

1

## Human Perception Inspired Multimodal Object Recognition by Considering Time-Dependent Reliance on Modalities

大阪大学大学院 情報科学研究科  
情報ネットワーク学専攻 村田研究室  
安藤 颯人

修士論文発表会 25.02.07

1

2

### 研究背景

- Smart Manufacturing (SM) の発展<sup>[1-4]</sup>
  - 労働力不足やエネルギー資源の有効活用
  - 協調ロボット (Collaborative Robot) の重要性
- 協調ロボット
  - 作業空間を最大限活用可能
  - 単調/力の必要な仕事を分担
  - 人間は高度な作業に集中可能
- 安全性
  - 複数のセンサを搭載
  - 危険停止エリア/緊急停止ボタン
- 課題
  - 作業員の安全確保のための認識技術

図: 協調ロボット導入のイメージ図 (Figure. 1)

表: 各種センサの特徴<sup>[5-8]</sup>

Sensor	Advantages	Disadvantages
Camera	色や形状の識別精度	照明条件の影響
Radar	速度の検知	形状の詳細な識別が困難
LIDAR	距離や形状の三次元把握	透明/反射物の影響

2

3

### 研究分野の概要

- マルチモーダル物体認識<sup>[8-12]</sup>
  - 各センサ情報から検出されたクラスと検出の精度を表すスコアを生成
    - Early Fusion : 特徴量の段階で統合
    - Late Fusion : 各認識結果を統合
- 研究分野における課題
  - 各センサデータの信頼度 (Reliability) への注目不足
    - 照明条件の変化に伴う映像から得られるスコアの信頼度は変動
    - 対象が近づくと映像やradarから得られるスコアの信頼度は変動
- センサの信頼性を加味した物体認識手法
  - 先行研究<sup>[17,18]</sup>では人の認知<sup>[19,20,36-38]</sup>を活用 (Late Fusion-Base)
    - 人はノイズを含むセンサ情報を環境に応じて重みづけし意思決定
    - 観測データをもとにセンサ情報を静的に重みづけし統合に活用
  - 実環境は動的に変化しているため観測データに基づき動的に重みづけする必要性

図: 検出の例

図: 信頼性の変化例

3

4

### 分散を用いた信頼度の定義

- 情報の品質を反映させた重みの設定<sup>[19,20]</sup>
  - 感覚器から得られる情報は感覚器により異なる
  - 統合の際の重みは情報の品質を反映させていなければならない<sup>[19]</sup>
  - 人は認識の際に線形に重みづけ・統合を行っているとする研究<sup>[19,20]</sup>
  - モダリティ  $i$  からのスコアの分散 ( $\sigma_i^2$ ) が大きい
  - そのモダリティからの認識が安定していないことを示唆
  - これを用いて信頼度 ( $r_i$ ) を定義

$$r_i = 1/\sigma_i^2$$

図: 信頼性の変化例

4

5

### 研究目的とアプローチ

- 目的
  - 時間変化するセンサの信頼性を考慮した物体認識手法の検証
- アプローチ
  - センサの信頼性を動的に評価する仕組みを取り入れたマルチモーダル統合手法の導入
    - 各モダリティからの認識スコアを統合
    - 人の認知モデル<sup>[19,20]</sup>を活用した統合方式
  - オブジェクト毎に閾値を設定
    - 対象に応じて柔軟な認識を実現
  - 実映像データを用いた認識結果とシミュレーションを用いた統合手法の実験的な検証
    - 実映像をもとに得られた認識結果とシミュレーションにより生成された仮想的なモダリティのスコアを統合
    - どのようなスコアが与えられた場合にどのように対応できるかを複数パターン検証
    - 算術平均を活用する場合や最大値を選択する場合と比較
    - 不安定なセンサスコアの影響度低減と認識の安定性向上

5

6

### 信頼度重み付き認識手法概要 (1/2)

- 2種類のモダリティからの入力
  - 例) Point Cloud-Base 認識手法, RGB-Base 認識手法
  - 同じオブジェクトに対する認識結果を統合

図: 信頼度変化のシチュエーション例

図: 提案手法概要図 (Figure. 2)

6

### 信頼度重み付き認識手法概要 (2/2)

7

- それぞれの平均/分散を推定
  - 信頼できるセンサの情報により反映
- 認識結果を統合
  - 得られた信頼度に基づき統合

図: 信頼度変化のシチュエーション例

図: 提案手法概要図 (Figure. 2)

7

### 信頼度重み付き統合手法<sup>[19,20]</sup>(1/3)

8

- 各モダリティの信頼度の算出
  - 時刻 $(t)$ において各モダリティから得られる平均  $\mu_{(t)}$
  - 時刻 $(t)$ において各モダリティから得られる分散  $\sigma_{(t)}^2$
- スコアを統合
  - 時刻 $(t)$ において各モダリティから得られるスコア  $I_{(t=T)} = \frac{\mu_{1(t)}/\sigma_{1(t)}^2 + \mu_{2(t)}/\sigma_{2(t)}^2 + \mu_{p(t)}/\sigma_{p(t)}^2}{1/\sigma_{1(t)}^2 + 1/\sigma_{2(t)}^2 + 1/\sigma_{p(t)}^2}$
- スコアの急な変動を抑制する調整項
  - $\mu_{p(t=T+1)} = I_{(t=T)}$
  - $\sigma_{p(t=T+1)}^2 = \left( \frac{1}{\sigma_{1(t)}^2} + \frac{1}{\sigma_{2(t)}^2} + \frac{1}{\sigma_{p(t)}^2} \right)^{-1}$

8

### 信頼度重み付き統合手法<sup>[19,20]</sup>(2/3)

9

- 各モダリティの信頼度の算出
  - $\mu_{(t)} = \frac{1}{W} \sum_{i=1}^W x_{(t-i)}$
  - $\sigma_{(t)}^2 = \frac{1}{W-1} \sum_{i=1}^W [x_{(t-i)} - \mu_{(t)}]^2 + \epsilon$
- スコアを統合
  - モダリティ1, モダリティ2, 前フレームまでの情報
  - 統合スコア  $I_{(t=T)} = \frac{\mu_{1(t)}/\sigma_{1(t)}^2 + \mu_{2(t)}/\sigma_{2(t)}^2 + \mu_{p(t)}/\sigma_{p(t)}^2}{1/\sigma_{1(t)}^2 + 1/\sigma_{2(t)}^2 + 1/\sigma_{p(t)}^2}$
- スコアの急な変動を抑制する調整項
  - $\mu_{p(t=T+1)} = I_{(t=T)}$
  - $\sigma_{p(t=T+1)}^2 = \left( \frac{1}{\sigma_{1(t)}^2} + \frac{1}{\sigma_{2(t)}^2} + \frac{1}{\sigma_{p(t)}^2} \right)^{-1}$

9

### 信頼度重み付き統合手法<sup>[19,20]</sup>(3/3)

10

- 各モダリティの信頼度の算出
  - $\mu_{(t)} = \frac{1}{W} \sum_{i=1}^W x_{(t-i)}$
  - $\sigma_{(t)}^2 = \frac{1}{W-1} \sum_{i=1}^W [x_{(t-i)} - \mu_{(t)}]^2 + \epsilon$
- スコアを統合
  - $I_{(t=T)} = \frac{\mu_{1(t)}/\sigma_{1(t)}^2 + \mu_{2(t)}/\sigma_{2(t)}^2 + \mu_{p(t)}/\sigma_{p(t)}^2}{1/\sigma_{1(t)}^2 + 1/\sigma_{2(t)}^2 + 1/\sigma_{p(t)}^2}$
- スコアの急な変動を抑制する調整項
  - スコアの急な変動を抑えられる  $\mu_{p(t=T+1)} = I_{(t=T)}$
  - $\sigma_{p(t=T+1)}^2 = \left( \frac{1}{\sigma_{1(t)}^2} + \frac{1}{\sigma_{2(t)}^2} + \frac{1}{\sigma_{p(t)}^2} \right)^{-1}$

10

### シミュレーションを用いた評価

11

- RGBベースの認識スコアと仮想的なモダリティのスコアを統合
  - RGBベースの認識スコア
    - 撮影したデータを Stream YOLO<sup>[29]</sup> を用いて解析したスコアを利用
    - 実機ベースの認識結果としてRGBベースの認識スコアは同じものを使用
  - 仮想的なモダリティの認識スコア
    - 区間に応じて平均と分散を変化させるスコアをシミュレーションで複数生成
    - Low / Medium / High の3段階で平均/分散を変化できるよう設計
- 信頼度が変化する入力を与え評価
  - 信頼度が低い状態から高くなる状態を以降で説明
  - 他の統合手法と比較
    - 算術平均を使う手法
    - 最大値を選択する手法

Parameter	Value
Detection Threshold ( $\theta$ )	0.65, 0.75, 0.85
Time Window ( $W$ )	5, 10
Prior Distribution ( $\mu_p, \sigma_p$ )	0.5, 0.05
Simulated Modality Pattern ( $p$ )	1, 2
Simulated Modality	$p=1, (0.15, 0.75, 0.90)$
Average Score ( $\mu_1, \mu_m, \mu_h$ )	$p=2, (0.25, 0.50, 0.80)$
Simulated Modality	$p=1, (0.015, 0.030, 0.13)$
Standard Deviation ( $\sigma_1, \sigma_m, \sigma_h$ )	$p=2, (0.030, 0.050, 0.15)$

表: 設定した主要パラメータ (Table. 1)

図: 撮影シーン

11

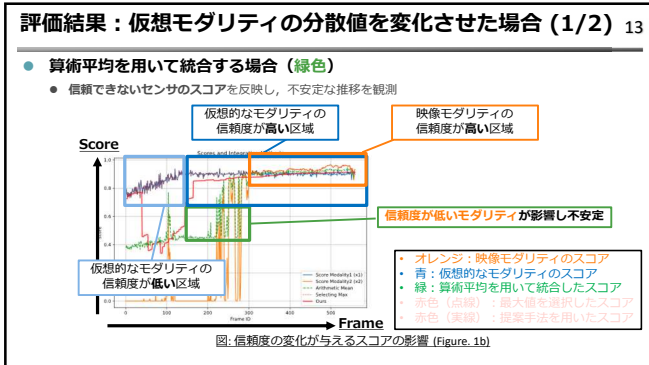
### 評価結果: 仮想モダリティの分散値を変化させた場合 (1/2)

12

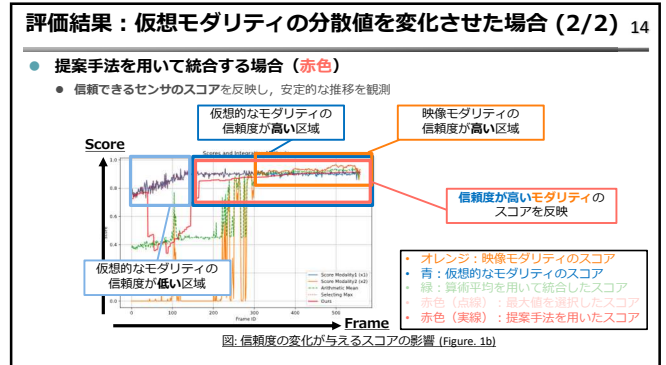
- 算術平均を用いて統合する場合
  - 信頼できないセンサのスコアを反映し、不安定な推移を観測

図: 信頼度の変化が与えるスコアの影響 (Figure. 1b)

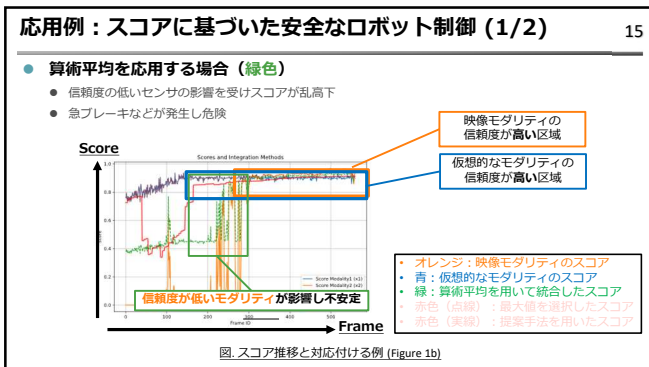
12



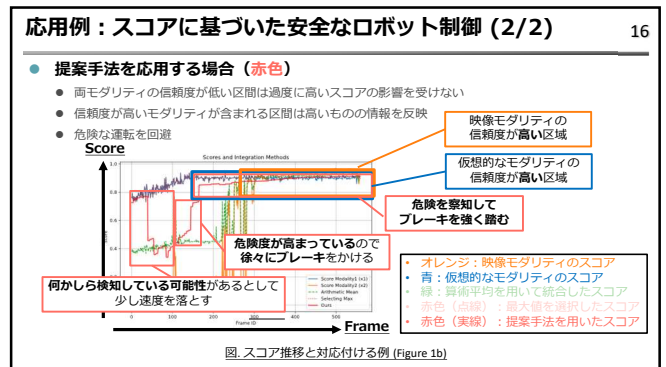
13



14



15



16

### まとめ 17

- 提案手法
  - 時間変化するセンサ信頼性を考慮したマルチモーダル認識手法
  - MLE model<sup>[19,20]</sup> を使った信頼性の重みづけ
- 実験結果
  - 実測データを用いたRGBベースの認識結果と仮想的に生成したモダリティのスコアを統合し検証
  - 統合手法は比較手法と比べて信頼度が低いスコアの影響度を抑えて統合することを確認
- 今後の課題
  - 各モダリティでの認識結果の組み合わせ
  - 実機を用いた検証
  - 安全なロボット制御への適用

17

### 付録 18

- 撮影環境
- 定性的な評価
  - 低信頼度・低いスコアと統合する場合
- 定量的な評価
  - 混合行列の定義
  - 評価指標の定義
  - 評価結果

18

### 撮影環境 19

- 撮影データ
  - <https://youtu.be/ZiwWpQXHO4>
- 撮影場所
  - 情報科学研究科A棟GF

クラスタリング処理結果 (縮小)

Route  
Camera  
Pedestrian

43.20m  
図: 撮影環境 (Figure. 4)

19

### 低信頼度/低いスコアと統合する場合 20

- 信頼できるセンサのスコアを反映
  - 算術平均や最大値を使用する場合にはスコアが低下

- オレンジ: 映像モダリティのスコア 20
- 青: 仮想的なモダリティのスコア
- 緑: 算術平均を用いて統合したスコア
- 赤色 (点線): 最大値を選択したスコア
- 赤色 (実線): 提案手法を用いたスコア

信頼度が高いモダリティのスコアを反映  
映像モダリティの信頼度が高い区域  
信頼度が低いモダリティの影響しスコアが低い  
仮想的なモダリティの信頼度が低い区域

Score  
Frame

Fig. 信頼度の変化が与えるスコアの影響 (Fig2c. 1f)

20

### 混合行列の定義 21

- 閾値を設定して評価
  - スコアが閾値を超えるべき場合:
    - 各フレームにおいて、信頼できるモダリティが存在し、かつそのスコアが閾値を超えている場合
    - 統合スコアも閾値を超えていることが望ましい
  - スコアが閾値を超えてはいけくない場合:
    - どちらのモダリティも信頼できない、もしくは信頼できるモダリティが閾値を超えていない場合
    - 統合スコアも閾値を超えていないことが望ましい
- 混合行列
  - スコアが閾値を超えるべき場合に超えている場合 : TP, 正しく統合できているフレーム数
  - スコアが閾値を超えるべき場合に超えていない場合 : FN, 見落としが発生したフレーム数
  - スコアが閾値を超えてはいけくない場合に超えている場合 : FP, 誤認識が発生したフレーム数
  - スコアが閾値を超えてはいけくない場合に超えていない場合 : TN, 正しく無視できたフレーム数

21

### 評価指標の定義 22

- 混合行列より以下を定義
  - Accuracy
    - 統合スコアが全体のフレームに対してどの程度正しく分類できているかを評価する指標
    - この指標が高いほど分類精度が高いことを示す
$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$
  - Precision
    - 統合スコアが閾値を超えたフレームのうち、実際に閾値を超えているべきフレームの割合を評価する指標
    - この指標が高いほど誤認識 (False Positive) を抑えながら正しく検出できていることを示す
$$Prec = \frac{TP}{TP + FP}$$
  - Recall
    - 実際に閾値を超えているべきフレームのうち、統合スコアが閾値を超えた割合を評価する指標
    - この指標が高いほど見落とし (False Negative) を抑えながら正しく検出できていることを示す
$$Rec = \frac{TP}{TP + FN}$$
  - Specificity
    - 実際に閾値を超えてほしくないフレームのうち、統合スコアが閾値を超えなかった割合を評価する指標
    - この指標が高いほど誤認識 (False Positive) を抑え、不要な検出を防ぐことができていることを示す
$$Spec = \frac{TN}{TN + FP}$$

22

### 定量的評価 23

- パラメータを変更し評価
  - 提案手法ではFPが抑えられ precision や Specificity が高いスコアを記録

仮想モダリティのパラメータ | 評価の閾値 | 統合のパラメータ

表: 定量的評価 (Table. 2)

p	μt	σt	θ	pop.op	W	Acc (Ours)	Prec (Ours)	Rec (Ours)	Spec (Ours)	Acc (Avg)	Prec (Avg)	Rec (Avg)	Spec (Avg)	Acc (Max)	Prec (Max)	Rec (Max)	Spec (Max)
1	hhhh	mmmm	0.75	5	0.05	0.94	0.87	0.95	0.92	0.86	0.86	0.49	0.47	0.04			
1	hhhh	rrrr	0.75	5	0.05	0.94	0.94	0.82	0.78	0.72	0.99	0.75	0.76	1.00			
1	hhhh	hhrr	0.75	5	0.05	0.96	1	0.91	0.86	0.93	0.99	0.82	0.75	0.87	1	0.95	
1	hhhh	hhrr	0.75	5	0.05	0.79	1	0.71	0.8	0.72	0.97	0.98	1	0.95			
1	hh	hhrr	0.75	5	0.05	0.91	1	0.82	0.95	0.94	0.99	0.94	0.94	0.9	1	0.99	
1	hh	hhrr	0.75	5	0.05	0.93	1	0.85	0.89	1	0.76	1	0.94	0.89	1	0.89	
1	hh	hhrr	0.75	5	0.05	0.92	1	0.83	1	0.54	0	0	0.94	0.89	1	0.89	
2	hhhh	hhrr	0.75	5	0.05	0.93	1	0.85	0.93	0.88	0.97	0.88	0.82	0.63	1	0.9	
2	hhhh	hhrr	0.75	5	0.05	0.93	1	0.85	0.93	0.88	1	0.89	0.91	0.83	1	0.82	
2	hh	hhrr	0.75	5	0.05	0.91	1	0.81	0.94	1	0.63	1	0.94	0.89	1	0.89	
2	hh	hhrr	0.75	5	0.05	0.93	1	0.85	0.88	0.73	1	0.94	0.89	1	0.89		
2	hh	hhrr	0.85	2	0.05	0.87	1	0.73	0.94	0.98	0.9	0.98	0.96	0.94	1	0.94	
2	hh	rrrr	0.65	6	0.05	0.86	1	0.81	0.76	1	0.73	0.94	0.87	0.96	1	0.91	
2	hh	mmmm	0.65	6	0.05	0.93	1	0.85	0.97	0.93	1	0.94	0.93	0.87	1	0.87	
1	hhhh	mmmm	0.75	5	0.05	0.82	0.94	0.58	0.97	0.92	0.86	1	0.86	0.49	0.47	1	0.04
1	hhhh	rrrr	0.75	5	0.05	0.48	0.91	0.48	0.86	0.79	1	0.72	0.69	0.75	0.76	1	0.04
1	hh	hhrr	0.75	5	0.05	0.84	1	0.68	0.94	0.94	0.99	0.94	0.94	0.9	1	0.89	
2	hh	hhrr	0.75	5	0.05	0.8	1	0.58	0.94	0.89	1	0.86	0.91	0.83	1	0.83	
2	hh	mmmm	0.75	5	0.05	0.8	1	0.58	0.54	1	0.03	1	0.94	0.89	1	0.89	
2	hh	hhrr	0.75	5	0.05	0.8	1	0.58	0.54	0	0	1	0.94	0.89	1	0.89	

Acc = (TP + TN) / (TP + TN + FP + FN)  
Prec = TP / (TP + FP)  
Rec = TP / (TP + FN)  
Spec = TN / (TN + FP)

23