

ユーザとモノのネットワーク体験を創作するための AR システム

- Monokko -

1 背景

人間の生活空間には、家具や家電、趣味や思い出の品など、感情や記憶と結びつく多様なモノが存在する。これらは単なる道具ではなく、生活体験を形成する大切な要素である。IT 技術を活用し、これらのモノをエージェントとして扱うことで、人工ペット型エージェントによる心の安らぎの提供や、スマートエージェントによる家事・コミュニケーションの補助など、生活の利便性と豊かさが向上する可能性が示される。

さらに、Apple 社の Vision Pro（2024 年発売）などの最新の高性能デバイスの進化により、物理的なハードウェアに依存せず、現実空間のモノをデジタルに拡張してエージェント化するアプローチが現実味を帯びている。これにより、モノ自体に新たな機能や存在感が付与され、従来の枠を超えた新しい生活体験が提案されることが期待される。

2 目的

本プロジェクトでは、日常生活のモノをエージェント化し、ユーザーがモノを介して愛着を育む仕組みを構築することを目指す。その XR 技術を活用することで、ユーザーがデバイスを装着していない時でも「そのモノにはデジタルの分身がいる」と感じられるような、新たな生活習慣や感性の変化を促すことを狙う。

このシステムでは、モノが単なる便利なツールとしてだけでなく、ユーザーの思い出や感情を引き出す「パートナー」のような存在になることを目指している。これらの要素を組み合わせ、本プロジェクトでは「モノのエージェント＝モノッコ」を開発し、ユーザーが日常の中で自然に対話し、愛着を育めるインタラクションを実現するシステム「Monokko」を構築することを目的としている。

3 開発の内容

Monokko では、生活空間の任意の物を画像解析技術を用いて検出したのち、XR 技術でモノッコの様相を空間に提示し、LLM 技術でそのモノッコと会話可能となるようなアプローチをとる。この際、複数種類のデバイスでモノッコを観察可能であり、それぞれ異なる体験が得られる。また、モノッコの性格や容姿などをユーザーが自由に創作可能である。そして、キューブンという独自デバイスを開発し、システムとしての世界観を深める。

3.1 管理アプリ

管理アプリは、モノッコの作成・登録、設定変更、見た目のカスタマイズという三大機能に大別される（図 1）。まず、ユーザーはカメラを用いて対象物体を撮影し、そのスキャンデータを取得することで、現実世界のオブジェクトを Monokko システム内で認識可能な状態にする。次に、サーバーに登録された各モノッコの名前、性格、対話スタイルなどの基本パラメータを編集する設定変更機能がある。モノッコの見た目のカスタマイズ機能では、瞳孔の大きさ、色、形状（丸型、ハート型、星型など）、眉の種類、アクセサリ、音声のトーン、全体の大きさなど多様な視覚的要素を調整可能である。これらのカスタマイズ要素は単なる外見の変更に留まらず、ユーザーがモノとの心理的距離を縮め、愛着を生み出す体験につながると考えている。

技術的実装としては、管理アプリは MonokoScope アプリと同様に、Unity ゲームエンジンを採用し、UI は Unity の uGUI を用いてレスポンス画面レイアウトを構築している。バックエンドとの通信は REST API を介して JSON フォーマットで行われ、モノッコの設定情報やカスタマイズデータをサーバー側で一元管理する仕組みとなっている。



図 1: 管理アプリ

3.2 MonokoScope

MonokoScope は、Monokko システムの中核をなすアプリケーションであり、ユーザーがモノッコと対話するための主要インターフェースである。iPad 版と Apple Vision Pro 版の 2 種類が存在し、両者は基本的な機能とコンセプトを共有しつつ、それぞれのデバイス特性を活かした独自の体験を提供する。iPad 版は手軽にモノッコ体験ができるよう最適化され、家庭や教育現場など幅広い環境での利用を想定している（図 2）。一方、Vision Pro 版は没入感の高い体験を実現するために、UI 要素を極力排除し、Hand Tracking や Object Tracking 技術を導入して、物体の 3D 形状に合わせたモノッコの配置と自然な対話を可能にしている（図 3）。



図 2: iPad 版 MonokoScope



図 3: Apple Vision Pro 版 MonokoScope

両バージョンは Unity を基盤に AR Foundation ライブラリを活用して開発した。iPad 版は ARKit、Vision Pro 版は ARKit に加え PolySpatial ライブラリを採用することで、共通基盤を維持しつつ各デバイス固有の機能を実装している。カメラやセンサーを用いて現実空間の物体を認識し、サーバー設定に応じたモノッコの 3D 座標を特定・重畳表示するほか、専用ジェスチャーやボタンで音声入力モードを起動し、Whisper API で音声をテキスト変換、入力に基づく適切な応答を生成する仕組みである。さらに、眼球の多層構造による豊かな表情表現や自然な瞬き、視線のランダムな動き、抽象的な鳴き声と改行コードを用いたフキダシ分割によって、機械的な印象を排除した自然な対話が実現されている。

iPad 版は横持ちで iPad スタンドに設置し、カメラ範囲内の対象物にモノッコが出現する仕組みとなっている。技術面では、iPad Pro の LiDAR カメラを活用し、後述する YOLO による物体認識と深度カメラによる 3D 座標特定を組み合わせている。UI はシンプルかつ直感的で、画面四隅に機能ボタンを配置し、モノッコが主役となるよう画面の大部分をカメラビューに割り当てている。

Vision Pro 版は、没入型体験を重視しており、immersive モードで全画面表示されることで、ユーザーが自らの生活空間にモノッコが存在しているかのような強い没入感を体験できるよう設計している。また、Hand Tracking を活用してモノッコに対する「撫で」インタラクションが実現され、これにより物理的な触れ合いを通じた愛着形成が促進される。Object Tracking 技術により、3D スキャンデータから生成された認識モデルを用いて、物体の正確な 3D 形状を把握し、モノッコの目玉を物体表面に自然に配置することを可能としている。

3.3 画像解析システム

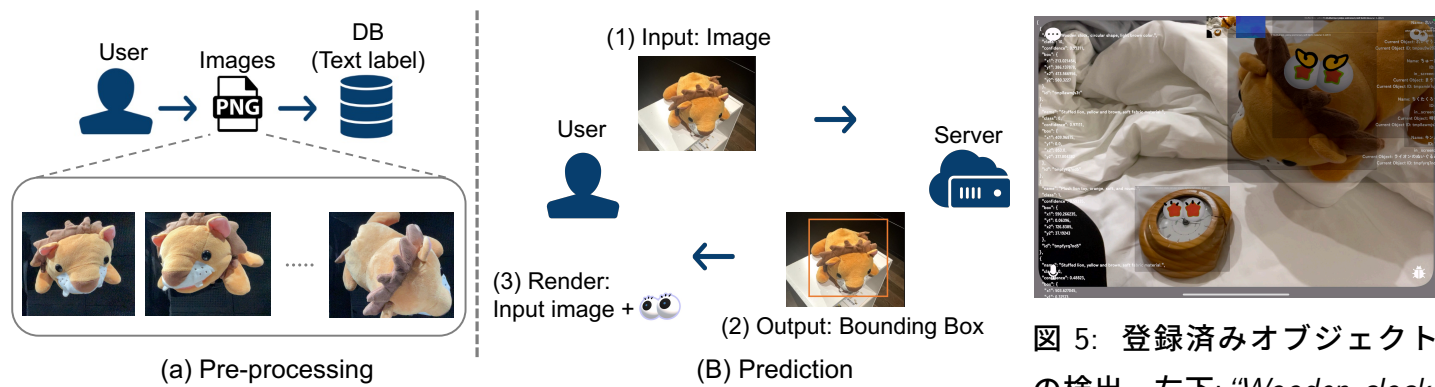


図 4: 画像解析システム概要。(a) 前処理：ユーザーが行うオブジェクト画像の登録。登録された画像から Image-to-Text 技術によりオブジェクトのラベルを生成し保存。(b) 推論時：登録されたオブジェクトが画像中にあればバウンディングボックスを取得。

図 5: 登録済みオブジェクトの検出。左下: "Wooden clock, circular shape, light brown color.", 右上: "Stuffed lion, yellow and brown, soft fabric material"

本システムは、図 4 に示すようにユーザーが身の回りの所望の物を（一枚以上）撮影、画像登録をすることで、ユーザーは上述のデバイスを介してモノッコを視認できる。具体的には、ユーザーが登録した画像について、Image-to-Text 技術を用いて、オブジェクトに関するテキストラベルを生成する。次に、Monokoscope により得られる周囲の画像について、サーバー側でオブジェクト検出モデルによる推論を行う。特に、登録されたオブジェクトについて画像中の位置情報を検出し、バウンディングボックスを獲得する。以上のプロセスを通して図 5 に示すように、登録済みオブジェクトに関するモノッコの可視化を行う。

オブジェクト検出では YOLO-World を用いて入力画像からバウンディングボックスを獲得する。YOLO-World では、登録されたオブジェクトのラベルを用いたクロスモーダル推論により、固定クラスラベルに依存しないオブジェクト検出が実現される。したがって、オブジェクトに関する画像を大量に用意する必要もなく、モデルの追加学習が不要である。これにより、ユーザーはオブジェクトの登録からモノッコの可視化まで即座に実行可能である。また、オブジェクトについて複数のラベルを用意することで、ある程度頑健な検出が可能であることを確認した。

3.4 会話システム

本プロジェクトでは、図 6 に示す仕組みでモノッコと対話可能であり、モノッコ同士の会話も行われる。具体的には、ユーザーが入力する文章について、モノッコのキャラクターを反映した LLM エージェントが応答を行う。プロンプトは、モノッコのペルソナおよび個々のキャラクターに設定されている取り憑くオブジェクト、性格や過去の思い出等の情報を用いて記述されている。例えば、図 7 に示すモノッコは冷蔵庫に取り憑いており、設定された話し方の特徴に合わせて応答が生成されている。また、ユーザーの与える文章のコンテキストを理解して、あたかも本当にその場にいるような、その物自身が話しているような体験を生み出せる。

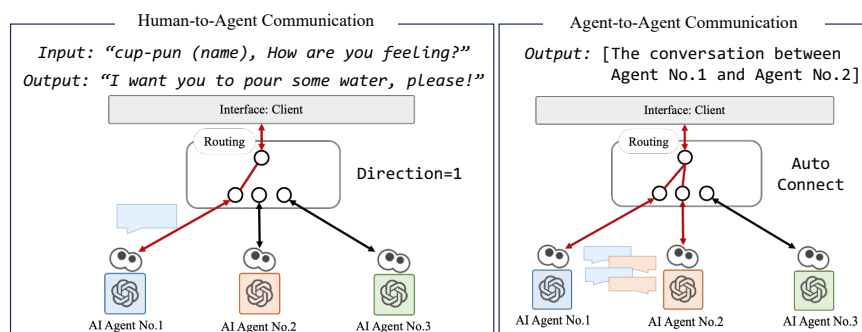


図 6: 会話システム概要。左: ユーザーによるモノッコの応答生成。右: モノッコ同士の会話生成。自動で対応するエージェントが選択される。

図 7: 冷蔵庫に取り憑くモノッコの応答例。
入力: 「今、冷蔵庫が開いたままです。」

ユーザー体験の品質向上のため、モノッコ同士の会話については、過去の会話に基づき自律的に停止する処理と会話履歴の外部評価により打ち切る処理を用意している。これにより、ユーザーが興味を失うほどに重複したトピックが続くことを未然に防ぎ、程よいタイミングで会話を締めくくることを実現している。

3.5 キューブン



図 8: キューブン外観



図 9: キューブン操作の流れ

キューブンは、Monokko システムの世界観を深化させ、ユーザーとモノッコの関係性を豊かにする専用デバイスである。Raspberry Pi 5 を基盤とし、マイク、スピーカー、タッチディスプレイを搭載している。箱状のキャラクターとして設計され、3D プリンタで製作されたプラスチックカバーを装着し、中央に目玉表示、右上にフキダシ表示を設けている（図 8）。ユーザーがタッチすると内部プログラムが起動し、音声合成で呼びかける（図 9）。録音・文字起こしされた会話はモノッコのバックエンドシステムと共有され、エピソードや思い出として蓄積される仕組みである。

4 従来の技術との相違

モノに一切の加工をせずエージェント化することができる点は、本システムの大きな特徴である。また、会話システムのみならず、触覚的な「撫で」インタラクションを、最新の Object Tracking 技術を導入して実装したことで、モノのエージェントと交流するという体験の質をより一層高めている。さらに、本システムの特徴として、システム独自の世界観を持っている点が挙げられる。これにより、技術的な部分と同時に世界観を深めていくことで、システムの質向上が行われている。

5 期待される効果

まず、チャットボットによる会話と「撫で」アクションへの反応で、ユーザーが自然にモノに興味や愛着を持つ行動が促進される。これにより、モノは単なる操作対象や情報ツールではなく、ユーザーのライフスタイルに溶け込み、感情や記憶と結びつく「パートナー」として機能することが期待される。

また、本システムは物理的なモノにデジタルな付加価値を付与し、従来の家電やツールでは得られなかったエンターテインメント性やコミュニケーションの幅を実現する。ユーザーは、普段使用するモノが生きたキャラクターのように振る舞い、親しみやすい会話や触覚的反応を示すことで、生活空間全体が温かみのあるものになり、行動や感情にポジティブな影響を受けると期待される。

さらに、デジタルとフィジカルの境界を意識せず、自然にエージェントと対話・触れ合える体験を提供することで、未来のスマートホームや生活支援システムとしての普及・活用が期待される。

6 普及の見通し

本システムの強みは、体験の面白さと可愛らしい世界観にあるため、特に若年層や家族、さらにはエンターテインメントやライフスタイルに敏感な層へアプローチすることで、広範に受け入れられると期待される。モノに加工を施すことなくそのままエージェント化できる柔軟なアプローチにより、XR 技術の普及に後押しされる形で、既存の生活空間にシームレスに溶け込んで行ける可能性が高い。加えて、本プロジェクトで構築した独自の世界観は、XR アプリケーションとしての利用に留まらず、マルチメディア展開やキャラクターグッズ、さらにはコンテンツとしての拡大展開も視野に入れたブランド戦略と連動し、幅広いマーケットへの普及を狙う戦略も考えられる。

7 クリエータ名（所属）

- 後藤汰誓（九州大学 芸術工学部 芸術工学科 音響設計コース 4 年）
- 大塚敏郎（九州大学 芸術工学部 芸術工学科 メディアデザインコース 4 年）
- 石山遼（九州大学 大学院システム情報科学府 情報理工学専攻 修士 1 年）

（参考）関連 URL

- ティザーサイト：<https://www.monokko.com/>
- ソースコード：<https://github.com/monokko-project>