

カメラ機能および移動機構を実装した音声認識付きマグボットの製作

稲村 虹輝¹⁾, 飯野 拓海¹⁾

指導教員 林 誠治¹⁾

1) 拓殖大学工学部 電子システム工学科 林研究室

【要約】

「マグボット」は、PC やスマートフォンから遠隔によりロボットの表情を操作したりや音声合成による音声出力が可能なソーシャルロボットである。本研究室ではこれまでに音声認識や雑談会話機能を追加実装することで、機能の拡張を行ってきた。本稿では Web カメラを用いた動画像認識により視覚的な操作の機能を追加するとともに、キャタピラの移動機構を搭載することで PWM 制御によるマグボット本体の移動動作を可能とした。動画像認識と移動機構の連携によりマグボットの更なる高機能化を図る。

キーワード：ソーシャルロボット，遠隔操作，動画像認識，移動機構，音声認識

1. はじめに

これからの世の中では身の回りにロボットが存在するのが当たり前になり、ロボットと人がコミュニケーションをとる機会が増えると考えられる。本研究室ではこれまで、マグボット[1]を基盤としたソーシャルロボットに音声認識などの機能を追加してきた[2]。本稿では、さらに Web カメラによる動画像認識およびキャタピラを搭載した移動機構の追加を実装したので報告する。

2. 音声認識を搭載したマグボット

2.1. マグボットについて

マグボット(図 1)はオープンソースのロボットで Arduino と Raspberry Pi を組み合わせて作られている。Raspberry Pi の中には Web サーバーを設置しており、PC・スマートフォン・タブレットから Web ブラウザ経由でアクセスし遠隔操作が可能である。



図 1: 本稿のマグボット

Arduino はハード制御を、Raspberry Pi はネットコミュニケーションと合成音声・音声認識・雑談会話を担当する。また電力は Raspberry Pi から USB ケーブルで Arduino に供給し、同時にシリアル通信を行う(図 2)。

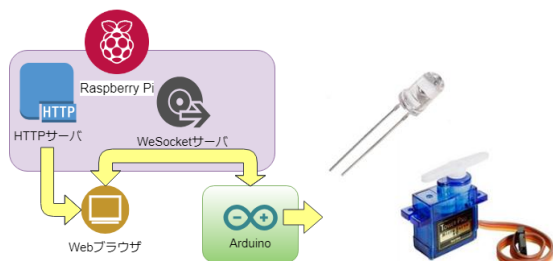


図 2: マグボットの構成図

2.2. ソフトウェアについて

2.2.1. HTTP サーバー

HTTP サーバーは Nginx(エンジンエックス)を使用する。Nginx は少ないメモリで高速に動く軽量サーバ

ーである。PC、スマートフォン、タブレット等の Web ブラウザから有線/無線 LAN 経由で Raspberry Pi 内の HTTP サーバーにアクセスし、マグボットの操作画面である HTML ファイルを開くことができる。

2.2.2. WebSocket サーバー

WebSocket サーバーはブラウザとサーバーの間で双方向通信、非同期通信を実行するためのプロトコルを使用したサーバーである。マグボットでは操作画面から WebSocket サーバーを経由して Arduino に接続した LED やサーボを制御するために用いる。

2.2.3. 日本語音声合成ソフトウェア Open JTalk

Open JTalk とは名古屋工業大学の徳田・李・南角研究室が開発している文字列を音声(Wav ファイルなど)に変換するオープンソースの日本語テキスト音声合成システムである。マグボットでは操作画面から文字列を入力して Open JTalk によりその Wav データを取得し、`aplay` コマンドによりスピーカから出力する。

2.2.4. Julius 音声認識

Julius は音声認識システムの開発・研究のためのオープンソース大語彙連続音声認識エンジンである。単語辞書のモジュールを組み替えることで、ソーシャルロボットに特化した音声認識語彙システムを構築できる。本稿では、音声の収録に USB マイクを使用し、Raspberry Pi 上に Julius 音声認識ソフトウェア(Julius 本体、ディクテーションキット、グラマーキット)を実装した[2]。

また、認識したい言葉を単語辞書に登録し、ディクテーションファイル(口述筆記)を生成した後、Julius ディクテーション実行キットを使用して Julius 音声認識を実行した。Python スクリプト上で、モジュールモードの Julius サーバーをバックグラウンドで動作させ、`socket` モジュールを用いて Julius からの認識結果を取得できるようにした。Julius サーバーの受信データから RECOGOUT キーを XML として随時パースし、単語辞書中の言葉を判別して、各々の言葉に対する動作(サーボ動作、LED 制御、音声合成)を行っている。

3. OpenCV による画像認識

OpenCV はオープンソースの画像処理のライブラリである[4]。RaspberryPi へ USB 接続した Web カメラより取り込んだ画像データから、OpenCV のライブラリを用いて認識を行う。本稿では人の顔を検出するために顔認識機能を使用する(図 3)。更に合成音声や音声認識の連動を行うプログラムを作成し、OpenCV で顔を検出した場合にマグボットが挨拶を行う機能を追加した。

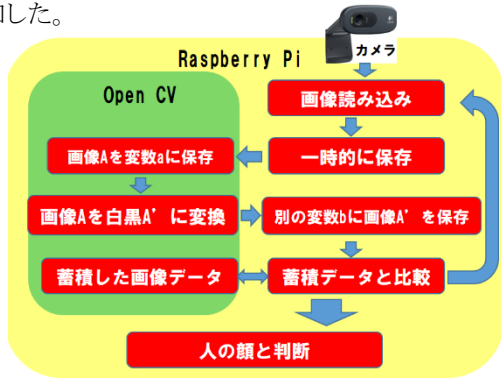


図 3: OpenCV の動作構成図

4. 移動機構

4.1. 移動機構本体について

マグボットの操作項目の中に移動オプションを追加し、遠隔で移動機構を操作することが可能である。移動機構本体は移動用のキャタピラ「Tri-Track Chassis Kit (Lynxmotion 社)」(図 4)、モータ× 2 (HSIANG NENG 社)、モータドライバ「Sabertooth dual 5A motor driver for R/C (DimensionEngineering 社)」で構成した。モータドライバの制御はマグボットの Arduino から行う。



図 4: 本稿の移動機構

4.2. 移動機構の動作について

Arduino の PWM を用いてモータドライバを制御する。PWM ではパルス幅を変化させることで出力電力を制御している(表 1)。PWM の値は、0~255 の範囲で各モータへ出力する。また PWM によって「前進」「後進」「右旋回」「左旋回」の 5 パターンの動作を行うことが可能である。さらにウェブ上のスライダを操作することで速度調整も可能である。

表 1: モータドライバを制御する各 PWM の値

動作	制御信号 1	制御信号 2
前進	120~170	170
後進	200~250	170
右旋回	180	200~250
左旋回	180	120~160
停止	0	0

4.3. 音声認識との連動について

音声認識による移動機構の操作を追加した。音声認識には Julius を利用し、予め単語辞書に単語を登録し、その単語の音声を検出させる。認識単語として「前進」「停止」「後進」「右旋回」「左旋回」を追加し、音声による移動機構を操作可能にした。音声認識から動作までの遅延が若干見受けられるが、実際に動作することを確認した。

5. システムの動作例

5.1. 画像認識の動作例

動作例として、まず Web カメラを使用し画像認識を行う動作について説明する。まず Raspberry Pi 内の HTTP サーバーにアクセスし、HTML ファイルを Web ブラウザに表示する。次に操作画面の「撮影」ボタンをクリックすることにより、Web カメラの動画が表示され、画像認識を開始する。認識する顔は、両目と口の 3 点と、顔の輪郭によって認識する。3 点と輪郭は色の違いで認識される。認識時は、色の明暗で 3 点と輪郭を認識する。

5.2. 移動機構の動作例

動作例として、マグボットの「前進」の動作について説明する。まず画像認識同様、操作画面の「前進」ボタンをクリックすることにより、WebSocket 通信を利用して Raspberry Pi に“@a”の文字列を送る。“@”から始まる文字列は Arduino の制御(動作させるコマンド)として認識させているので、シリアル通信を利用して Arduino に送り、対応するキャタピラに相当するモータを動作させることで「前進」の動作を実現する(図 5)。

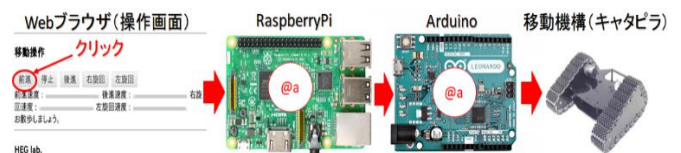


図 5: 移動機構のシステム動作例

6. 今後の予定

本稿では Web カメラによる認識機能とキャタピラによる移動機構を新たに実装した。今後はカメラ機能と移動機構の連動を図り、マグボットが移動しても常に人物の顔領域をトラッキングできるような仕組みを導入したいと考える。

参考文献

[1]小池星多,“おしゃべりロボット「マグボット」ラズパイと Arduino で電子工作”,リックテレコム, 2016 年
 [2]“Raspberry Pi と Arduino を用いたソーシャルロボットに関する研究”,第 9 回大学コンソーシアム八王子学生発表会, P001
 [3]“大語彙連続音声認識エンジン Julius”,
<http://julius.osdn.jp/>
 [4]北山洋幸,“OpenCV で始める簡単動画プログラミング”,カットシステム, 2010 年