

Unity と ROS を統合したクラウド型マルチモーダル対話経験蓄積プラットフォーム

Cloud-based Platform Utilized Unity and ROS Frameworks
for Collecting Multimodal Human-Robot Interaction Experiences

水地 良明 *1 稲邑 哲也 *1*2
Yoshiaki Mizuchi Tetsunari Inamura

*1国立情報学研究所 National Institute of Informatics
*2総合研究大学院大学 SOKENDAI(The Graduate University for Advanced Studies)

In this paper, we propose a new software architecture of a Human-Robot Interaction (HRI) simulator in virtual reality (VR) environments. Since collecting and storage of massive multimodal interaction experience is important for research in HRI, we have proposed a cloud-based VR platform named SIGVerse, to reduce the cost of development of real robots and cost for interaction experiments in real world. However, the reusability of virtual robot controllers is restricted in the conventional platform due to difference between controller architecture for the VR and real robots. To solve this issue, we propose a new platform architecture utilizing the Unity and ROS frameworks. To demonstrate the feasibility of the proposed architecture, we evaluated the performance of simulated multimodal information in actual interaction applications and checked the latency between synchronized avatars in cloud computing.

1. はじめに

ロボットの知能獲得において、軌跡の模倣や感覚運動のマッピングはシンプルな学習方法の一つであるが、日常生活行動のモデル獲得には、単純なセンサ情報のみならず、身体動作や発話などのマルチモーダルな情報を含めた人間とロボットによる対話経験が重要である。しかしながら、日常生活行動における身体的・社会的情報の蓄積は、膨大な実験コストやハードウェアの維持管理による困難さが問題となる。

この問題を解決する方法の一つは、シミュレータを利用したハードウェアの省略によるコスト削減である。これまで様々なシミュレータがロボット開発のために提案されているが、シミュレーション環境において人間の行動を再現するためには大きな課題がある。また、大規模な対話経験の蓄積のために、多数の被験者を実験環境に収集することも困難な課題となる。このような課題を解決するアプローチの一つは、クラウドコンピューティングと VR システムの利用である。

稲邑らは、人間がアバターとしてログインし、仮想ロボットとの対話を可能とする SIGVerse と呼ばれる VR システム [Inamura 10] を提案している。SIGVerse はサーバ/クライアント形式を採用しており、サーバ上の仮想環境内に異なるコンピュータ上から複数のクライアントが参加することを可能としている。しかしながら、独自の API/SDK によりロボットソフトウェアの再利用性が乏しいことや VR インタフェース毎にドライバソフトウェアを開発しなければならないことが課題となっていた。

本研究では、上記の課題を解決するため、Unity と ROS を活用した新たなアーキテクチャを提案する。Unity の利用により、ドライバソフトウェアの開発を必要とせず、多種の VR デバイスの利用を可能とする。さらに、Unity と ROS をブリッジする方法の構築により、データ量の大きな画像情報などの仮想センサフィードバックおよび実機ロボットのためのソフトウェアを利用した仮想ロボットの制御を可能とする。これにより、ソフトウェア開発におけるコストの大幅な削減が期待される。

る。本稿では、新たなアーキテクチャについて詳述するとともに、基本的な対話実験への応用における性能を評価する。

2. マルチモーダル対話経験蓄積プラットフォームに求められる機能

本研究では、身体的・社会的対話が求められる日常生活行動におけるマルチモーダル情報の収集が目的となる。ロボットが人間の振る舞いを観測、学習し、経験に基づいて曖昧性を解消していくためには、以下のようなマルチモーダル情報の収集が要求される。

- 1) 身体情報 (視線を含む動作・ジェスチャ)
- 2) 視覚情報 (エージェントが見ている画像情報)
- 3) 空間情報 (エージェントや物の位置・姿勢)
- 4) 音声対話 (エージェントの発話)

また、VR 環境において、それらの大規模な収集を実現するシステムには以下のような機能が要求される。

- i) 被験者が VR 環境内のアバターにログインできること
- ii) 複数のユーザがインターネットを介して一つの VR 環境に同時にログインできること
- iii) 時間的に連続なマルチモーダル対話データを記録・生成できること
- iv) 実機ロボットの制御プログラムを仮想ロボットに適用できること

現在、様々なロボットシミュレータが存在するものの、1)~4) の要求を同時に満たすものは提案されていない。従来の SIGVerse(ver.2) は、1)~4) のマルチモーダル情報および i)~iii) の機能における要求を満たしているものの、独自の API/SDK による制御プログラムの再利用性の制約が課題となっている。

3. 提案する SIGVerse のアーキテクチャ

図 1 に提案する SIGVerse(ver.3) の構成を示す。SIGVerse は、サーバ/クライアント形式となっており、サーバを介して

連絡先: 水地 良明, 国立情報学研究所, 〒101-8430 東京都千代田区一ツ橋 2-1-2, mizuchi@nii.ac.jp

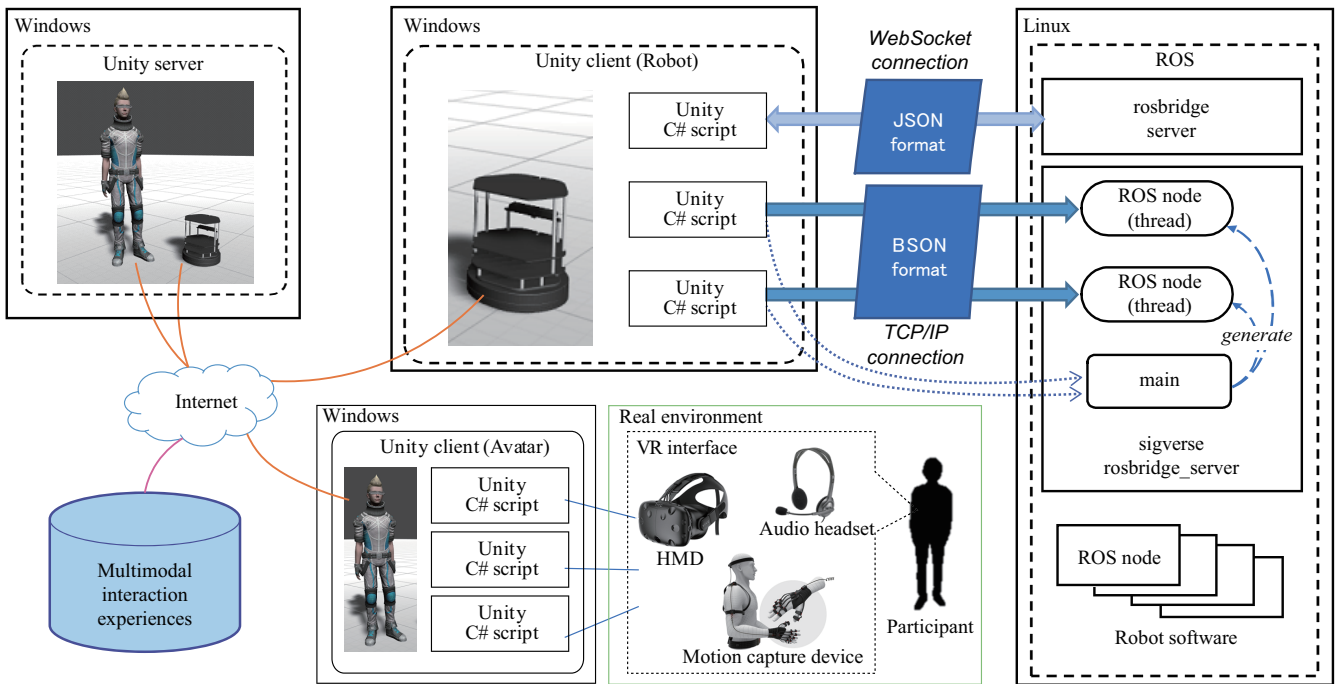


図 1: SIGVerse (ver.3) のシステム構成

各クライアント上のアバターやロボット、物体の状態が同期される。被験者は、ヘッドマウントディスプレイ (HMD) などの VR インタフェースを利用してアバターにログインすることができる。仮想ロボットには、ROS 環境とのブリッジ機能により、実機ロボットを制御するためのソフトウェアを再利用することができる。また、独自のブリッジ機能により、画像情報などのデータ量の大きなセンサ情報のリアルタイムなフィードバックも可能とする。これらにより実現される VR 空間内での身体的・社会的な対話経験をデータベースに蓄積し、共有することで、大規模なマルチモーダル情報に基づくロボットのスキル獲得や学習が可能となる。

3.1 仮想環境の構築

日常生活の行動における身体的・社会的な情報を蓄積するためには、大規模で複雑な 3 次元環境を構築しなければならない。Unity では、その構成要素となる物体の 3D モデルやスクリプトが Assets として豊富に共有されている。さらに、Unity は操作性の高いユーザインタフェースを備えており、それらを活用することで容易に大規模かつ複雑な仮想環境が構築できる。また、高精細な仮想のセンサ情報の生成や物体間の物理演算シミュレーションもサポートしている。

3.2 VR インタフェースの利用

仮想空間内でマルチモーダルな対話を行うため、被験者は没入型の VR インタフェースを利用してアバターにログインする必要がある。近年では、VR ヘッドセット (HTC Vive, Oculus Rift and Touch, FOVE, etc.) や簡易運動計測装置 (Kinect V2, Leap Motion, Perception Neuron, etc.) などの多様な VR インタフェースが開発されている。多様な対話経験を収集するためには、実施する対話経験に合わせ、これらの VR インタフェースを適切に利用する必要がある。Unity では、ほとんどの VR インタフェースのドライバソフトウェアが Assets として既に提供されており、マルチモーダルなタスクの実演や対話経験の観測が容易に実現可能となっている。

3.3 サーバ/クライアントシステムの構築

本シミュレータは、Unity の既存機能を活用したサーバ/クライアント形式となっている。サーバとクライアントは、アバターやロボット、家具などのオブジェクトから構成される同一のシーンを持つ。ネットワークを経由して、環境内の登録されたオブジェクトの情報を通信することで、それぞれのシーンが同期される。クライアントがログインすると、アバターやロボットなどの動的なオブジェクトがサーバによって VR 環境中に生成される。各クライアントの所有するアバターやロボットの状態は逐次サーバに送信され、それらの情報がブロードキャストされることで、それぞれのアバターやロボットによる対話が実現される。

3.4 Unity と ROS 間のブリッジ機能

近年、多くのロボットがロボットミドルウェアを利用して開発されており、ROS は最もスタンダードなミドルウェアとして利用されている。実機ロボットと仮想ロボットの制御プログラムにおける互換性を保つためには、Unity と ROS 間で制御コマンドやセンサ情報を通信するためのブリッジ機能が最も重要な要素となる。ROS では、JSON 形式データを利用した WebSocket 通信により、外部プログラムとのデータ通信を可能とするブリッジ機能が提供されている。しかしながら、画像情報などのセンサフィードバックを実現するためには通信速度がボトルネックとなる。Unity と ROS をブリッジすることでデータサイズの小さい制御コマンドやセンサ情報を通信するシステム [Codd-Downey 14, Hu 16] も提案されているものの、膨大なデータ通信については議論されていない。そこで、BSON 形式データを利用した TCP/IP 通信によるブリッジ機能を提案することで、ロボットの制御に必要なリアルタイムな仮想センサフィードバックを実現する。

3.5 マルチモーダルな対話経験の記録・再生

マルチモーダルな対話経験データを共有するためには、仮想空間内での対話経験の実演のみでなく、それらを記録・蓄積

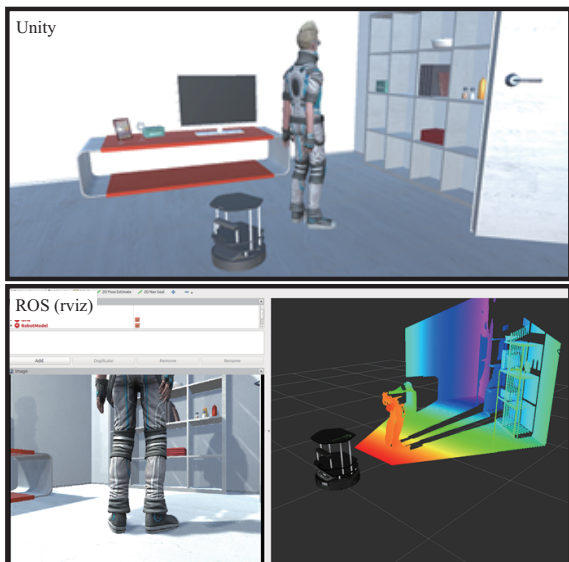


図 2: Unity 内の VR 環境および ROS 側で受信された画像情報

する機能が重要となる。その実現のためには、エージェントのセンシング情報を記録するのみで不十分であり、マルチモーダルな対話経験を再現できるレベルで記録する必要がある。本シミュレータでは、能動的に行動するエージェントの言動を記録しておけば、VR 環境での事象をシミュレートすることで、マルチモーダルな対話経験を再現することができる。これらをネットワーク経由でデータベースに蓄積し、共有することで、クラウドコンピューティングと VR システムを併用した大規模な対話経験の収集と収集データの相互利用が可能となる。

4. 提案プラットフォームの性能評価

4.1 仮想センサ情報のフィードバック

ロボット制御に必要な仮想センサによるリアルタイムなセンサフィードバックが可能であるか検証するため、仮想の RGB-D センサ情報の生成と Unity-ROS 間での通信性能を評価した。図 2 に Unity 内の VR 環境と ROS 側で受信された RGB 画像、Depth 情報を示す。

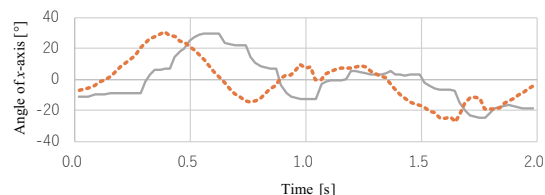
実験では、Xeon E5-2687W CPU を搭載したデスクトップ PC と同一マシン上の仮想マシンを用いて評価した。画像サイズは 640×480 ピクセルであり、RGB 画像、Depth 画像のデータサイズはそれぞれ 900 [KB]、600 [KB] である。JSON 形式を利用した WebSocket 通信では、ROS 側で受信される両画像データの頻度は 0.55 [fps] であった。一方で、BSON 形式を利用した TCP/IP 通信では、57.60 [fps] となった。この結果は、提案したブリッジ機能により、リアルタイム性を十分に満たすセンサフィードバックが実現されていることを示している。

4.2 ネットワーク経由で同期したクライアント間の遅延

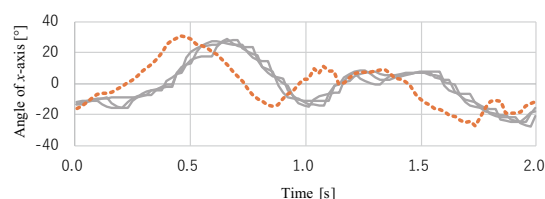
サーバ/クライアント形式による複数アバターによる対話が可能であるか評価するため、遠隔のサーバを経由して同期されたローカルアバターと他のアバター間の動作遅延を計測した。実験では、同一環境に設置した複数 PC 上にクライアント、遠隔地に設置した PC 上にサーバを起動した。クライアントとサーバ間の距離は 33.5 [km] である。クライアントと同じ場所に設置した PC 上で Axis Neuron ソフトウェアを利用し、再生した動作データを各クライアント内のアバターに反映した。



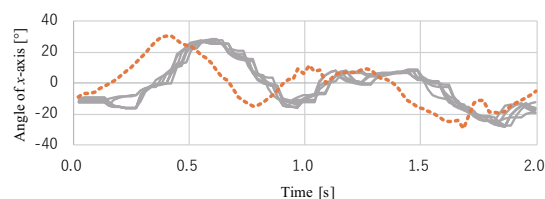
図 3: ローカルアバターと同期されたアバターの見かけのずれ



(a) 2クライアント時のアバターの動作



(b) 4クライアント時のアバターの動作



(c) 8クライアント時のアバターの動作

図 4: ローカルアバターと同期されたアバター間の動作

クライアントの 24 関節の状態を 10 [fps] で同期した。各関節の状態データのサイズは 36 [B] であり、それらは圧縮して通信される。また、同期された動作は補間され、アバターに反映される。

図 3(a)~(c) がそれぞれ 2 クライアント、4 クライアント、8 クライアントがログインした際に同期されたアバターの動作である。図中の色付きのアバターがローカルのアバター、白色のアバターがサーバ経由で同期された他のアバターである。

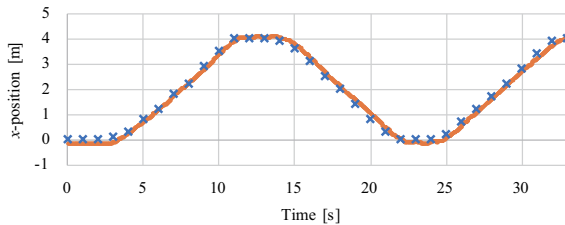
あるクライアントにおけるローカルアバターとサーバを経由して同期される他のアバターを動作を記録し、その遅延を計測したものを図 4 に示す。図の横軸は時間、縦軸は右手首の x 軸の角度であり、破線がローカルアバターの動作、実線が他のアバターの動作を表している。結果から、他のアバターの動作データは僅かに揺らいでおり、約 0.2 [s] の遅延があるものの、インターネットを経由して動作が同期されていることがわかる。これらは、アバター同士が基本的な対話を行う上では十分な性能といえる。

4.3 提案プラットフォームを用いたアプリケーション

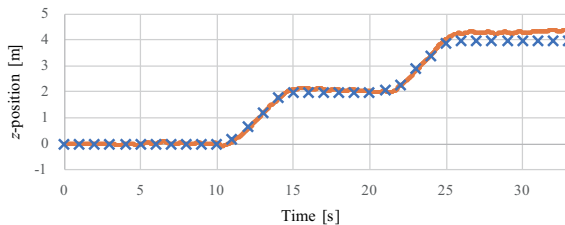
VR 環境における人とロボットによる対話アプリケーションの一例として Turtlebot を利用した人の追従タスクを実装し、同一の制御プログラムを利用したときの実機ロボットと仮想ロボットの振り舞いを比較した。実験の様子を図 5 に示す。本



図 5: 人の追従タスクにおける実機ロボットと仮想ロボットの振る舞い



(a) 人間とアバターの水平方向の位置



(b) 人間とアバターの垂直方向の位置

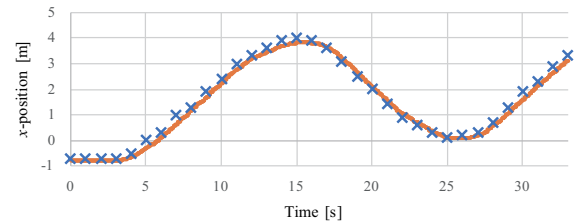
図 6: 人間とアバターの位置

実験では、まず、実世界で運動計測装置 (Perception Neuron) を装着した被験者を実機ロボットに追従させた。この時に記録された動作をアバターに反映し、同一の制御プログラムを用いて仮想ロボットにアバターを追従させた。ロボットの制御プログラムには、Turtlebot 用の ROS パッケージに含まれる `turtlebot_follower` を利用した。

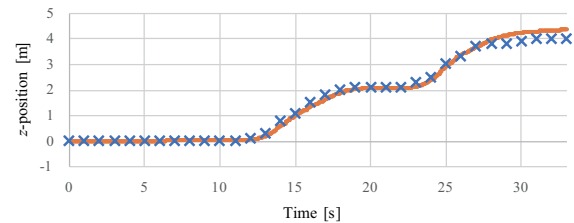
図 6 に時間毎の被験者とアバターの位置を示す。図 6(a) が図 5 における垂直方向の位置、図 6(b) が水平方向の位置を表している。また、図中の×が被験者の位置、実線がアバターの位置を表している。結果から、実世界の人間の動作が大きな誤差なく、アバターの動作に反映できていることがわかる。25 [s] 付近からアバターの水平方向の位置が僅かにずれているが、これは Perception Neuron の持つ慣性センサの累積誤差によるものである。図 7 に実機ロボットと仮想ロボットの位置を示す。図 7(a) がロボットの垂直方向の位置、図 7(b) が水平方向の位置を表しており、×が実機ロボットの位置、実線が仮想ロボットの位置を表している。結果から、人間とアバターの動作と同様に実機ロボットと仮想ロボットの動作がほぼ一致していることがわかる。また、アバターの動作に従って、仮想ロボットの水平方向の位置も僅かにずれている。これらの結果は、実機と同じ制御プログラムを再利用することで、仮想のセンサフィードバックにより VR 空間内でロボットの振る舞いを同じように再現できることを示している。

5. まとめ

本稿では、従来の SIGVerse において課題となっていたロボット制御プログラムの再利用性の向上、VR インターフェース



(a) 実機ロボットと仮想ロボットの水平方向の位置



(b) 実機ロボットと仮想ロボットの垂直方向の位置

図 7: 実機ロボットと仮想ロボットの位置

のドライバソフトウェア開発コストの低減を目的として、新たなアーキテクチャを提案した。上記の目的を実現するため、VR アプリケーション開発を容易にする Unity と実機ロボットと互換性の高い制御プログラムの利用を可能とする ROS を採用した。また、VR 空間における人とロボットによる対話アプリケーションの開発に必要なリアルタイムなセンサフィードバックを可能とする Unity と ROS 間のブリッジ機能を提案した。提案するアーキテクチャの基本性能を評価するとともに、基本的なアプリケーションの実現において十分な性能を持つことを示した。

謝辞

本研究の成果の一部は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務、および JST CREST の支援を受けたものです。

参考文献

- [Inamura 10] Inamura, T., Shibata, T., Sena, H., Hashimoto, T., Kawai, N., Miyashita, T., Sakurai, Y., Shimizu, M., Otake, M., Hosoda, K., Umeda, S., Inui, K., and Yoshikawa, Y.: Simulator platform that enables social interaction simulation - SIGVerse: Sociointelligence simulator, in the IEEE/SICE Int. Symp. on System Integration, pp. 212-217 (2010)
- [Codd-Downey 14] Codd-Downey, R., Forooshani, P. M., Speers, A., Wang, H., and Jenkin, M.: From ROS to unity: Leveraging robot and virtual environment middleware for immersive teleoperation, in IEEE Int. Conf. on Information and Automation, pp. 932-936 (2014)
- [Hu 16] Hu, Y., and Meng, W.: ROSUnitySim: development of a low-cost experimental quadcopter testbed using an arduino controller and software, J. Simulation, Vol. 92, Iss. 10, pp. 931-944 (2016)