が離れているほど関連性は低く、予測の精度も低下する。
- この関連性に対して閾値処理を行うため、カルマンフィルタと閾値付きコスト行列の処理は独立していると言える。
- つまり、手法 2 ではカルマンフィルタのパラメータ調整によって得られた高精度な結果に対して閾値処理を行っており、手法 2.1 での IDF! や IDS<sub>w</sub> は手法 2.2 によってより高精度化されていると考えられる。