

Human Gesture Analysis and Generation from Text for Communication Robot

手嶋, 仁志

<https://hdl.handle.net/2324/7363830>

出版情報 : Kyushu University, 2024, 博士 (情報科学) , 課程博士
バージョン :
権利関係 :



氏 名 : 手嶋 仁志

論 文 名 : Human Gesture Analysis and Generation from Text for Communication Robot

(対話ロボット向けのテキストに基づく人間ジェスチャーの解析および生成)

区 分 : 甲

論 文 内 容 の 要 旨

Human gesture analysis and generation play a critical role in enhancing the interaction capabilities of communication robots. While existing data-driven methods generate gestures from textual inputs, they often lack coherence and intuitiveness. To address these limitations, this research proposes a systematic approach for gesture generation using natural language, focusing on semantic alignment with textual content. This work integrates deep learning techniques, gesture libraries, and contextual understanding to achieve natural and diverse gesture generation for communication robots or human avatars.

We first introduce a gesture transformation system that leverages gesture types as intermediate representations based on psycholinguistic theories. The system clusters imagistic gestures using Labanotation-based metrics and maps these gestures to corresponding words in a gesture library. Implemented on a DIY conversational robot, our system was evaluated against various conversational media. Results demonstrated that our approach significantly improved user impressions, making robots more intuitive and friendly for conversation.

To enhance gesture quality, we introduced a type-specific gesture generation system that categorizes gestures explicitly into Beat, Imagistic, and No-Gesture categories. The method uses a pre-trained network to predict gesture types and generates type-specific gestures by consulting dedicated libraries. User studies confirmed that this approach produces human-like gestures and improved appropriateness compared to end-to-end baseline methods. The system also showed versatility for avatars and humanoid robots.

We further propose ACT2G, an attention-based contrastive learning framework for text-to-gesture generation. This method maps gestures and textual features into a unified latent space, enabling the synthesis of content-aware gestures with significant diversity and realism. Through user studies and comparisons, ACT2G outperformed state-of-the-art methods, achieving higher ratings for gesture appropriateness and naturalness. Additionally, ACT2G demonstrated the ability to produce context-specific gestures by adjusting attention

weights.

Also, building on the foundation of gesture generation from textual descriptions, this research explores an advanced approach to synthesizing human motions that involve interactions with objects in 3D scenes. The proposed method, called SceneMDM, decomposes motion generation into key steps: extracting key poses informed by textual and scene contexts, and synthesizing full trajectories aligned with these poses and planned paths. This disentangled framework eliminates the dependency on hybrid datasets combining text, motion, and scene, leveraging independent datasets instead. By integrating motion diffusion models and affordance-based techniques, SceneMDM enables the generation of diverse, contextually accurate motions, advancing the possibilities for interaction-rich, human-like animations in applications ranging from robotics to VR environments.

This research advances the state-of-the-art in gesture analysis and generation by introducing a systematic, type-specific, and content-aware approach. Our contributions enable communication robots to generate natural, diverse, and contextually relevant gestures, bridging the gap between textual semantics and non-verbal expressions. Our future work aims to enhance real-time interaction by incorporating facial expressions and adapting gestures dynamically based on speech patterns.