

2023 年度 卒業論文



知識選択型転移強化学習を用いた自律型移動ロボット における障害物回避

Obstacle Avoidance in Autonomous Mobile Robots Using
Knowledge-Selective Transfer Reinforcement Learning

指導教員 准教授 河野 仁

東京電機大学 工学部 情報通信工学科

学籍番号 20EC060

須賀 哉斗

目次

第 1 章	序論	1
1.1	背景	2
1.2	自動運転の事故事例	3
1.2.1	自動運転車による死亡事故	3
1.2.2	自動運転レベル 4 の運行を行っていた車両の接触事故	3
1.3	関連研究	5
1.3.1	動的障害物回避に注目した電動四輪車の知的自動運転システム	5
1.3.2	強化学習による仮想環境と実環境における自動走行車いすの障害物回避	5
1.4	本研究の目的	6
1.5	本論文における研究 3 要素	7
1.5.1	研究の学術性	7
1.5.2	研究の新規性	7
1.5.3	研究の有用性	7
1.6	本論文の構成	8
第 2 章	学習アルゴリズム	9
2.1	はじめに	10
2.2	学習アルゴリズム	11
2.2.1	強化学習	11
2.2.2	転移学習	11
2.2.3	行動選択	12
2.2.4	知識選択	12
2.3	物理演算シミュレーション	13
2.4	おわりに	14

第 3 章	提案手法	15
3.1	はじめに	16
3.2	提案手法	17
3.2.1	知識選択手法 (SAP-net)	17
3.2.2	シミュレーションによる実験	17
3.2.3	実機による実験	17
3.3	おわりに	17
第 4 章	シミュレータ実験	19
4.1	はじめに	20
4.2	実験環境	21
4.3	知識選択型転移強化学習	22
4.3.1	目的と実験条件	22
4.3.2	実験結果	22
4.4	実機実装	22
4.4.1	目的と実験条件	22
4.4.2	実験結果	22
4.5	おわりに	23
第 5 章	結論	25
5.1	結論	26
5.2	今後の展望	27
	謝辞	30
	参考文献	31
	研究業績	36

図目次

1.1	自動運転のレベル分け [ROHM 2020]	2
1.2	事故を起こした Uber Technologies の自動運転車 [米国家運輸安全委員会 2019]	3
1.3	自動運転レベル4の運行を行っていた車両 [ソリトンシステムズ 2023]	4
1.4	電動四輪車の知的自動運転システム [2005]	5
2.1	SAP-net のイメージ図	13

表目次

第 1 章

序論

Contents

1.1	背景	2
1.2	自動運転の事故事例	3
	1.2.1 自動運転車による死亡事故	3
	1.2.2 自動運転レベル 4 の運行を行っていた車両の接触事故	3
1.3	関連研究	5
	1.3.1 動的障害物回避に注目した電動四輪車の知的自動運転システム ...	5
	1.3.2 強化学習による仮想環境と実環境における自動走行車いすの障 害物回避	5
1.4	本研究の目的	6
1.5	本論文における研究 3 要素	7
	1.5.1 研究の学術性	7
	1.5.2 研究の新規性	7
	1.5.3 研究の有用性	7
1.6	本論文の構成	8

1.1 背景

自動車業界は AI や IoT のような最先端技術によって「100 年に一度の大変革の時代」に突入しているとされ、その中心には自動運転技術がある。特に、2023 年 4 月 1 日には日本で自動運転レベル 4 の公道走行が解禁された。これは限定された走行領域でシステムが全ての運転操作を実施し、ドライバーが運転席を離れることができる段階である。自動運転のレベル分けを Fig. 1.1 示す。2025 年を目途に完全自動運転、レベル 5 の実現が目指されているが、そのためには周囲の物体を正確に識別し、適切な判断を下す能力が不可欠である。現在は機械学習技術を用いて状況判断や障害物検知などを行うことで自動運転技術をじつげんしているが、障害物回避技術に関しては完全な解決には至っていない。現時点では、この課題への対応として、単眼カメラと三次元地図を用いて動的な障害物を三次元的に検出し、その形状を復元するための手法や、動的な障害物を回避するための軌道生成法が研究されているが、いずれも実機を用いた障害物の回避は実現されていない [1][2]。このため、完全自動運転を実現するためには、環境認知と行動計画を人工知能に任せると共に、さらに人間の賢さに似た知的能力が必要とされている [3]。障害物回避を含め、様々な配置パターン環境や場面に対応する手法の 1 つとして転移強化学習が注目されている。中でも知識選択型転移強化学習を用いて開発した Kono らの SAP-net に注目する [3]SAP-net を用いて物理演算シミュレータで走行経路の強化学習をした後、学習結果をシニアカーに転移させて自律走行の実現に成功している [4]。本研究では、知識選択型転移強化学習を用いた移動ロボットにおける障害物回避の実現を提案する。具体的には

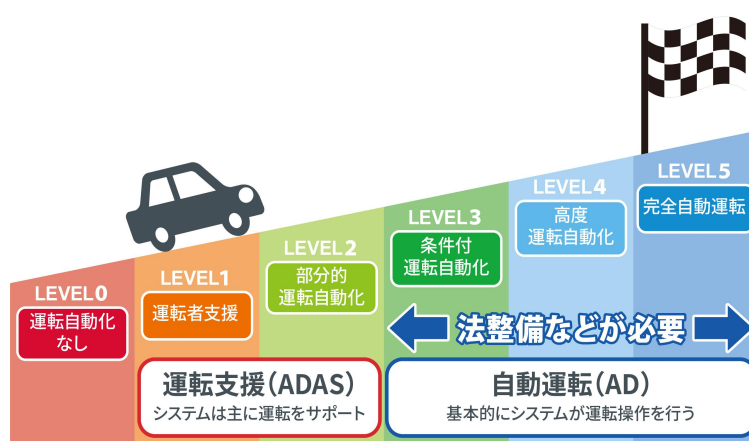


Fig. 1.1: 自動運転のレベル分け [ROHM 2020]

1.2 自動運転の事故事例

1.2.1 自動運転車による死亡事故

2018 年 3 月，米ライドシェア大手の Uber Technologies（ウーバー・テクノロジーズ）の自動運転車がアリゾナ州を走行中，自転車を押しながら道路を横断していた歩行者をはね，死亡させる事故を起こした．事故を起こした車両を Fig1.2 に示す．当時の米の自動運転のレベルは 3 である．自動運転レベル 3 は，限定された条件下において，システムが全ての運転操作を実施する．ただし運転自動化システム作動中であっても，システムからの要請があればドライバーはいつでも運転に戻れる状態である必要があるレベルである．事故の主要因として，同乗していたセーフティドライバーが事故発生時に携帯端末で動画を視聴していたことが挙げられる．また，ウーバーの自動運転システムが歩行者を検知できず，「自動車」や「自転車」，「その他のオブジェクト」などと認識していたことも発覚している．その挙動についても「左側の車線を走行」「静止中」などと判断したため，ギリギリまで衝突の危険性を検知できなかった．



Fig. 1.2: 事故を起こした Uber Technologies の自動運転車 [米国家運輸安全委員会 2019]

1.2.2 自動運転レベル 4 の運行を行っていた車両の接触事故

2023 年 10 月 29 日には自動運転レベル 4 の運行を行っていた車両が自転車に接触するという事故が発生した．自動運転レベル 4 は，限定された条件下において，システムが全ての運転操作を実施する・ドライバーが運転席を離れることができるレベルである．その車両に

は障害物を避けるためのセンサーやレーダーが車両に取り付けられていた. 事故の最大の原因は学習データ不足により自動ブレーキが作動せず, 自転車を十分に認識できなかったことである.



Fig. 1.3: 自動運転レベル4の運行を行っていた車両 [ソリトンシステムズ 2023]

1.3 関連研究

1.3.1 動的障害物回避に注目した電動四輪車の知的自動運転システム

これまで運転技術・知識の不十分な利用者にも，安心して利用できるために，電動四輪車を用いて人間の運転知識をコントローラに組み込んだ自動運転システムが研究されてきた [中川 2005]. その電動四輪車を Fig1.4 に示す. この電動四輪車の自動運転システムを基に，予測的ファジー制御を運転アルゴリズムに組み込むことで，実際の環境（例えば歩道）でのシステムの有効性を示すシミュレーション実験を行った. しかし，システムを開発しただけで，実機への実装はまだ出来ていない.



Fig. 1.4: 電動四輪車の知的自動運転システム [2005]

1.3.2 強化学習による仮想環境と実環境における自動走行車いすの障害物回避

Unity 3D を用いたシミュレーション環境と WHILL 社の電動車いすを活用し，DDPG を採用した強化学習によって，動的な環境下での自動走行車いすの障害物回避技術を開発し，実機での実験を通してその有効性を検証した. しかし，ルールベースを用いての障害物回避を行ったため，

1.4 本研究の目的

1.1 節より自動運転システムは日々進化しているが、1.2 節のような自動運転にまつわる事故が多発している。このことから、自動運転システムに搭載された環境認知をするためのセンサーやレーダーに頼るだけでは現時点では限界があることがわかる。

知識選択型転移強化学習 (SAP-net) を活用した移動ロボットにおける障害物回避の実現

1.5 本論文における研究 3 要素

1.5.1 研究の学術性

本研究の学術性は、LIDAR を用いて様々な静的障害物の配置パターンを回避する強化学習を行い、強化学習で得られた学習データを知識（行動価値関数）として保存し、

1.5.2 研究の新規性

本研究の新規性は、

1.5.3 研究の有用性

研究の有用性は、自動運転車両やその他の自律型移動ロボットの安全性と効率性を向上させることにある。特に、交通事故のリスクを減らし、人々の安全を確保することに大きな貢献をすると期待されている。またこの技術は、都市計画や公共交通システムの最適化にも応用可能であり、広範な影響を与える事が可能となる。

1.6 本論文の構成

第 2 章

学習アルゴリズム

Contents

2.1	はじめに	10
2.2	学習アルゴリズム	11
	2.2.1 強化学習	11
	2.2.2 転移学習	11
	2.2.3 行動選択	12
	2.2.4 知識選択	12
2.3	物理演算シミュレーション	13
2.4	おわりに	14

2.1 はじめに

2.2 学習アルゴリズム

本研究では、移動ロボットにおける動的障害物の回避のため強化学習、転移強化学習、知識選択を用いる。それぞれ式を示しながら述べていく。

2.2.1 強化学習

タスク達成を目指し、繰り返し最適解を試行錯誤的に学習する手法のことである。試行錯誤的に学習していくため、知識が全くない状態からの知識の取得が可能である。しかし、試行錯誤的に学習するため、学習が遅いという欠点がある [Sutton1998]。以下に強化学習に用いられる Q 学習の式を示す。

$$Q(s_t, a) \leftarrow Q(s_t, a) + \alpha \left\{ r + \gamma \max_{a' \in A} Q(s_{t+1}, a') - Q(s_t, a) \right\} \quad (2.1)$$

ここで、 $Q(s_t, a)$ は時刻 t における状態 s から行動 a を選択したときの報酬の期待値を表す行動価値関数を示している。学習率 α ($0 < \alpha \leq 1$) は、更新される価値の大きさおよび学習の速度に影響を与え、高い値では学習が速くなるものの、最適解を見つける確率が低下する可能性がある。割引率 γ ($0 < \gamma \leq 1$) は、将来の報酬の現在価値を計算する際に用いられ、 γ の値が小さい場合は未来の価値を低く評価し、大きい場合は高く評価する。状態 s_{t+1} での最大行動価値を求めるために $\max_{a' \in A} Q(s_{t+1}, a')$ を用い、この値を γ で割り、現在の状態 $Q(s_t)$ の価値として加算し、さらに報酬 r を加え、 α で割引いた値を $Q(s, a)$ に加えることで、行動価値を更新していく。

2.2.2 転移学習

強化学習である程度学習した知識を別のタスクに適用させる手法である。知識を再利用して学習するので学習の短縮や精度を向上させることが可能である [Taylor2009]。以下に転移学習の式を示す。

$$Q_c(s, a) = Q_c(s, a) + \tau Q_s(s, a) \quad (2.2)$$

$Q_s(s, a)$ は Source task から転移された行動価値関数を示している。 $Q_s(s, a)$ に転移率 τ ($0 < \tau \leq 1$) を掛けることで、新たな環境が再利用される行動価値関数を獲得した環境と異なる場合においても適応できるようになる。 $Q_t(s, a)$ は転移先のタスクで更新する行動価値関数を示している。さらに新たな環境とで学習した行動価値関数も $Q_t(s, a)$ に更新していく。 $Q(s_c, a)$ は転移された行動価値関数と Target Task で獲得した行動価値関数を統合した行動価値関数であり、Target Task で行われる行動選択は $Q(s_c, a)$ を用いて行われる。

2.2.3 行動選択

Q テーブルにはある状態 s で取り得る行動 a とそれに対応する価値が記録されている。これに基づき、どの行動を取るかを決定するために行動選択関数を用いる。アルゴリズムが存在する [河野 2022]。本研究では、行動選択関数ボルツマン選択を用いる。複雑で不確実性の高い環境や、エージェントが幅広い行動から学習による必要がある場合に有効である。ボルツマン選択を用いることで、エージェントが受け取る報酬を行動選択に反映させることができる。以下にボルツマン選択の式を示す。

$$p(a|s) = \frac{\exp(\frac{Q(s,a)}{T})}{\sum_{b \in A} \exp(\frac{Q(s,b)}{T})} \quad (2.3)$$

$p(a|s)$ は状態 s において行動選択 a を選択する確率で、 $Q(s,a)$ は行動 a を選択したときの価値、 T は温度定数でボルツマン選択におけるランダム生成を調整するパラメータである。 T が高いほど選択はランダムに近くなり、 T が低いほど最も価値の高い行動が選択されやすくなる。

2.2.4 知識選択

人の脳内における概念の選択手法と言われている活性化拡散モデルを元にした SAP-net (Spreading Activation Policy-network) を用いた知識選択型の転移学習手法を使用し、転移強化学習を行う。使用する知識、いわゆる方策は強化学習で獲得した学習結果を指し、それらを選択することで行動を決定する。SAP-net のイメージ図を Fig.2.1 に示す。

この手法では、強化学習で得られた方策や行動価値関数をグラフ構造で管理し、外部刺激に反応して関連するポリシーの活性値が増加するメカニズムを持っている。この活性値の増加は、グラフ内の他のノードへ伝播し、一定の閾値を超えた方策や行動価値関数が「想起」されて転移学習に利用される。また、SAP-net には時間経過に伴う活性値の減少を考慮する設計が取り入れられており、効率的な情報処理を可能にしている。この手法を式で表したのが活性化拡散方程式である。以下に活性化拡散方程式を示す。

$$A_{j,t+\Delta t} = A_{j,t} + a_\delta + \sum_{k=1}^q \eta_k - D_k \quad (2.4)$$

ある時点 t での方策が持つ活性値 $A_{j,t}$ を外部からの入力や内部メカニズムによって加算される a_δ 、周囲の状況や他の方策から拡散されてくる値の総和 $\sum_{k=1}^q \eta_k$ を加えて、さらに時間の経過に伴う自然な減少や外部からの抑制的な影響を反映する減衰定数 D_k を引くことで、時間 $t + \Delta t$ で活性値 $A_{j,t+\Delta t}$ を更新していく。

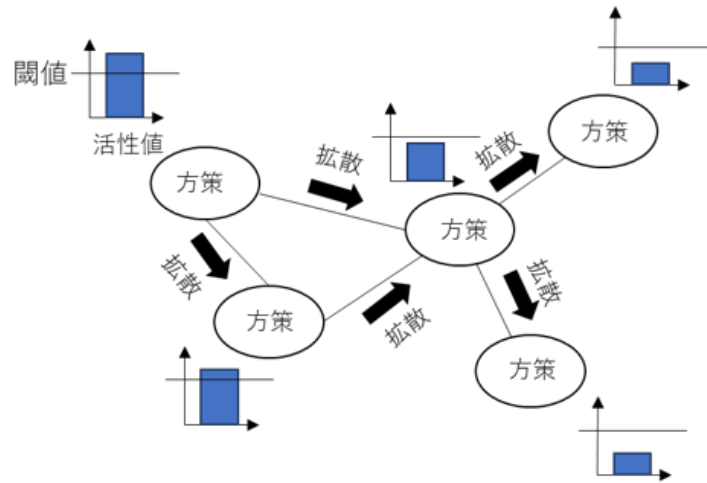


Fig. 2.1: SAP-net のイメージ図

2.3 物理演算シミュレーション

本研究では強化学習，転移学習，知識選択手法を用いる．強化学習は試行錯誤的に行動を何千と繰り返して学習するため，実環境でやると時間がかかってしまう．また実機で使用するロボットが破損する恐れがある．そこで Webots という物理演算シミュレーションを用いる [Webots1998]. webots の操作画面を Fig に示す．Webots を使用することで，シミュレーション内のロボットが強化学習を行っていくため，実環境で行うよりも安全に効率よく学習することが可能である．シミュレーション環境を Fig に示す．

2.4 おわりに

第 3 章

提案手法

Contents

3.1	はじめに	16
3.2	提案手法	17
3.2.1	知識選択手法 (SAP-net)	17
3.2.2	シミュレーションによる実験	17
3.2.3	実機による実験	17
3.3	おわりに	17

3.1 はじめに

3.2 提案手法

本研究では、移動ロボットの動的障害物回避能力を向上させることを目的とする。この目的を達成するために、Q 学習を基礎とする強化学習及び、知識選択型転移強化学習メカニズムである SAP-net を活用する。まず Webots という物理演算シミュレータを使用して、複数の障害物配置が含まれる環境下で Q 学習による強化学習実験を行い、SAP-net を利用して強化学習で取得した行動価値関数が新たな環境に迅速に適用される能力を検証する。次に実機を使用して、シミュレータで開発されたモデルが現実世界にどう適応するか、そして SAP-net が活性化する知識を元にした実際の障害物回避の性能を評価する。

3.2.1 知識選択手法 (SAP-net)

SAP-net はロボットの初期座標を基準に障害物までの角度と距離を保存させる。この情報はロボットが環境を理解し、障害物を正確に把握するための基盤となる。次に強化学習から得た行動価値関数の情報を取得し、それらの類似度を計算する。類似度計算は異なる状況における最適な行動の一貫性を確認する重要なステップである。そして計算された類似度情報を基にネットワークを構築する。このネットワークは活性化拡散モデルとして機能し、異なる行動価値関数を統合し、環境への柔軟な対応を可能にする。そして構築されたネットワークは知識を選択し、動的な状況において障害物回避の戦略を提供する。これによりロボットは瞬時の判断により、適切な知識を活用して効果的な障害物回避を実現する。

3.2.2 シミュレーションによる実験

Webots を使う 二輪ロボットを作成 障害物の回避の強化学習をやったのちに知識選択手法を用いて転移強化学習を行う

3.2.3 実機による実験

使用する実機はラズベリーパイマウスを使用する。ラズベリーパイマウスに障害物認識のための RPLIDAR を接続する。

3.3 おわりに

第 4 章

シミュレータ実験

Contents

4.1	はじめに	20
4.2	実験環境	21
4.3	知識選択型転移強化学習	22
	4.3.1 目的と実験条件	22
	4.3.2 実験結果	22
4.4	実機実装	22
	4.4.1 目的と実験条件	22
	4.4.2 実験結果	22
4.5	おわりに	23

4.1 はじめに

本章では，提案したアプローチの有用性を示すために行った．

4.2 実験環境

実験環境を図で説明

4.3 知識選択型転移強化学習

4.3.1 目的と実験条件

目的：転移強化学習が障害物回避に有用であるか
実験条件を書く

4.3.2 実験結果

4.4 実機実装

4.4.1 目的と実験条件

目的：転移強化学習で得られた学習データを実機に転移させ，障害物回避が可能であるか
二輪ロボットと同じような環境で行う

4.4.2 実験結果

4.5 おわりに

第 5 章

結論

Contents

5.1	結論	26
5.2	今後の展望	27

5.1 結論

本研究では従来の自動運転システムに知識選択手法（SAP-net）を組み込むことで、動的障害物の回避の性能向上を目的とした。第 1 章では、本研究の背景第 2 章では、第 3 章では、第 4 章では、第 5 章では、

5.2 今後の展望

謝辭

本論文を締めくくるにあたり、ご指導、ご協力をいただいた全ての方々に、深く感謝いたします。

本研究の指導教員である東京電機大学知能情報システム研究室准教授河野仁先生には、有意義な研究の機会を与えていただくとともに、終始熱心なご指導と適切な助言を頂きました。卒業研究の内容の相談や、専門的なアドバイスをいただきこの論文を執筆するのに大きく影響を受けました、この経験は非常に有意義で、今後の人生に大きく役立つと確信しています。ここに深く感謝いたします。

最後に、私の大学での学びを経済的、精神的に支えてくれた家族、そして友人の方々に深く感謝いたします。本当にありがとうございました。

令和6年2月 須賀哉斗

参考文献

<和文文献>

[浅間 2014]

浅間 一: “災害時に活用可能なロボット技術の研究開発と運用システムの構築,” 日本ロボット学会誌, vol. 32, no. 1, pp. 37–41, 2014.

[田所 2012]

田所 諭: “災害対応ロボットのあるべき姿,” 建設の施工企画, vol. 753, pp. 23–31, 2012.

[敷島 2021]

敷島 惇也, 田崎豪: “単眼カメラと三次元地図を用いた動的障害物の検出と三次元復元,” 計測自動制御学会論文集, vol. 57, no. 1, pp. 37–46, 2021.

[金原 2022]

金原 翔, 米谷 昭彦: “自律走行車の軌道生成における不確定な動的障害物への対処方法,” 2022 年第 65 回自動制御連合講演会, pp. 462–467, 2022.

[河野 2022]

河野 仁, 坂本 裕都, 温 文, 藤井 浩光, 池 勇勲, 鈴木 剛: “知識選択型転移強化学習を用いたシニアカーの自律運転,” 2022 年電気学会電子・情報・システム部門大会, no. 1151, pp. 714–718, 2022.

[齋藤 2014]

齋藤 碧, 小林 一郎: “強化学習における効率的な転移学習適用に関する一考察,” 2014 年度第 28 回人工知能学生全国大会, no. 3, pp. 1–3, 2014.

[石倉 2016]

石倉 裕貴, 岸本 良一, 堀内 匡: “CPG と強化学習を用いた多脚ロボットの目標到達行動の獲得,” 電気学会論文誌 C, vol. 136, no. 3, pp. 333–339, 2016.

[Sutton 1998]

R. S. Sutton, A. Gbarto (三上貞芳, 皆川雅章訳) : 強化学習, 森北出版, 2000.

[木村 1999]

木村 元, 宮崎 和光, 小林 重信: “強化学習システムの設計指針,” 計測と制御, vol. 38, no. 10, pp. 1–6, 1999.

[広瀬 1998]

広瀬 茂男, 塚越 秀行, 米田 完: “不整地における歩行機械の静的安定性評価基準,” 日本ロボット学会誌, vol. 16, no. 8, pp. 1076–1082, 1998.

[米田 1996]

米田 完, 広瀬 茂男: “歩行機械の転倒安定性,” 日本ロボット学会誌, vol. 14, no. 4, pp. 517–522, 1996.

[中岡 2013]

中岡 慎一郎: “拡張可能なロボット用統合 GUI 環境 Choreonoid,” 日本ロボット学会誌, vol. 31, no. 3, pp. 12–17, 2013.

[中村 2016]

中村 晋也, 吉灘 裕, 倉鋪 圭太, 谷本 貴頌, 近藤 大祐: “複合ロボットのための動力学シミュレータの開発,” 日本機械学会ロボティクス・メカトロニクス講演会’16 講演論文集, 2016.

[内閣府 2016]

内閣府: “平成 28 年版防災白書,” http://www.bousai.go.jp/kaigirep/hakusho/h28/honbun/3b_6s_10_00.html, 2016, 閲覧日 2016.12.15.

[気象庁 2016]

気象庁: “日本付近で発生した主な被害地震（平成 8 年以降）,” <http://www.data.jma.go.jp/svd/eqev/data/higai/higai1996-new.html#higai2016>, 2016, 閲覧日 2016.12.15.

[原木 2013]

原木俊彦, 大川一也, 加藤秀雄, 樋口静一: “軽微な故障を抱えた脚型ロボットにおけるゴール

到達のための自律動作修正,” 日本機械学会ロボティクス・メカトロニクス講演会’13 講演論文集, 2013.

[防衛省 自衛隊 2016]

防衛省・自衛隊: “平成 28 年熊本地震に係る災害派遣について（最終報）,” <http://www.mod.go.jp/j/press/news/2016/05/30b.html>, 2016, 閲覧日 2016.12.16.

[米国家運輸安全委員会 2019]

米国家運輸安全委員会: “事故を起こした Uber Technologies の自動運転車,” <https://japan.cnet.com/article/35145765/>, 2019, 閲覧日 2023.12.20.

[ROHM 2020]

ROHM SEMICONDUCTOR: “ADAS. 自動運転, 安全を守る自動車のテクノロジーを解説,” <https://www.rohm.co.jp/blog/-/blog/id/8030502>, 2016, 閲覧日 2023.12.15.

<英文文献>

[Matsuno 2004]

F. Matsuno and S. Tadokoro: “Rescue Robots and Systems in Japan,” *Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics*, pp. 12–20, 2004.

[Murphy 2004]

R. R. Murphy: “Trial by Fire [Rescue Robots],” *IEEE Robotics & Automation Magazine*, vol. 11, no. 3, pp. 50–61, 2004.

[Carlson 2005]

J. Carlson and R. R. Murphy: “How UGVs Physically Fail in the Field,” *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 423–437, 2005.

[Messuri 1985]

D. A. Messuri and C. A. Klein: “Automatic Body Regulation for Maintaining Stability of a Legged Vehicle during Rough-terrain Locomotion,” *IEEE Journal on Robotics and Automation*, vol. 1, no. 3, pp. 132–141, 1985.

[Platt 1991]

J. Platt: “A Resource-Allocating Network for Function Interpolation,” *Neural Computation*, vol. 3, no. 2, pp. 213–225, 1991.

[Nagatani 2011]

K. Nagatani, S. Kiribayashi, Y. Okada, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi and Y. Hada: “Redesign of Rescue Mobile Robot Quince,” *Proceeding of 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pp. 13–18, 2011.

[Kawatsuma 2012]

S. Kawatsuma, M. Fukushima and T. Okada: “Emergency Response by Robots to Fukushima-Daiichi Accident: Summary and Lessons Learned,” *Industrial Robot: An International Journal*, vol. 39, no. 5, pp. 428–435, 2012.

[Kober 2013]

J. Kober, B. J. Andrew and Jan Peters: “Reinforcement Learning in Robotics: A survey,” *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[Haykin 2009]

S. Haykin: *Neural Networks and Learning Machines*, Pearson Upper Saddle River, 2009.

[Mnih 2013]

V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller: “Playing Atari with Deep Reinforcement Learning,” *Proceedings of NIPS 2013 Deep Learning Workshop*, 2013.

研究業績

査読有り国内会議

1. 伊藤 翼, 河野 仁, 田村 雄介, 山下 淳, 浅間 一: “アーム搭載移動ロボットの駆動系故障時のための強化学習を用いたリカバリモーション獲得,” 第 22 回ロボティクスシンポジウム予稿集, 2017, 発表予定.

査読有り国際会議

1. **Tasuku Ito**, Hitoshi Kono, Yusuke Tamura, Atsushi Yamashita, and Hajime Asama: “Recovery Motion Learning for Arm Mounted Mobile Crawler Robot in Drive System’s Failure,” *The 20th World Congress of the International Federation of Automatic Control*, 2017, 査読中.