

1

Proposal and Evaluation of Multi-Object Recognition Method Inspired by Multimodal Information Processing in the Human Brain

— — — — —

関 良我
大阪大学 大学院情報科学研究科 村田研究室
修士論文発表資料

1

2

- 1 研究背景と目的
- 2 アプローチ
- 3 手法の評価方式
- 4 まとめと今後の課題

2

1 研究背景と目的

- 2 アプローチ
- 3 手法の評価方式
- 4 まとめと今後の課題

3

3

研究背景と目的

- デジタルツインへの期待
 - 現実世界のあらゆる物体をリアルタイムに仮想空間上にマッピングすることで環境を再現
 - 仮想空間でのシミュレート結果を現実世界へフィードバック
 - 倉庫管理、安全管理、自動操縦など様々な場面で活躍が期待される
- ロバストなセンシングデータ処理が必要
 - 画像分析などによる物体とその状態の識別とセンサー機器自体の自己位置情報が必要
 - 従来は物体推定手法ではクラウド上に集約するため広域でリアルタイムなマッピングは困難

変化し続ける現実世界の環境に対応するため、端末単位でのロバストな処理が必要

4

研究背景と目的 - エッジクラウド型の確率的デジタルツイン

4

- AR クラウドを利用したエッジクラウド型のデジタルツイン構想
 - 各端末で「この物体は〇〇であるらしい」という周辺環境認識を行う
 - AR クラウド上で各端末の認識結果を統合し、巨大な仮想世界を構築
 - 統合された情報を端末にフィードバック
- 本論文ではクラウド処理の軽量化のため、各端末で周辺環境の認識を行いその結果を共有
 - 小さな端末での処理 → **軽量性**
 - 1 端末から得られる不安定な情報の処理 → **ロバスト性**

カメラの手ブレ/光の加減や角度など

5

6

- 1 研究背景と目的
- 2 アプローチ
- 3 手法の評価方式
- 4 まとめと今後の課題

6

- 1 研究背景と目的
- 2 アプローチ**
- 3 手法の評価方式
- 4 まとめと今後の課題

7

アプローチ

目的 **端末単位で処理可能な軽微性と現実世界の環境変化に適応可能なロバスト性を持った物体推定手法の作成**

アプローチ

- 脳の良い認知機構に着目した物体推定手法
 - 目や耳、三半規管などから得られた不確実な情報のマルチモーダル統合処理
 - 階層構造の**軽微かつロバスト**な意思決定
- 脳の認知機能の数理モデルを利用
 - 脳や生体の「ゆらぎ学習」によるユニモーダル処理
 - Bayesian Attractor Model (BAM)
 - 脳の知覚による因果推論に基づいたマルチモーダル処理
 - Bayesian Causal Inference (BCI)
- 複数のオブジェクトの認識結果を統合
 - 条件付き確率場 (CRF) を利用

8

アプローチ - BCIによるBAMの拡張

- カメラから得られた物体の三次元位置モダリティと映像モダリティの2つの入力刺激 $x_{Location}$, x_{Visual} を用いて、それぞれ個別に認知結果 S_L , S_V を計算
 - 本研究ではBAMによる認知をするよう設計
- 両モダリティの入力刺激を統合した認知結果 S_{LV} を計算
- 両モダリティで同じものを認知しているか別のものを認知しているかを確率的に計算する (Causal inference)
- その割合に応じた認知結果を出力 (Model Average)

9

アプローチ - マルチオブジェクト認識

- 観測データは安定しているに限らないので、複数物体を認識してより安定した結果を出力する
 - BCIの出力結果を入力として作成されたノードを持つグラフを作成
 - グラフは条件付き確率場 CRF で各ノードの状態と対応する認識結果を物体のラベルとして出力
 - この対応関係が最も良いベアとなるように CRF を解く
- 入力 I に対してギブスエネルギー $E(x)$ を最小化する x を探す → 確率伝播法を用いる
 - $E(x|I) = \sum_i \psi_u(x_i|I_i) + \sum_{i < j} \psi_p(x_i, x_j)$

10

- 1 研究背景と目的
- 2 アプローチ
- 3 手法の評価方式
- 4 まとめと今後の課題

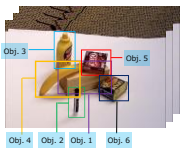
11

- 1 研究背景と目的
- 2 アプローチ
- 3 手法の評価方式**
- 4 まとめと今後の課題

12

手法の評価方式

- 公開データセットの映像データを元に映像モダリティ / 位置モダリティの特徴量を抽出
 - 既存手法 SiameseRPN を用いて bounding box を切り出す
 - 各 box から映像に関する特徴量を圧縮・抽出
- BAM のみでのユニモーダル精度評価
 - 入力は連続する各フレームの映像特徴量
 - BAM は初めての 1 枚目の特徴量のみをアトラクターに記憶
- BCI の結果からマルチモーダル評価
- 制約を加えたマルチオブジェクトでの評価
 - 本文献では「同じ物体は存在しない」という単純な制約のみを課した

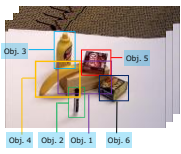


13

手法の評価方式 - 評価結果

Modality	Obj. 1	Obj. 2	Obj. 3	Obj. 4	Obj. 5	Obj. 6	Total
Unimodal video	95.3	99.4	95.9	91.7	96.9	78.7	96.3
Unimodal location	39.8	92.8	67.8	92.8	74.6	66.6	72.2
Multimodal video	75.2	99.4	96.5	94.8	96.9	83.0	91.0
Multimodal location	39.8	92.7	67.9	92.2	74.8	66.4	72.3
CRF version	94.5	83.8	83.2	92.4	86.6	86.8	87.8

※ bounding box のオブジェクト認識の正答率 (%)



- Unimodal は BAM のみでの正答率
- Multimodal は BCI を行った結果の正答率
- CRF version はマルチオブジェクト認識の正答率
- Total は各 bounding box の正答率の平均
- カメラ角度や手ブレなど観測データの不確実性に左右されない認識結果
 - 途中で見切れるなどしても、他の良い観測データを得られているオブジェクトからの推論結果を元に認識精度を向上
- 元々観測の精度の良いオブジェクトは、悪い方に引っ張られてしまう
 - 単純な平均ではなく、悪い影響を抑えられている

14

- 研究背景と目的
- アプローチ
- 手法の評価方式
- まとめと今後の課題

15

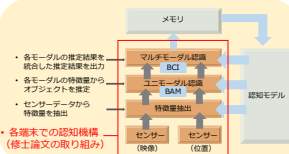
- 研究背景と目的
- アプローチ
- 手法の評価方式
- まとめと今後の課題

16

まとめと今後の課題

これまでの達成状況

- 複数モダリティの観測情報を元に物体推定を行う手法を提案
- 端末レベルで推定を行うことが可能な**軽量性**、複数モダリティを組み合わせることで**ロバスト性**を獲得
 - 実行速度: 1.18 ms / frame (CPU: Core-i7 8700, RAM: 16.0 GB)
 - ロバスト性: 観測情報がうまく得られなかった場合にも高い精度での推定を行うことができた
- 得られた観測データの良し悪しに大きく差がある場合でも悪い方にあまり強く影響を受けない



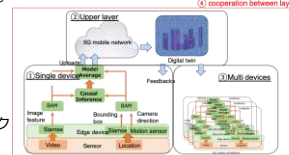
各端末での認知機構 (修士論文の取り組み)

17

まとめと今後の課題

今後の課題

- ① 端末レベルでの認識
 - より複雑な制約を課し、同様のパフォーマンスが発揮できるか検証
 - 現在の処理速度のボトルネックは BAM 部分であるが、制約が複雑になると CRF 部分のオーダーが増加すると考えられる
- ② 上位層に端末から得られた認知を蓄積
- ③ 複数の端末からの認知を統合する
- ④ 蓄積された認知を端末にフィードバック



① cooperation between layers

18