

論文審査の要旨および学識確認結果

報告番号	甲 第 号	氏 名	細見 直希
論文審査担当者：	主査	慶應義塾大学教授	博士（情報学） 杉浦 孔明
	副査	慶應義塾大学教授	博士（工学） 青木 義満
		慶應義塾大学教授	博士（工学） 今井 倫太
		慶應義塾大学教授	工学博士 斎藤 博昭
(論文審査の要旨) 学士（工学）、修士（工学）細見直希君提出の学位請求論文は「マルチモーダル情報に基づく移動指示理解モデルの構築」と題し、全6章からなる。 本論文は、人々の自由で便利な移動を支援するエージェント（モビリティ）の基盤となる、視覚と言語を用いたマルチモーダル情報に基づく移動指示理解モデルの構築に取り組んだものである。具体的には、人間が自律移動可能なモビリティに対して「黒いトラックの向かい側に停まってください」という移動指示を与えた場合に、モビリティが指示内容を適切に解釈し、黒いトラックを目印として、その向かい側の道路上における適切な目標位置や領域を予測可能なモデルの構築を目的としている。 本論文では、特に実世界の移動タスクに特有の問題に対処する、マルチモーダル情報に基づく目標位置または領域の予測に関する手法を提案している。具体的には、(1)画像と言語のような異なるモダリティ間で情報の性質に違いがあるうえに、情報の信頼性が環境条件によって変動し得る点、(2)目標位置や領域は道路上の任意の地点に存在する可能性があり、予測のための明確な視覚的境界が存在しない点、に対処する手法を提案し、実験を通じてその有効性を示している。 第1章では、研究の背景、目的、および論文構成を示している。 第2章では、言語理解および記号接地、視覚と言語を中心としたマルチモーダル情報の統合的処理に関する先行研究、およびモビリティ分野における応用事例を整理し、本研究の位置づけを述べている。 第3章では、本論文で用いるマルチモーダル情報の統合的処理の基盤技術について詳述し、提案手法の背景にある原理および理論的根拠を示している。 第4章では、移動指示文とRGB画像を入力として、モビリティの目標領域をセグメンテーションマスクとして予測する Referring Navigable Region タスクを対象とし、言語、RGB画像、およびRGB画像から生成したセマンティックセグメンテーションマスクの3つのモダリティを扱う Trimodal Navigable Region Segmentation Model (TNRSM)を提案している。実験を通じて、TNRSMがモダリティ間の情報の性質の違いを補完し、さらに、視覚に関する情報の信頼性が環境条件によって変動する問題に対して効果的に対処できることを明らかにしている。 第5章では、移動指示文とRGB画像を入力として、モビリティの目標位置を画像上の二次元の座標として予測する Vision-and-Language Target Positioning タスクを対象とし、移動指示で参照されるランドマークと目標位置の相対的位置関係を明示的にモデル化する Target Regressor in Positioning (TRiP)を提案している。実験を通じて、TRiPがランドマークと目標位置の相対的位置関係を考慮し、目標位置の予測において明確な視覚的境界が存在しない問題に効果的に対処できることを明らかにしている。 第6章では、本論文の総括を行い、今後の展望について述べている。 以上要するに、本論文は、視覚と言語を用いたマルチモーダル情報に基づく移動指示理解に関するモデルおよびその基盤技術を提案し、実験を通じてその有効性を明らかにしたものであり、工学上寄与するところが少なくない。 よって、本論文の著者は博士（工学）の学位を受ける資格があるものと認める。			
学識確認結果	学位請求論文を中心にして関連学術について上記審査会委員で試問を行い、当該学術に関し広く深い学識を有することを確認した。 また、語学（英語）についても十分な学力を有することを確認した。		