ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Humanoid Shadow Robot Using Kinect Technology

Pranjal Katara [*1], Ratul Aggarwal [2]
[*1,2] B.Tech , School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
pranjal.katara2011@vit.ac.in

Abstract

The Shadow bot is a basic humanoid with 19 degrees of freedom. At first site, it appears to be just like any other humanoid robot but the real difference comes in when we go inside its working mechanism. Shadow bot works on the principle of wireless human-robot interface. It will take user's body motion as an input and move just like him/her. It's more like giving sight to the robot and it'll mimicries user's motion. Unlike any other humanoid robot, the controller in this bot's architecture is the user himself. Manipulators can direct communicate with the computer just relying on body movements, and thus control the actions of remote robot in real time, coordination, and harmonization. The design is a combination of new human motion capture methods, intelligent agent control method, network transmission, multiple sensor fusion technology. The name 'Shadow bot' is used because just like our shadow, it'll move precisely in the same fashion as we are. This paper briefly describes all the segments of the humanoid robot using Kinect technology.

Keywords: Shadow bot, humanoid, kinect, wireless human-robot interface.

Introduction

The Shadow bot is a basic humanoid (as shown in figure 1) with 19 degrees of freedom, 39 cms height, 1.65 kgs weight and made up of light aluminium. At first site, it appears to be just like any other humanoid robot but the real difference comes in when we go inside its working mechanism. Shadow bot works on the principle of wireless human-robot interface. It will take user's body motion as an input and move just like him/her. It's more like giving sight to the robot and it'll mimicries user's motion. Unlike any other humanoid robot, the controller in this bot's architecture is the user himself. Manipulators can direct communicate with the computer just relying on body movements, and thus control the actions of remote robot in real time, coordination, and harmonization. The design is a combination of new human motion capture methods, intelligent agent control method, network transmission, multiple sensor fusion technology. The name 'Shadow bot' is used because just like our shadow, it'll move precisely in the same fashion as we are.

The principle of Human-Machine interaction to humanoid robot mainly includes the four types of machinery, acoustics, electromagnetism and optics. To achieve the sight function of the bot, we are using Microsoft's 'Kinect Technology' developed with PrimeSense at its heart. This method not only makes people no longer wear all kinds of sensors, but also decline costs of equipment. Kinect is a new Controller technology introduced by Microsoft in 2010(1). Kinect has the ability to track movements and voices, and even identify faces without need of any additional devices (2).

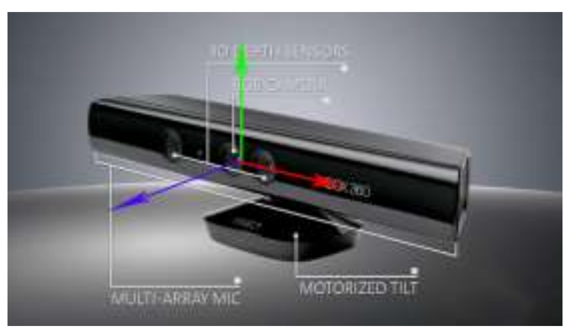

Figure 1: The Kinect device by Microsoft with the Kinect or camera reference frame. The z–axis is pointing out of the camera (courtesy of Microsoft).

This paper presents an attempt to use Kinect technology in designing humanoid shadow robot with 19 degrees of freedom.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

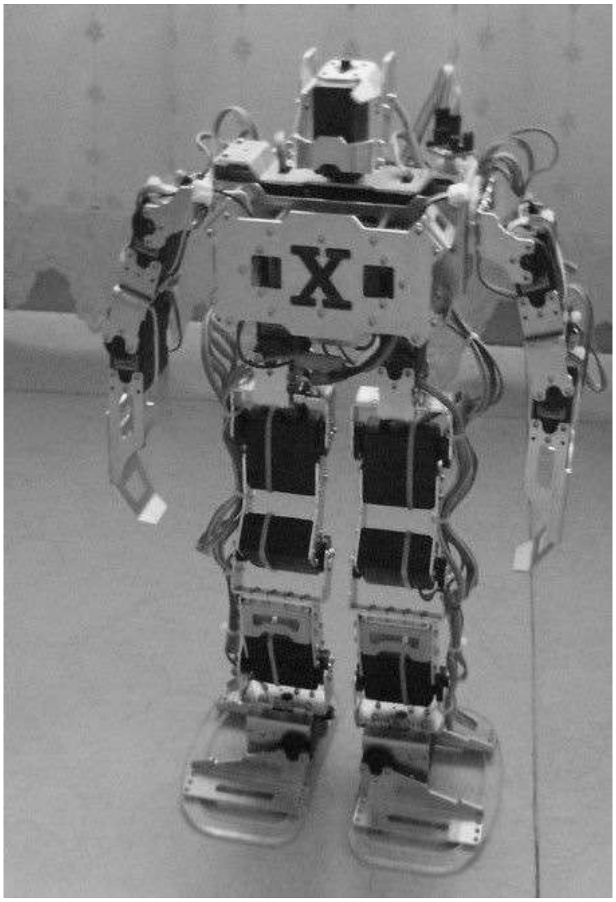

Figure 2: Basic Structure

Kinect

To recognize body gesture and motion, it makes use of depth mapping and then infer the body position. For making a depth map, structured light is used which is projected using Infrared Laser Light Projector at 830 nm, and the reflection is captured by Infrared CMOS camera with matching bandpass filter. This light has a specific pattern and any change in the reflected pattern is be detected. This can be done under any ambient light condition and can detect up to 20 joints of human body, thus making it an extremely robust technique.

A smart chip named PrimeSense (PS1080), which processes all the data from the IR camera and generate the map, calculates the depth map.

A. PrimeSense

The depth sensor technology was developed by Israeli *Prime Sense*.(3) The depth recognition mechanism of PS1080 is very simple but effective, and is based on triangulation like a standard stereo camera. [4]

The stereo cameras take two views of the same thing from different angles, find the difference between them, and extract depth by triangulation. But occlusion, repeated patterns, and lack of textures make it difficult to find the difference between two views.

To counter this problem, fixed dot pattern/texture is used in laser. This texture is carefully designed to present unique patterns along the epipolar lines to enforce that the correspondence can be found simply by correlation. This approach is known as projected texture. [5]

B. Kinect Depth Flow

The main question arises here is, How to get the depth of the pattern? The solution to this is very intelligent. To solve this, a technique called *depth from focus*. The dot texture consists of circular dots. When projector projects the pattern, the circular dots become ellipse depending upon the depth of the area. The cameras have special *astigmatic lenses* with different focal lengths in x-axis and y-axis, which improves the accuracy of capturing the reflected pattern from the surface. This makes the edges of surface closer to the projector look sharper as compared to the edges of far surfaces. This helps in detection of depth and its mapping.

From the PrimeSense native drivers about Kinect, a reference image is pre stored And its size is 1280x1024. And the speed of the IR sensor is coherent with 1280x1024 at 30fps. This clearly suggests that the depth extraction algorithm takes 1280x1024 images. However, Kinect provides depth extraction at 632x480 with 3 bits of subpixel precision. Usually subpixel precision is obtained using quadratic interpolation, but this is expensive in hardware. Block matching can be efficiently implemented instead, so we suggest that the internal reference image is upsampled at 5120x1024 internally and the SAD is done on an assimetrical window, where 17x17 pixels from the IR image are matched against 68x17 pixels of the reference. [6]

OPENNI

Open Natural Interaction, the term *Natural Interaction* points towards the concept of Human-machine interaction modeled on human senses, mostly focused on sight & speech. This module includes:

- Voice and command recognition, where the target machine receives vocal instructions.
- Hand gestures, where the pre-programmed hand gestures are used to control the machine's activity. Like activating and deactivating certain abilities, etc.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

- Body motion, where full body movement is tracked, analyzed and interpreted for controlling the machine's motion.

OpenNI is a framework, which provides its Application Programming Interfaces' (API) for building applications utilizing Natural Interaction. The OpenNI standard API enables natural-interaction application developers to track real-life (3D) scenes by utilizing data types that are calculated from the input of a sensor (for example, representation of a full body, representation of a hand location, an array of the pixels in a depth map and so on). Applications can be written regardless of the sensor or middleware providers. [7]

A. NiTE (OpenNI module)

OpenNI module NiTE is used for skeleton tracking, to track the motion of the user. In OpenNI, human body is made up of joints, and each joint have the information about location and orientation of the human. So, skeleton mapping is thus very important. In dynamic gesture track recognition, gesture recognition rate has always been a problem, Because the traditional hand gesture recognition in how to separate the hand type from the surrounding environment, and then determine the palm position has been one of the difficult problems. A Kinect based dynamic gesture trajectory identification method, use OpenNI where palm is defined as a Kinect node, quickly and accurately get the palm position coordinates, and then use the HMM technology to do gesture model training to improve the recognition rate. (8)

B. Gain Control

The game starts when the controlling human, what we will call a User, gain control over the device. This is offered in two by this module, 'click' and 'wave'.

These are basic gestures user will perform in front of the sensor to get its focus. And once a user gains the control, no other person can get it until the user releases the control (which generally happens when the user leaves the field-of-view for more then 10 seconds). To be brief, 'Click' gesture is like raising your hands up (till your eye level, with a finger clearly visible to the sensor), and swings them to and fro (with clear large forward and backward motion, 20 cm at least). And 'Wave' gesture is like raising one of your hands straight up and move it from left to right (5 times to complete the gesture). This motion completion is important because it distinguishes the control-gesture from rest.

This control-gesture recognition can potentially be misjudged with other hand movements, so to distinguish these two gestures from rest of them is how their recognition starts. Click gesture recognition starts when both the hands does a large enough motion in forward direction. While wave gesture recognition starts when one controlling hand is twice in left-to-right direction.

Now, to keep the device in your control, keep your controlling hand away from other objects, which includes your body too. This is because doing this will lead to hand-point misplacement (on the object which came in between your hand and sensor). So, obviously, user is not supposed to touch anything while controlling.

C. Sticking to User

There can be many people in front of the sensor. To distinguish the user from all of them, each one in the scene is assigned with a unique and persistent ID based on the label map. And based on this label map, the skeleton of user will build-up and skeleton tracking will begin. So, it is clear that if any problem occurs in label mapping algorithm, then the whole process will malfunction and will lead to an inaccurate orientation of the skeleton. The error can come with few more things like

- User moving too close to the objects in the scene, example walls.
- Moving the sensor while user segmentation is active.
- When two users are moving close to each other.

D. Skeleton Tracking

Skeleton of the user is created to track the motion of the joints of the user's body and control the device. To create the skeleton, user has to come in the field-of-view of the sensors, and that's it. The algorithm of label mapping will compile as soon as the user enters the view field and ID will get assigned to him. And as soon as the control-gesture completes, the algorithm of skeleton making makes a skeleton of the user's body. There is function for auto-calibration, which makes the skeleton more accurate as the time passes.(9)

Auto-calibration works when most of the user's body is visible to sensor, means the user should be standing instead of sitting, with both hands out, and no fast movements.

This tracking generates a Confidence code according to which

1 = surety of recognition, and 0 = recognition fail. Also, mid values like 0.5 can occur in case of joints, when the skeleton heuristics is enabled and heuristic adjusted their position.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

Wireless network

Wireless device uses new IEEE 802.11n standards, which has significantly increased data rate to 600Mb/sec with the use of four spatial stream of channel width of 40MHz [10].

Microprocessor

BeagleBone Black is used for receiving commands from the Wireless Network Card and sending them to the required sections of the robot. This is done by the set of program running on the microprocessor and will decode the binary command for necessary actions.

Conclusion

This paper briefly describes all the segments of the complete architecture of Humanoid Shadow Bot using Kinect. The detailed description of Kinect, its building components, and the functions of algorithm behind will help in understanding the working mechanism of the system. The wireless network is placed to eliminate the need of camera and projector on the robot. This will make the robot more robust, efficient, and controllable from a single position. We are convinced that the preliminary result of this paper will open new horizons to the use of kinect technology in the designing of a humanoid shadow robot.

References

1. *Wikipedia. Kinect — Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Kinect, 2011.*
2. *J. Giles. Inside the race to hack the Kinect. The New Scientist,208(2789):22–23, 2010.*
3. *Prime Sense. Prime sense kinect's hardware. http://www.primesense.com, 2010*
4. *Weibo Song, Xianjiu Guo, Fengjiao Jiang, Song Yang, Guoxing Jiang, Yunfeng Shi, Teleoperation Humanoid Robot Control System Based on Kinect Sensor, 4th International Conference on Intelligent Human-Machine Systems and Cybernetics, vol. 2, 2012, pp. 264-267.*
5. *http://irl.postech.ac.kr/class/mr_2013/Kinect_sensor/references/kinect.pdf*
6. *Manuel Martinez and Rainer Stiefelhagen, Kinect Unleashed: Getting Control over High Resolution Depth Maps, IAPR Conference on Machine Vision Applications, May 2013.*
7. *Jae-Han Park, Yong-Deuk Shin, Ji-Hun Bae, and Moon- Hong Baeg, Spatial uncertainty model for visual features using a kinect sensor, Sensors, 2012.*
8. *Youwen Wang Cheng Yang ; Xiaoyu Wu ; Shengmiao Xu ; Hui Li, Coll. of Inf. Eng., Commun. Univ. of China, Beijing, China Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference Nanchang, Jiangxi (Volume:1) 26-27 Aug. 2012 Page(s):274 – 279, Publisher:IEEE*
9. *http://www.openni.org/wp-content/uploads/2013/02/NITE-Algorithms.pdf*
   David Fiedler and Heinrich Mu ̈ller, Impact of thermal and environmental conditions on the kinect sensor, International Workshop on Depth Image Analysis at the 21st International Conference on Pattern Recognition, 2012.