

# Designing of Humanoid Robot with Voice Recognition Capability

Widodo Budiharto<sup>a,\*</sup>, Alexander Agung S.G<sup>a</sup>, Azani Cempaka Sari<sup>a</sup>, Heri Ngarianto<sup>a</sup>

<sup>a</sup>School of Computer Science, Bina Nusantara University, Jakarta-Indonesia

\*Corresponding Author: wbudiharto@binus.edu

## Abstract

Intelligent Humanoid Robots for education and entertainment in uncontrolled environments need to be based on vision and voice recognition. This paper propose a high speed embedded system based on Raspberry Pi 2 and voice recognition method for Humanoid robot, because the ability to accurately recognize commands is important feature for education and entertainment. An audio module PiAudio from Python is used in this system. The proposed method is able recognize commands and face recognition. We evaluate and present the performance of the system.

**Keywords:** humanoid, voice recognition, tracking, education

## 1. Introduction

The humanoid robots are popular nowadays for education and entertainment. The important features of humanoid robot, such as accuracy, robustness and recognize commands from user; has proven to be a challenging subset of this research area. With the evolution of robotics hardware and subsequent advances in processor performance in recent years, the temporal and spatial complexity of feature extraction algorithms to solve this task has grown accordingly [1].

In the case of Humanoid robot for education, vision systems are one of the main sources for environment interpretation. Many problems have to be solved such as voice and face recognition. First of all, the robot has to get information from the environment, mainly using the camera. The robot has to self-localize and decide the next action: move, walking, search another object, etc. It makes

no sense within this environment to have a good localization method if that takes several seconds to compute the robot position or to decide the next movement in few seconds based on the old perceptions [2]. At the same time many other topics like human-machine interaction, robot cooperation and mission and behavior control give humanoid robot a higher level of complexity like no any other robots [3]. So the high speed processor with efficient algorithms is needed for this issue.

One of the performance factors of a humanoid robot is that it is highly dependent on its tracking ball and motion ability. The vision module collects information that will be the input for the reasoning module that involves the development of behavior control. Complexity of humanoid robot makes necessary playing with the development of complex behaviors, for example situations of coordination or differ rent role assignment during the match. There are many types of behavior control, each with advantages and disadvantages: reactive control is the simplest way to make the robot play, but do not permit more elaborated strategies as explained for example in [4].

In this paper we propose a low cost humanoid robot compared with the well-known humanoid robots for education such as DarwIn-OP and NAO and test its ability for voice recognition and face recognition based on OpenCV.

## 2. Proposed Method

### 2.1 Architecture

Humanoid robot design based on the vision involves the need to obtain a mechanical structure with a human appearance, in order to operate into a human real world. AT

previous version we use Odroid as main controller and we migrate to Raspberry pi 2 Model B for this research. The specification of Raspberry pi 2 Model B as shown in fig.1(a) is [6]:

- Broadcom BCM2836 ARMv7 Quad Core Processor powered Single Board Computer running at 900MHz
- 1GB RAM so you can now run bigger and more powerful applications
- Identical board layout and footprint as the Model B+, so all cases and 3rd party add-on boards designed for the Model B+ will be fully compatible.
- 40pin extended GPIO
- 10/100 Ethernet Port to quickly connect the Raspberry Pi to the Internet

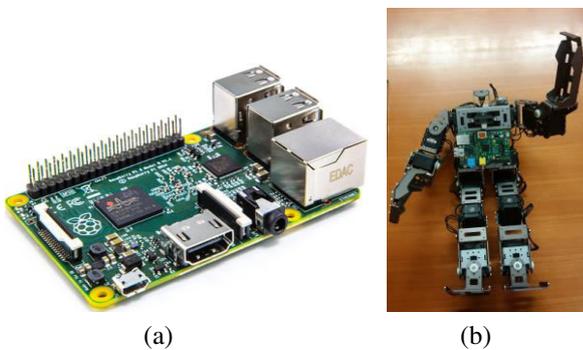


Fig.1. Raspberry Pi 2 Model B and our prototype of humanoid robot (b)

We propose an embedded system with audio input for voice and able to handle high speed image processing for track and kick a ball, so we use main controller based on the ATMEL Processor. Camera and servo controller used to track an object or face, and the output of the main controller will communicate with the CM530 controller to control the actuators and sensors of the robot as shown in fig. 2.

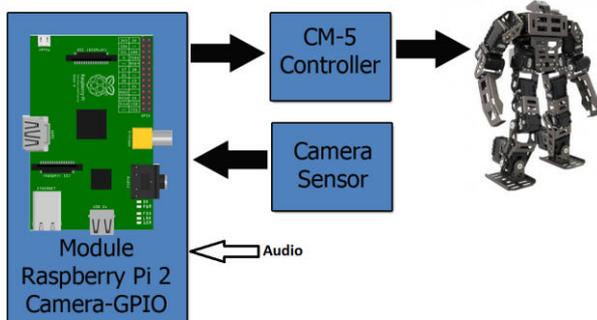


Fig.2. Architecture of our humanoid robot for voice recognition and object/face tracking using camera.

Computer vision for robot is one of the most challenging applications in sensor systems since the signal is complex from spatial and logical point of view[5]. An active camera tracking system for humanoid robot tracks an object of interest (ball) automatically with a pan-tilt camera. At previous work, we detect ball based on color (color-based object detector) which is not robust [7].

## 2.2 Voice Recognition Module

We use PyAudio and Speech recognition 2.11 based on Python for programming the Raspberry Pi 2, and Google modules for translator. This module is able to recognize commands from user. PyAudio provides Python bindings for PortAudio, the cross-platform audio I/O library. With PyAudio, you can easily use Python to play and record audio on a variety of platforms. Example:

```

importspeech_recognition as sr
r = sr.Recognizer()
# use the default microphone as the audio source
withsr.Microphone() as source:
audio = r.listen(source)          # listen for the first
phrase and extract it into audio data
try:
# recognize speech using Google Speech Recognition
print("You said " + r.recognize(audio))
exceptLookupError:
# speech is unintelligible
print(" Robot could not understand audio")

```

## 2.3 Track and Kick Ball algorithm

We use OpenCV for convert to HSV (Hue Saturation-Value), extract Hue & Saturation and create a mask matching only the selected range of hue value [9]. OpenCV with Machine Learning class for SIFT and FLANN methods run very well in the Raspberry's board. To have a good estimation, the object must be in the center of the image, i.e. it must be tracked. Once there, the distance and orientation are calculated, according to the neck's origin position, the current neck's servomotors position and the position of the camera respect to the origin resulting of the design [8]. The ball will be track based on the color and webcam will track to adjust the position of the ball to the center of the screen based on the Algorithm 1.

### Algorithm 1: Ball Detection, Track and Kick the ball[8]

```
Get input image from the camera
Detect ball using SIFT Keypoint detector
If detected then
    Get the center position of the ball
    Centering the position of the ball
    Move robot to the ball
    If ball at the nearest position with the robot then
        Kick the ball
endif
endif
```

### 3. Experimental Results

The approach proposed in this paper was implemented and tested on a humanoid Robot named Bee Humanoid Ver 3.0 based on Bioloid GP and OpenCV in Python. For test case I, when a ball in front of the robot and detected, the robot try to track the ball, and if the ball at the nearest position with the robot, robot will kick it as shown in fig. 3.

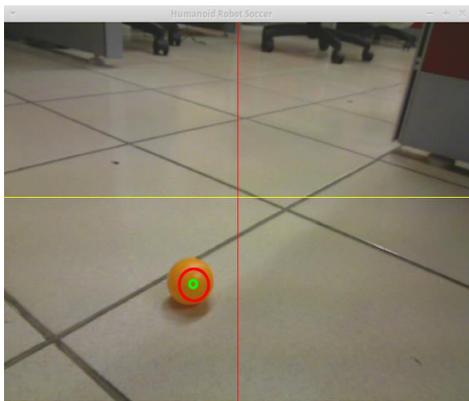


Fig. 3. The ball detected by our vision-based system for humanoid robot.



Fig. 4. The robot tracks and kicks a ball when at the correct position.

The overall result of our system shown in table 1 below:

Parameter	status
Detect a ball	10ms
Tracking object without error	90%
Kick a ball	> 80%

For a test Case 2, the robot able to recognize commands such as Start, speak, stop, move, dancing etc.

### 4. Conclusions

In this paper, we introduced the hardware architecture implemented on our humanoid robot. They are based on Raspberry Pi 2 that has powerful ability for high speed image processing. We propose robust system using PyAudio for voice recognition and SIFT keypoint detector to detect and track a ball, then kick the ball after getting the nearest position of the robot with the ball. The FLANN based matcher suitable to be used for real situation. For future work, we want improve the ability voice recognition with multiple sources of sound.

### Acknowledgment

We say thanks to Indonesian government for granting Hibah Ristek 2015 for this research.

### References

- (1) I. Ha, Y. Tamura, Y., H. Asama: Development of open humanoid platform DARwIn-OP. In: SICE Annual Conference (SICE), 2011 Proceedings of. pp. 2178–2181. IEEE, 2011.
- (2) F. Martin, C. Aguero, J. Maria Canas and E. Perdices, Humanoid. Soccer Player Design, Robot Soccer, VladanPapi (Ed.), ISBN: 978-953-307-036-0, InTech, 2011.
- (3) F. Blanes, Embedded Distributed Vision System for Humanoid Soccer Robot, Journal of Physical Agents, vol. 5(1), 2011
- (4) S. Behnke and R. Rojas. A hierarchy of reactive behaviors handles complexity. In Balancing Reactivity and Social Deliberation in Multi Agent Systems, volume 2103 of LNAI, pages 125-136. Springer, 2001.

- (5) R. Szeliski, Computer vision: algorithms and applications. Springer-Verlag New York Inc, 2010.
- (6) <https://www.raspberrypi.org/products/raspberry-pi-2-model-b/>
- (7) T. Bräunl, Embedded Robotics, Mobile Robot Design and Applications with Embedded System, Springer, 2008
- (8) W. Budiharto, B. Kanigoro, Viska Noviantri, High Speed Embedded System for Ball Tracking of Humanoid Soccer Robot, ISRSC 2013 Semarang, 2013.
- (9) <https://pypi.python.org/pypi/PyAudio>