

Title	Recognizing Human Internal States through Deep Learning Approaches
Author(s)	時, 嘉祺
Citation	大阪大学, 2024, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/96118
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、 〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉 大阪大学の博士論文について 〈/a〉 をご参照ください。

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

論 文 内 容 の 要 旨

氏 名 (時 嘉 祺)	
論文題名	Recognizing Human Internal States through Deep Learning Approaches (ディープラーニング手法を用いた人間の内的状態の認識)
論文内容の要旨	
<p>For service-oriented robots, the capability to recognize and interpret human internal states, such as emotions and intentions, is paramount. Recognizing these states not only enhances their ability to effectively communicate but also allows them to provide tailored services closely aligned with human needs. This recognition ability is not exclusive to robots but extends to other AI systems and devices, such as autonomous vehicles and smart home systems. This dissertation addresses the critical capability of robots and AI systems to recognize human internal states, with a primary focus on emotions and intentions, as they directly affect human actions and might be inferable from observable behaviors even after short-term observation.</p> <p>The existing research on emotion recognition, while extensive, reveals certain shortcomings. These works have predominantly focused on facial expressions and vocal intonations, often overlooking the rich information conveyed through gestures. Furthermore, many existing methodologies neglect modeling the human body's spatial structure. Additionally, there is a scarcity of efficient multimodal fusion methods, which are vital for comprehensive emotion recognition. To address these gaps, this research introduced innovative deep learning models that leverage gesture-based modalities for emotion recognition. Utilizing the IEMOCAP database, a self-attention enhanced spatial-temporal graph convolutional network was developed that accurately models the skeletal structure and dynamic connections between body joints. This approach significantly outperformed existing models, highlighting the untapped potential of gesture-based emotion recognition. Furthermore, a graph attention-based fusion method was proposed, integrating audio, textual, and skeletal data, demonstrating the efficacy of comprehensive data fusion in emotion classification.</p> <p>On intention recognition, existing research is constrained by limited real-world scenarios and has not reached its full potential in terms of accuracy. To address these limitations, this research collected data on human behaviors observed at a building entrance, which include intentions related to the use of hand dispensers and temperature scanners. A spatial-temporal graph convolutional network was proposed for skeleton-based intention recognition, outperforming baseline models and surpassing human predictive abilities in our experiments. In the realm of intelligent transportation systems, a deep-learning model harnessing multimodal features was proposed for automatic pedestrian intention inference, improving accurate predictions essential for the safety and convenience of road users.</p> <p>In this research, new methods for recognizing human internal states, specifically emotions and intentions, using deep learning were proposed and verified for their effectiveness through experiments. With the ongoing refinement and adaptation of these methodologies in diverse scenarios, robots and AI systems are expected to achieve more effective communication with humans, enhancing their capabilities across various applications.</p>	

論文審査の結果の要旨及び担当者

氏 名 (時 嘉 祺)			
	(職)		氏 名
論文審査担当者	主 査	教 授	石黒 浩
	副 査	教 授	飯國 洋二
	副 査	教 授	佐藤 宏介

論文審査の結果の要旨

本論文は、サービスロボットおよびAIシステムが人間の内部状態、特に感情と意図を認識するための新たな方法を提案し、その有効性を実証したものである。

感情認識に関して、本研究は従来の顔表情や音声のトーンに重点を置いた手法を超え、ジェスチャーを含む身体動作の豊富な情報を利用することの重要性を明らかにした。特に、自己注意機能を強化した空間-時間グラフ畳み込みネットワークを用いて、人体の骨格構造とその動的な関係をモデル化し、感情認識の精度を大幅に向上させた。さらに、音声、テキスト、骨格データのグラフ注意ベースの融合方法を提案し、これにより感情分類における総合的なデータ融合の有効性を実証した。

意図認識においては、本研究は実世界のシナリオを拡大し、建物の入口で観察された人間の行動から意図を収集し、これを分類する新しいアプローチを採用した。提案された空間-時間グラフ畳み込みネットワークは、スケルトンベースの意図認識において、基準モデルを上回り、実験においては人間の予測能力をも超越した。さらに、歩行者の横断意図を予測するための深層学習モデルが提案され、自動運転システムにおける安全性と便利性の向上に貢献する可能性が示された。

本研究は、感情および意図認識の分野において、新たな視点と技術的進歩をもたらした。特に、ジェスチャーと身体動作の分析に重点を置いたアプローチは、人間とロボットまたはAIシステムとのより自然で効果的なコミュニケーションを可能にする。この研究成果は、サービスロボットやAIシステムが人間の感情と意図をより正確に理解し、それに応じた対応をすることの重要性を強調している。

総じて、本論文は、感情および意図認識に関する深層学習に基づく革新的な方法論を開発し、その効果を実験的に検証することで、人間とロボットまたはAIシステムとの相互作用の向上に貢献するものであると評価される。そのため、この論文は博士（工学）の学位論文として価値のあるものと認める。