Form 3

# Abstract of Thesis

| Name （Amalia Istiqlali Adiba） | |
|---|---|
| Title | Sharing Cognition Between Human and an Intelligent Machine, Based on Deep Learning System -"See What I See"-<br>(ディープラーニングを用いた人間とロボットの間の認識共有) |


## Abstract of Thesis

The main focus of this system is for supporting a disable people. Regenerative medicine can repair the damage tissues but it can't replace the total function of human's arm. Brain machine interface comes with the idea of Electromyography (EMG) to control the robot. The electrodes information was very limited, and the method require user to manage high concentration for single movement. This study proposed a new concept of Brain Machine Interface. The concept's name is "see what I see" system. In this thesis, three topics are discussed to explain the detail of "See what I See" system as an interface for sharing human's cognition. At first, a wearable Gaze Tracking (GT) was made to estimate gaze position in 2D coordinate space. Commercial GT is readily available, but they are usually fabricated at the same size. A three-dimensional (3-D)-printable frame and an open-source architecture was made to fabricate a wearable GT with low-cost configuration and reasonable performance. The output of this GT stays very steady at 75 cm or more with an accuracy of 2.58° . This GT has a 24-Hz sampling rate, which can analyze human interest points in real time. The output of the GT was used to discriminate the object. As a result, the system has a 7-Hz sampling rate and a 3.8% classification error. In this system, the object was detected under solid color background in order to simplify the environmental setup. Three dimensions (3D) object detection was developed to make the system work in a real environment. The model uses Convolution Neural Network (CNN) on four channels formed using RGBD information (splitting into R, G, B, Depth in separate channels). The Multichannel CNNs decomposes the task into five layers: two sets of convolution-pooling pairs and a fully connected-output layer. The filters are taken from random patches of the images in an unsupervised way by using k-means clustering. The learned filters are fed into a convolution layer. Each convolution layer is followed by a pooling layer, to reducing the resolution of the feature map, and reducing the sensitivity of the output to shifts and distortions. In the end, fully-connected layers can be used as a classifier with a feedforward based process to classify the household objects. The evaluation results show that the combination of RGB and depth improve the accuracy of object recognition. This integrated framework is compared with related architecture such as combination CNN and recursive neural networks (RNNs) in same RGB-D dataset. The result shows that the Multichannel CNN model obtains 91 % accurate in 3D object detection. Gaze tracking is the important tool for "see what I see" system to identify the 3D object of the person's attention in the visual world coordinate. The wearable GT requires calibration of scene geometry and camera to make it applicable in visual world coordinates. By using a depth sensor camera, a non-intrusive GT was developed to obtain robust and accurate gaze vector in 3D space coordinates. The eye-in-head gaze direction information is obtained by training the visual data from eye image with a Neural Network model. The model uses Convolution Neural Network (CNN) that also was used for 3D object detection. The gaze vector is reconstructed from a set of head and eye pose orientation. The result of this approach reports that the gaze estimation error is 5 degrees. The output of this GT was combined with the 3D object detector to automatically discriminate an interest object. As a result, the system has 80% accuracy in interest object detection. Moreover, the system also giving the object's position in 3D space which is important information for sharing human's cognition.

様式7

# 論文審査の結果の要旨及び担当者

| 氏　名 | （Amalia Istiqlali Adiba） | | |
|---|---|---|---|
| 論文審査担当者 | | （職） | 氏　　名 |
| | 主 査<br>副 査<br>副 査 | 教　授<br>教　授<br>教　授 | 三宅 淳<br>和田成生<br>出口真次 |

論 文 審 査 の 結 果 の 要 旨

上記論文については、平成 28 年 2 月 13 日に大阪大学大学院・基礎工学研究科において審査を行った。

本研究では、3D 空間内において、人間の意図を共有するためのインタフェースシステムの概念を提案した。このシステムは、3D 空間注視追跡と 3D 物体追跡を組み合わせて、実際の視覚世界における関心対象を検出するためのものである。

視線トラッカーを自作して視線を計測し、 SOINN を用いてオブジェクト分類を行ったところ、オブジェクトの 67.2%を正しく予測できることを示された。システムの性能を向上させるために、3D 視線追跡装置と深い畳み込みニューラルネットワーク（CNN）モデル適用したところ、3D データの RGB と奥行き画像との組み合わせが、オブジェクト分類性能を改善し、人間の環境に適した 3D 物体を認識することができることが明らかとなった。さらに進めてフリーヘッド状態での 3 次元空間座標における視線ベクトルを推定するシステムを開発した。奥行きセンサカメラとコンボリューションネットワークアーキテクチャを使用して開発した。評価結果は、頭部姿勢の制約下で視線追跡が高い精度を有することが示された。様々な姿勢や位置の変化のもとでも、対象物を正確に 80%検出が可能となった。

発表に対して主査副査との間で質疑応答が行われた。質問に対する回答を含め、当該研究論文は、科学研究として高い質の実験を実施し、高度レベルでまとめられたものと判断されたことから、博士（工学）の学位論文として価値のあるものと認める。