

2016

# Development of Markerless Augmented Reality System Based on Machine Learning

Sittidetchanchai Chanayot

(Student ID Number : 81434596)

Supervisor Professor Ogi Tetsuro

September 2016

Graduate School of System Design and Management,  
Keio University  
Major in System Design and Management

## SUMMARY OF MASTER'S DISSERTATION

Student Identification Number	81434596	Name	Sittidetchanchai Chanayot
Title: Development of Markerless Augmented Reality System Based on Machine Learning			
<p>Abstract</p> <p>Creating a markerless system for augmented reality is quite a difficult task. Current Augmented Reality (AR) systems relies on the controlled variables of markers to present information, without these markers the AR system lacks the input with which it would show the output (information) to the user. This research aims to investigate two methods of successfully developing a markerless AR system. These methods are highly inspired by allowing a user to have the ability to receive information via an AR program on a mobile phone without the use of markers. The simplistic method is to use multiple photos to create various angled picture markers of a single object in order to make a 3D marker of that object. This 3D marker can be viewed from multiple perspectives yet still show the same result as a singular marker. The only requirements for this would be a mobile phone, a computer, and a program for creating (In this study, Vuforia was utilized) an object, making this an easy to access method for many people. In the alternative methods, AI technology of machine learning was applied to recognize an object instead of a marker for AR systems. Particularly, in this study, visual recognition function in IBM Bluemix was used for the machine learning. By connecting the mobile device to the cloud computer of IBM Bluemix, the heavy calculation load of machine learning can be performed and the AR system based on the machine learning can be constructed. This paper will discuss about various elements affecting the creation of the picture marker bases AR system and the machine learning based AR system, also presents tests on the recognition percentage of AR systems and a demo program</p>			
<p>Key Word(5 words)</p> <p>Augmented reality, Machine learning, Image processing, Markerless System, Cloud system</p>			

## Table of Content

<b>CHAPTER 1 INTRODUCTION .....</b>	<b>10</b>
1.1 INTRODUCTION.....	10
1.2 PROBLEM.....	12
1.3 LITERATURE REVIEW.....	13
1.3.1 <i>General Concept of AR Markers</i> .....	13
1.3.2 <i>Visual Recognition Architecture</i> .....	15
1.3.3 <i>Current research efforts</i> .....	16
1.3.4 <i>Academic Contribution of Study</i> .....	17
<b>CHAPTER 2 CONCEPTS AND PROPOSE .....</b>	<b>18</b>
2.1 CONCEPT.....	18
2.1.1 <i>Multiple picture marker</i> .....	18
2.1.2 <i>Machine learning</i> .....	18
2.1.3 <i>Difference between Multiple picture marker and Machine learning</i> .....	19
<b>CHAPTER 3 IMAGE MATCHING METHOD .....</b>	<b>20</b>
3 VUFORIA EXPERIMENT .....	20
3.1 SYSTEM DIAGRAM.....	20
3.1.1 <i>Storing Phrase</i> .....	20
3.1.2 <i>Scanning Phrase</i> .....	20
3.2 NUTELLA.....	21
3.2.1 <i>Calculation of Detectable Point</i> .....	23
3.2.2 <i>Effect of surface pattern</i> .....	27
3.3 STATUE .....	29
3.3.1 <i>High focal Point</i> .....	29
3.3.2 <i>Low Focal Point</i> .....	31
3.4 EFFECT OF BACKGROUND.....	33
3.4.1 <i>Effect of Background on High focal point</i> .....	33
3.4.2 <i>Effect of Background on Low Focal Point</i> .....	36
3.5 <i>Effect of focal point</i> .....	40
3.6 <i>Effect of backlit image</i> .....	44
3.7 <i>Effect of Saturation</i> .....	45
3.8 PROBLEM OF THE METHOD .....	47
3.9 DISCUSSION .....	48
<b>CHAPTER 4 MACHINE LEARNING METHOD .....</b>	<b>50</b>
4 VISUAL RECOGNITION BLUEMIX.....	50
4.1 WHAT IS VISUAL RECOGNITION BLUEMIX (IBM WATSON, 2016).....	50
4.2 THE PURPOSE OF USING VISUAL RECOGNITION BLUEMIX.....	51
4.3 SYSTEM DIAGRAM .....	52
4.3.1 <i>Learning Phrase</i> .....	52
4.3.2 <i>Scanning Phrase</i> .....	53
4.4 <i>Confidence score</i> .....	54
4.4.1 <i>Learning phase</i> .....	54
4.4.2 <i>Testing</i> .....	54
4.5 VARIABLE AND FORMULA.....	56
4.7 STATUE .....	58
4.7.1 <i>Times Affects Confidence score</i> .....	58
4.7.2 <i>Side Effect score</i> .....	68
4.7.3 <i>Effect of lens on confidence score</i> .....	75
4.8 NUTELLA.....	79
4.9 DISCUSSION .....	81
<b>CHAPTER 5 DISCUSSIONS .....</b>	<b>83</b>

5.1 PROTOTYPE SYSTEM .....	83
5.1.1 Webprogram.....	83
5.1.2 Mobile App.....	85
5.2 LIMITATION AND FUTURE WORK .....	87
5.3 VALIDATION.....	88
5.3.1 Description.....	88
5.3.2 Questionnaire.....	89
5.3.3 Questionnaire Result.....	90
5.4 DIFFERENCE BETWEEN IMAGE MATCHING METHOD AND MACHINE LEARNING METHOD .....	92
5.5 BENEFITS AND ADVANTAGES .....	93
5.5.1 Image Matching Method.....	93
5.5.2 Machine learning Method.....	93
<b>CHAPTER 6 CONCLUSION .....</b>	<b>94</b>
6 CONCLUSION.....	94
<b>CHAPTER 7 REFERENCE AND ACKNOWLEDGEMENT .....</b>	<b>97</b>
7.1 REFERENCE .....	97
7.2 ACKNOWLEDGEMENT.....	99
<b>CHAPTER 8 APPENDIX.....</b>	<b>100</b>
8.1 BASED ON TIME.....	100
8.1.1 Time 6:00.....	100
8.1.2 Time 7:00.....	101
8.1.3 Time 8:00.....	102
8.1.4 Time 9:00.....	103
8.1.5 Time 10:00.....	104
8.1.6 Time 11:00.....	105
8.1.7 Time 12:00.....	106
8.1.8 Time 13:00.....	107
8.1.9 Time 14:00.....	108
8.1.10 Time 15:00.....	109
8.1.11 Time 16:00.....	110
8.1.12 Time 17:00.....	111
8.1.13 Time 18:00.....	112
8.1.14 Time 19:00.....	113
8.1.15 Time 20:00 without flash.....	114
8.1.16 Time 20:00 with flash.....	115
8.2 BASED ON SIDE .....	116
8.2.1 Camera .....	116
8.2.2 Zoom Lens.....	125
8.2.3 Phone Camera .....	134

## Table of Figures

Figure 1: Showing various type of AR marker.....	14
Figure 2 : System Diagram for Storing Phrase.....	20
Figure 3 : System Diagram for Scanning Phrase.....	20
Figure 3: Ten pictures taken of Nutella container in controlled environment.....	21
Figure 4: The ten images post-processed by the Vuforia program. ....	22
Figure 5: An enlarged portion of the previously mentioned Fukuzawa Bronze Statue, showing detectable points represented by yellow crosses.....	23
Figure 6: Graph of the number of detectable pixels on each side of the Nutella container.....	25
Figure 7: The number of detectable points on each side of Nutella. ....	25
Figure 8: Comparison of detectable points and pixels on each side of the Nutella container.....	26
Figure 9: The effect of patterns on the surface to the creation of marker.....	27
Figure 10: Comparison of number of pixel between four sides of Nutella.....	28
Figure 11: Picture of Fukuzawa bronze from various angles with high focal point.....	29
Figure 12: The number of pixel and point on each side of Fukuzawa bronze statue with background and high focal point.....	30
Figure 13: The number of pixel and point on each side of Fukuzawa Bornze statue without background in high focal point.....	30
Figure 14: Picture of Fukuzawa bronze from various angles with low focal point.....	31
Figure 15: The number of pixels and points on each side of Fukuzawa Bronze statue with background in low focal point. ....	32
Figure 16: The number of pixels and points on each side of Fukuzawa bronze statue without background in low focal point. ....	32
Figure 17: Processed images of Fukuzawa bronze statue with high focal length from various sides.....	33
Figure 18: Processed images of Fukuzawa bronze statue with high focal length from various sides without background.....	33
Figure 19: The comparison of the detectable points of high focal point between with background and without background.....	35
Figure 20: Processed images of Fukuzawa bronze statue with low focal length from various sides.....	36
Figure 21: Processed images of Fukuzawa bronze statue with high focal length from various sides without background.....	36
Figure 22: The comparisons of the detectable point in low focal point between with background and without background.....	39
Figure 23: The effect of background High focal point on the left and Low focal point on the right. ....	40
Figure 24: The comparison of the detectable point statue picture with background.....	41
Figure 25: The comparisons of detectable point statue picture without background between low and high focal point.....	42
Figure 26: The normal backlit image (right) as well as the augmented backlit image (left) .....	44
Figure 27: The effect of saturation.....	45

Figure 28: The comparison of number of pixels and detectable points between normal image and increased in saturation image. ....	46
Figure 29: The problem of Vuforia .....	47
Figure 30: The system diagram of Learning Phrase.....	52
Figure 31: The system diagram of Scanning Phrase .....	53
Figure 32: Comparing between the Average and Subaverage confidence scores for the camera. ....	60
Figure 33: Comparing the Minimum, Maximum and Minimum Detected image confidence score for the camera .....	60
Figure 34: Comparison between the average and subaverage confidence score for Zoom lens .....	63
Figure 35: Comparison the Maximum, Minimum and Minimum Detect confidence score for Zoom lens.....	63
Figure 36: Comparison between the average and subaverage confidence score for Phone .....	66
Figure 37: Comparison between the Maximum, Minimum and Minimum Detect confidence score for phone. ....	66
Figure 38: Comparison between average confidence score from different sides	69
Figure 39: Comparison between subaverage confidence score from different side .....	70
Figure 40: Comparison of maximum confidence score from different side .....	71
Figure 41: Comparison of minimum confidence score from different side .....	72
Figure 42: Comparison of minimum detect confidence score from different side .....	73
Figure 43: Comparison of number of non-detect image from different side .....	74
Figure 44: Comparison of confidence score at different time from different lens .....	75
Figure 45: Comparison of average, subaverage, maximum, minimum, minimum detect confidence score from different lens.....	77
Figure 46: Comparison of detection percentage from different type of lens .....	78
Figure 47: Comparison of detection percentage from different type of lens .....	79
Figure 48: The first page of the web program. ....	83
Figure 49: The Creating interface of the web program.....	83
Figure 50: The testing interface.....	84
Figure 51: The result of the test.....	84
Figure 52: These pictures show different interface of the mobile app program..	85
Figure 53: The positive and negative result .....	86
Figure 54: Showing the Confidence score at 6:00 .....	100
Figure 55: Showing the Confidence score at 7:00 .....	101
Figure 56: Showing the Confidence score at 8:00 .....	102
Figure 57: Showing the Confidence score at 9:00 .....	103
Figure 58: Showing the Confidence score at 10:00 .....	104
Figure 59: Showing the Confidence score at 11:00 .....	105
Figure 60: Showing the Confidence score at 12:00 .....	106
Figure 61: Showing the Confidence score at 13:00 .....	107
Figure 62: Showing the Confidence score at 14:00 .....	108
Figure 63: Showing the Confidence score at 15:00 .....	109
Figure 64: Showing the Confidence score at 16:00 .....	110
Figure 65: Showing the Confidence score at 17:00 .....	111

Figure 66: Showing the Confidence score at 18:00 .....	112
Figure 67: Showing the Confidence score at 19:00 .....	113
Figure 68: Showing the Confidence score at 20:00 without flash.....	114
Figure 69: Showing the Confidence score at 20:00 with flash.....	115
Figure 70: Showing the Confidence for camera front side .....	118
Figure 71: Showing the Confidence for camera front right side .....	118
Figure 72: Showing the Confidence for camera right side.....	119
Figure 73: Showing the Confidence for camera back right side.....	119
Figure 74: Showing the Confidence for camera back side .....	120
Figure 75: Showing the Confidence for camera back left side .....	120
Figure 76: Showing the Confidence for camera left side.....	121
Figure 77: Showing the Confidence for camera front left side .....	121
Figure 78: Showing the Confidence for camera by average.....	122
Figure 79: Showing the Confidence for camera by Subaverage.....	122
Figure 80: Showing the Confidence for camera by Maximum .....	123
Figure 81: Showing the Confidence for camera by Minimum.....	123
Figure 82: Showing the Confidence for camera by Minimum Detect .....	124
Figure 83: Showing the Confidence for camera Zoom lens front side.....	127
Figure 84: Showing the Confidence for camera Zoom lens front right side .....	127
Figure 85: Showing the Confidence for camera Zoom lens right side .....	128
Figure 86: Showing the Confidence for camera Zoom lens back right side .....	128
Figure 87: Showing the Confidence for camera Zoom lens back side .....	129
Figure 88: Showing the Confidence for camera Zoom lens back left side .....	129
Figure 89: Showing the Confidence for camera Zoom lens left side .....	130
Figure 90: Showing the Confidence for camera Zoom lens front left side.....	130
Figure 91: Showing the Confidence for camera Zoom lens by average .....	131
Figure 92: Showing the Confidence for camera Zoom lens by Subaverage .....	131
Figure 93: Showing the Confidence for camera Zoom lens by Maximum.....	132
Figure 94: Showing the Confidence for camera Zoom lens by Minimum .....	132
Figure 95: Showing the Confidence for camera Zoom lens by Minimum detect	133
Figure 96: Showing the Confidence for Iphone camera front side .....	136
Figure 97: Showing the Confidence for Iphone camera front right side .....	136
Figure 98: Showing the Confidence for Iphone camera right side.....	137
Figure 99: Showing the Confidence for Iphone camera back right side.....	137
Figure 100: Showing the Confidence for Iphone camera back side .....	138
Figure 101: Showing the Confidence for Iphone camera back left side.....	138
Figure 102: Showing the Confidence for Iphone camera left side .....	139
Figure 103: Showing the Confidence for Iphone camera front left side.....	139
Figure 104: Showing the Confidence for Iphone camera by Average.....	140
Figure 105: Showing the Confidence for Iphone camera by Subaverage .....	140
Figure 106: Showing the Confidence for Iphone camera by Maximum.....	141
Figure 107: Showing the Confidence for Iphone camera by Minimum .....	141
Figure 108: Showing the Confidence for Iphone camera by Minimum Detect...	142

## Table of Tables

Table 1: The number of pixels and point on each side of Nutella bottle. ....	24
Table 2: The number of pixels and points on each side of Fukuzawa Bronze statue: high focal point with background in comparison to without background.....	35
Table 3: The number of pixels and points on each side of Fukuzawa bronze statue in Low focal point with background and without background. ....	38
Table 4: The confidence score at different times as well as the average, subaverage, Maximum, Minimum and Minimum Detect for Camera.....	61
Table 5: The confidence score at different time as well as the average, Subaverage, Maximum, Minimum and Minimum Detect for Zoom lens.....	64
Table 6: The confidence score at different time as well as the Average, Subaverage, Maximum, Minimum, and Minimum Detect for Phone.....	67
Table 7: Average confidence score from different side .....	69
Table 8: The Subaverage confidence score from different side .....	70
Table 9: The Maximum confidence score from different side.....	71
Table 10: The Minimum confidence score from different side. ....	72
Table 11: The Minimum Detect confident score from different side.....	73
Table 12: The number of non-detect image from different side.....	74
Table 13: The confidence score at different time from different type of device...	76
Table 14: The confidence score from different type of device as well as the average, subaverage, Maximum, Minimum and Minimum Detect for Phone. ....	77
Table 15: The confidence score of Nutella with different background.....	79
Table 16: Result from ten participant .....	90
Table 17: Showing the confidence score of three type of lens and device at 6:00. ....	100
Table 18: Showing the confidence score of three type of lens and device at 7:00. ....	101
Table 19: Showing the confidence score of three type of lens and device at 8:00. ....	102
Table 20: Showing the confidence score of three type of lens and device at 9:00. ....	103
Table 21: Showing the confidence score of three type of lens and device at 10:00. ....	104
Table 22: Showing the confidence score of three type of lens and device at 11:00. ....	105
Table 23: Showing the confidence score of three type of lens and device at 12:00. ....	106
Table 24: Showing the confidence score of three type of lens and device at 13:00. ....	107
Table 25: Showing the confidence score of three type of lens and device at 14:00. ....	108
Table 26: Showing the confidence score of three type of lens and device at 15:00. ....	109
Table 27: Showing the confidence score of three type of lens and device at 16:00. ....	110
Table 28: Showing the confidence score of three type of lens and device at 17:00. ....	111



Table 29: Showing the confidence score of three type of lens and device at 18:00. .....	112
Table 30: Showing the confidence score of three type of lens and device at 19:00. .....	113
Table 31: Showing the confidence score of three type of lens and device at 20:00 without flash.....	114
Table 32: Showing the confidence score of three type of lens and device at 6:00 with flash.....	115
Table 33: Showing the detail confidence score for all direction and times from normal camera lens .....	116
Table 34: Showing the detail calculated confidence score for all direction and times from normal camera lens.....	117
Table 35: Showing the detail confidence score for all direction and times from Zoom camera lens.....	125
Table 36 : Showing the detail calculated confidence score for all direction and times from Zoom camera lens.....	126
Table 37: Showing the detail confidence score for all direction and times from Iphone camera. ....	134
Table 38 : Showing the detail calculated confidence score for all direction and times from Iphone camera.....	135

# Chapter 1 Introduction

## 1.1 Introduction

Museums have been around for a thousand of years and yet the interactions between people and antique have not been a subject of interest. Visitors simply look at art and read the descriptions, having little to no interactions with the pieces, sometimes losing the interest of those visitors. This is not new problem, but it is high since the past time where existing technology has yet to solve this problem. Recently, the amounts of visitors that museums receive overall are decreasing in number. The interests in museum and history have been declining in these past three years, because all museums in the world are all nearly the same. The minimal amounts of interaction between people and antiques are a large challenge when matched up against the more convenient Internet, removing most incentive to go to a museum at all. A change is needed in order to increase the popularity of museum and rekindle the popularity of history.

The introduction of the technology Augmented reality could be enough to improve, if not solve, this problem. The term Augmented reality was coined in the year 1990 but augmented reality itself has been around for more than 30 years. The interaction between people and antique is very crucial and plays a very important role in creating and maintaining interest in history. Augmented reality has already begun to be implemented into some museums and the use in some museums improves the experience of visiting a museum greatly. It already helps young adults better enjoy learning about history and allow children to understand more about history.

Augmented realities have improved the interaction between people and antiques by introducing the items in a more interesting way to people. Now people can interact or touch or play with the augmented antique on their screen. They can

also see the original shape of a broken antique, courtesy of the Augmented reality program. Augmented reality use in museum is still young and under development so it still has several limitations like how it's difficult to properly create models in places like outdoor exhibitions where antiques are shown outside with natural lighting. This paper will discuss about how to create a 3D marker for augmented reality that will work for outdoor museums.

Augmented reality have improved a lot in the past 30 years, previously Augmented reality only worked if based on an AR marker, but presently many different objects can be used as a marker such as barcode, QR, and picture marker. There are also markerless systems called markerless tracing. This system is not very reliable and, due to its complexity, needs a lot of resources as well as computing power. This research aims to find an alternative solution for this. The proposed alternative solution is to use multiple photos to create various angled picture markers of a single object in order to make a 3D marker of that object. This multiple picture marker method tends to have some problem, thus in the later stages of this research a machine learning program is implemented into this multiple picture marker method. The main contribution of this paper is to find whether this method new machine learning method is able to create markerless AR system as well as whether that method is reliable for practical use or not. This paper will present conducted experiments under various conditions of lighting; time both day and night, as well as artificial and natural lighting. The time, amount of light, and whether if artificial light is needed and how it affect the confidence score will be discussed a well.

## 1.2 Problem

This research is conducted regarding to two main problems as mentioned in the introduction section: the problem in museum visitors declining, and inconvenience of AR markers usage. According to the report, the museums around the world are facing new challenge; the audience has been tremendously decreasing. One of the reasons is that the museums lost their attractive, and the audience does not find that the traditional museums are interesting anymore, especially for the children and teenagers. To solve this problem, this research is proposed to increase the attraction of museums by utilizing the technology to create interaction between the exhibits and audience as new features represented on the audience's smartphone or tablets. With this application, the users can interact with the exhibits via their own devices, which make them enjoy and learn new knowledge in more pleasure way.

AR makers are very useful, but it is undeniable that it does not suit every object, including the antiques in the museums. Since the AR markers need to be attached to the object or require the space to place it, it may cause damage to the antiques or less convenient in case the markers are installed separately. In addition, the AR makers cannot be properly used in the outdoor exhibition. Consequently, this research also aims to get rid of this dilemma by making the objects as the markers. The users' device can track the exhibits directly without the AR markers usage.

Regarding to the problems and solutions ascribed above, it is highly expected that this research will significantly benefit the museums service to attract more people to visit them, and more appropriated to display the information than ever.

## 1.3 Literature Review

This chapter provides the description of background, relevant researches and theories. It is comprised of three sessions: general concepts of AR markers, visual recognition architecture, and current research efforts. The first explains basic idea of AR markers, its characteristics and role in various aspects. The next one presents the area of computer vision: visual recognition. Since this research requires the related researches as the foundation, those researches are illustrated in the third session. Lastly, the originality of this research is asserted as the conclusion of this chapter.

### 1.3.1 General Concept of AR Markers

In order to understand the concept of markerless, it is necessary to recognize the basic idea of the Augmented Reality markers (AR markers). AR markers were generated as the innovation for enormous benefits including optical tracking.

There are various kinds of markers based on its nature of the tracking algorithms (Koch et al, 2014). The first markers are ID markers (figure a). It is basically 2D markers used for simple AR application, and its structure is quite distinctive containing black border, which makes it easy to be detected and tracked. The next type is similar to the first one, Barcodes and Quick Response codes (QR codes) (figure b and c respectively). This type of markers appear in black lines or square modules arranged on the white background. It is optical machine-readable 2D, which can be tracked by imaging devices and interpreted to information. The third kind has been improved from the first and second, it is not only the black and white configuration, but it contains visual image or picture (figure d). Even though its appearance is color image, it still requires the black distinctive border as well. Thus,

the markers in initiative period generally require enough distinctive boundaries around the markers in order to be detected and tracked efficiently.



**Figure 1: Showing various type of AR marker.**

The markers were significantly developed in the next period; the borderline was removed. The first Markerless is in the form of 2D, but there is no boundary like previous markers (figure e). As it still requires distinctive visual images and advance algorithms, the improved one was generated, Markerless 3D (figure f). This is the most advance optical tracking method, as it can detect the real world object. However, the distinctive visual features are needed, and it must be scanned from various perspectives.

Apart from the type of markers explained above, the markers are occasionally divided by its separation with the object. If the objects are recognized by a markers intentionally installed on it, it is artificial markers. Basically the first generation of markers is included in this type. In contrast, when the device recognizes the objects acted as the markers itself, it is the natural markers (Edmund Ng Giap Weng et al, 2013)

Using marker based and markerless augmented reality systems can recognize the real world objects. Mostly, the previous developers used markers based augmented reality systems, but those systems actually hide the reality and it was also

difficult to keep the markers everywhere, especially outdoor. (Weng, 2013, p.1) There are several efforts from researchers in order to reduce the cost and time consuming as explained later in this chapter.

### 1.3.2 Visual Recognition Architecture

A key problem in multimedia data analysis is discovery of effective representations for sensory inputs-image, sound wave, haptic, etc, (Jia et al, 2014, p.1), as the machine performance cannot visual recognize and classify the objects as the human do. The model of deep learning or architecture has achieved a commanding lead on the state-of-the-art after success of digit classification in the 90's. Regarding to this phenomenon, large-scale visual recognition methods have been adopted by many software company for image searching (ibid, 2014) For the image processing methods for tracking, there are two tracking techniques classes in the computer vision: feature-based and model-based (Ibid, 2013). Finding a correspondence between 2D image feature and their 3D world frame coordinates is the technique of feature-based method. Meanwhile, the model of features of tracked objects such as CAD model is mostly the technique of model-based method.

Nowadays, there are many deep learning frameworks generated to support the visual recognition architecture containing the state-of-the-art algorithms such as Torch7, Caffe, and IBM Bluemix, which is employed in this research as a platform. Generally, they can provide the toolkit for training, testing, or deploying models, and support binding to programming language, for instance; Python, MATLAB, and C++. With their features, the users can make an application of object detection and classification, and even learning to extract the images.

### 1.3.3 Current research efforts

Several researches recently study on AR markers and visual recognition and classification in various aspects. The experiment of Dan Cireşan, Ueli Meier and Juergen Schmidhuber (2012) is one of those efforts. They ascribe that traditional methods of computer vision and machine learning cannot recognize handwriting or signs, like the human performance. Upon, they propose biologically deep artificial neural network as found in mammals between retina and visual cortex. They claim that their method is the first achievement near human performance. They focus on deep convolutional neural networks (DNN), which is proved the mettle on data sets including MNIST (handwritten characters), NIST SD19 (Latin characters), Chinese characters, CIFAR10 and NORB (3D toys).

The experiment of Christian Koch, Matthias Neges, and Markus Konig Abramovici (2014), is also the effort to utilize AR markers. Their purpose is to reduce the cost and time on the operation and maintenance (O&M or FM) phase. According to the research, the facility maintenance takes 50% of the on-site work, even though some software tools have been employed. The framework in their research is called Building Information Models (BIM) based Augmented Reality (AR). Normally the markers are designed, and made as the AR markers, but in their research, they use the operation and maintenance signs such as exit, smoke detector sign, which are the natural markers. Their experiment reveals that natural markers is effective under the conditions that the markers have strong and distinctive border with high contrast to the background, artificial lighting must be turned on, and proper frame rate applied. However, this research focuses on 2D markers, and results probably cannot be applied for other kinds of markers, especially the markerless.



Two researchs explained above focus on 2D visual object, the research of Zhirong Wu et al (2015) extends the field of machine learning by proposing a convolutional deep belief network to represent a geometric 3D shape (Ibid, 2015). Their model can recognize and reconstruct the objects by a 2.5D depth map and CAD model dataset, presented as a model namely 3D ShapeNets. Their experiment on it has been improved the state-of-the-art deep machine learning in various perspectives.

### **1.3.4 Academic Contribution of Study**

According to the literature review, there is plenty of research studying on machine learning, ranging from the basic concepts to the advanced deep learning called the state-of-the-art. However, the most of those researches focus on the markers. Although this research adopts the methods and results in order to contribute to a new research, it extends the field of machine learning by testing and proposing a demo application, which can detect and extract the information without the markers as the objects act as the marker itself. Futhermore, this research coordinates between the machine learning development, and AR markerless.

# Chapter 2 Concepts and Propose

## 2.1 Concept

### 2.1.1 Multiple picture marker

The method is to use multiple photos to create various angled picture markers of a single object in order to make a 3D marker of that object. This 3D marker can be viewed from multiple perspectives, yet still show the same result as a singular marker. The only requirements for this would be a mobile phone, a computer, and a program for creating (In this study, Vuforia was utilized) an object, making this an easy to access method for many people.

### 2.1.2 Machine learning

In this concept, AI technology of machine learning was applied to recognize an object instead of a marker for AR systems. Particularly, in this study, visual recognition function in IBM Bluemix was used for the machine learning. By connecting the mobile device to the cloud computer of IBM Bluemix, the heavy calculation load of machine learning can be performed and the AR system based on the machine learning can be constructed.

### 2.1.3 Difference between Multiple picture marker and Machine learning

The different between these two methods is how the program matches things. The Vuforia program uses image matching tool which matches the scanned image to the nearly similar image from the storage data. The Visual recognition Bluemix works differently, this visual recognition Bluemix uses machine learning to understand the structure of the object by multiple images given to the program prior to scanning. So the image scan by this program is then matches to the nearest similar structure instead of images like how Vuforia works. To put it in simple term Vuforia matches picture to picture but Bluemix matches picture to structure.

# Chapter 3 Image Matching Method

## 3 Vuforia Experiment

### 3.1 System Diagram

#### 3.1.1 Storing Phase

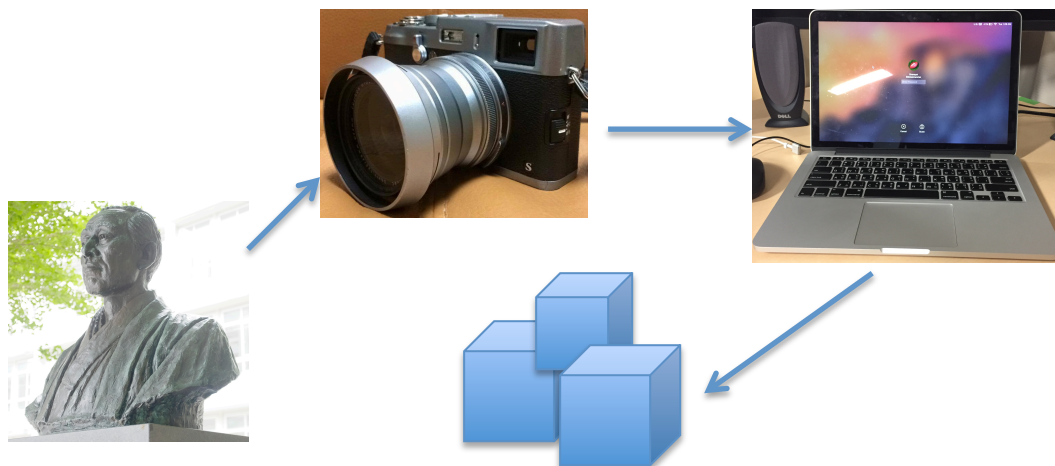


Figure 2: System Diagram for Storing Phase

First take pictures of the object and stored the image in the computer. The computer then processes the image into augmented reality marker or picture marker later store it into the local data based.

#### 3.1.2 Scanning Phase



Figure 3: System Diagram for Scanning Phase

Use the computer camera to view the object later on the image of the object is then compared with the local data based and the result will show up on the screen.

### 3.2 Nutella



Figure 4: Ten pictures taken of Nutella container in controlled environment

On the above picture, Figure 4 is a sample picture taken of Nutella chocolate container. Eight of the pictures are taken in distinct 45 degree angles around the Nutella. The remaining two pictures are of the top and bottom of the Nutella container. These Nutella pictures are taken in a room with a singular artificial light source, originating from the ceiling. In order to keep the experiment as brief as possible, the purpose of understanding the basic functions and how the program works, the maximum number of pictures used for processing by the Vuforia program was reduced to ten. After the method of processing would be better understood, the next experiment and the next object would be conducted upon, which was, in question, the Fukuzawa Bronze Statue.



**Figure 5: The ten images post-processed by the Vuforia program.**

These are the results after the program has used these pictures from Figure 4 to create the marker. As shown in the figure above (Figure 5) the picture has then been processed into grey scale picture composition (black and white), and yellow dot have been created within the pictures on the surface of the Nutella container. The yellow dots represent the marker detectability of the Marker. The higher the count points, the faster program can recognize the object as well as control the stability of the Marker. Therefore, the most ideal circumstance was a situation where there was a high volume of yellow dots on the pictures as shown in Figure 5, different sides of the Nutella container have different amount of yellow dots. This emphasizes that although the object is the same in all pictures taken in Figure 5, the detectability varies depending on the angle that the pictures were taken, regardless of the controlled environment where the pictures were taken. The number of yellow dot and/or detectable point are vary from the picture due to various reasons, therefore the processing of the pictures proved difficult for the program The worst examples of the disarrangement of processing are the pictures of the top and bottom of the Nutella container.

### 3.2.1 Calculation of Detectable Point



**Figure 6: An enlarged portion of the previously mentioned Fukuzawa Bronze Statue, showing detectable points represented by yellow crosses.**

The program used to create the augmented image shown in Figure 6 is Vuforia, a program that converts normal images to augmented images and also creates these yellow dots that represent the detectability rate of a picture. The Vuforia website provides a detectability rating, ranging from 1 star to 5 stars, for an image that is submitted for processing. Since this rating is quite rough and the program itself does not show the number of detectable points, the next paragraph will explain the process of how a more precise count was brought about.

The above picture (Figure 6) is showing a zoomed augmented image. This enlarged image shows us that these yellow dots are actually crosses. One cross sign is equal to a single detectable point, about 9 pixels of the image. Therefore, in order to count these detectable points another program has been used. The program uses to count these yellow pixels is Adobe Photoshop. As these augmented images are in

grey-scale, it is very easy to distinguish and count the number of yellow pixels in the picture with the help of Adobe Photoshop program. After the numbers of yellow pixels on the image have been estimated, the number of detectable point can be calculated from the number of yellow pixels on the image. As mention and shown (figure 6) before, 9 yellow pixels are equal to a cross or one detectable point.

The table below (table 1) shows the amount of yellow pixels on different sides of the Nutella container. The table also shows the estimated number of points calculated by the following formula:

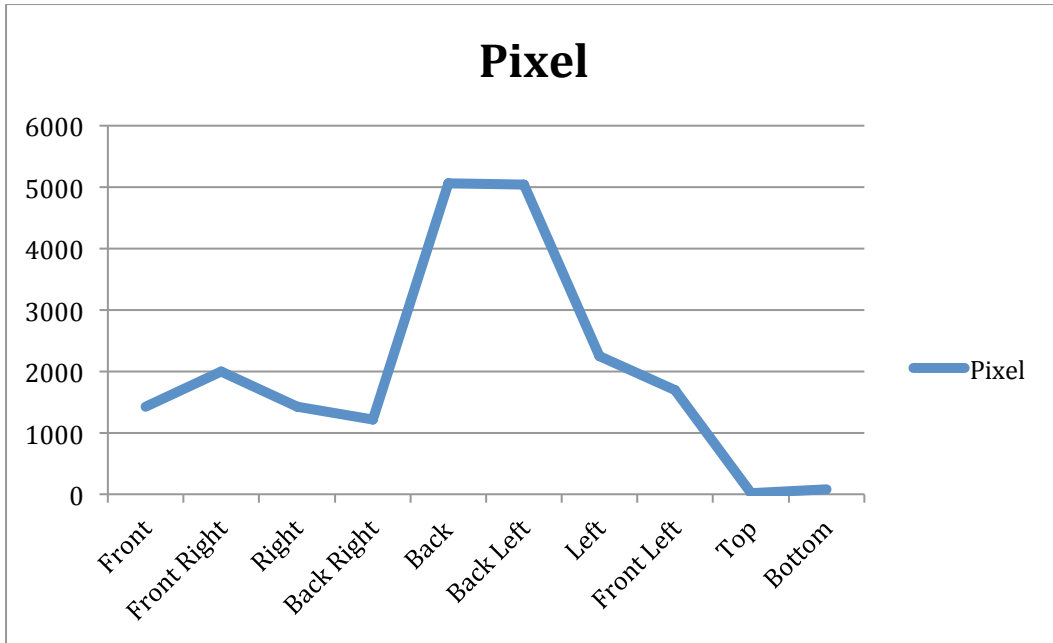
$$\text{Estimated number of detectable point} = \text{Number of yellow pixel} \times 9 \quad (1)$$

This formula was realized by counting the number of yellow pixels in across sign within the augmented reality image.

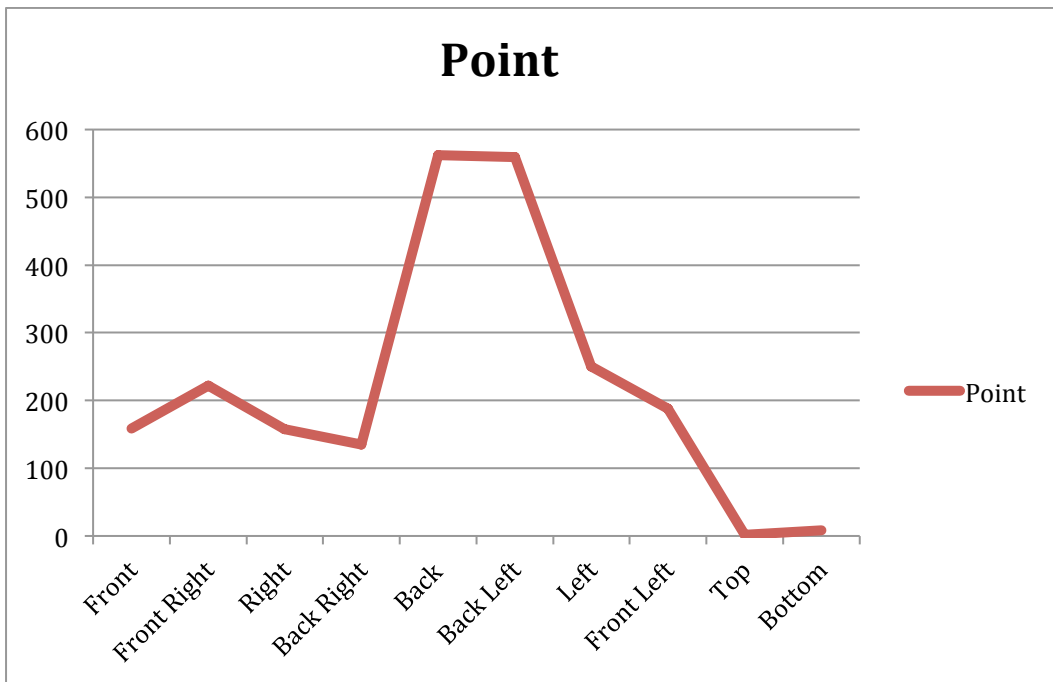
**Table 1: The number of pixels and point on each side of Nutella bottle.**

Side	Pixel	Point
<b>Nutella Front</b>	1426	158.4444444
<b>Nutella Front Right</b>	1997	221.8888889
<b>Nutella Right</b>	1422	158
<b>Nutella Back Right</b>	1212	134.6666667
<b>Nutella Back</b>	5061	562.3333333
<b>Nutella Back Left</b>	5036	559.5555556
<b>Nutella Left</b>	2253	250.3333333
<b>Nutella Front Left</b>	1695	188.3333333
<b>Nutella Top</b>	15	1.666666667
<b>Nutella Bottom</b>	76	8.444444444

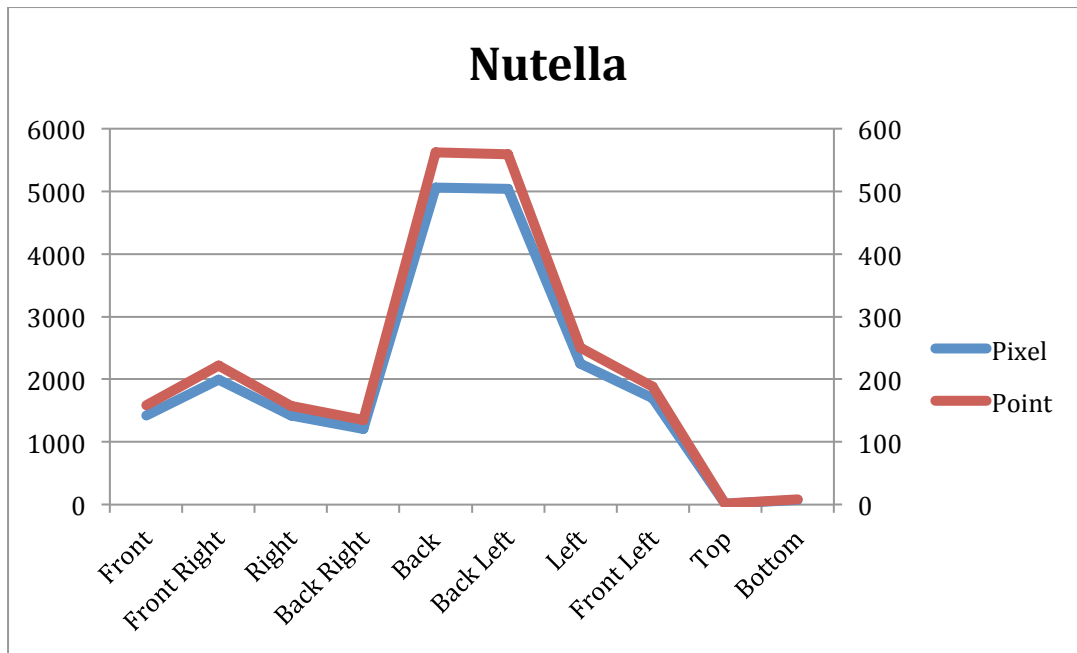




**Figure 7: Graph of the number of detectable pixels on each side of the Nutella container.**



**Figure 8: The number of detectable points on each side of Nutella.**



**Figure 9: Comparison of detectable points and pixels on each side of the Nutella container.**

In Figure 7 and Figure 8, the amount of pixels and detectable points in each side of the Nutella container are shown. It can be seen that the two sides with the most detectable points and pixels are the back side and the back left side while the two sides with the least detectable point are bottom and top of the container.

Figure 9 There are the comparison between numbers of pixels and points on each side of Nutella container. The X-axis represents the side of the Nutella Container and there are two Y-axis: on the left main Y-axis is the number of pixels while on the right hand side is a secondary Y-axis that represents the number of points. They are a slight difference between the two plotted lines because of an error between points and pixels, during the calculation since the formula uses was just estimation. As shown in Figure 6 some of the crosses do not contain 9 pixels. This is because when two detectable points are near each other, they can cause an overlapping between pixels. Due to this reason, the exact amount of detectable points cannot be perfectly calculated by this method.

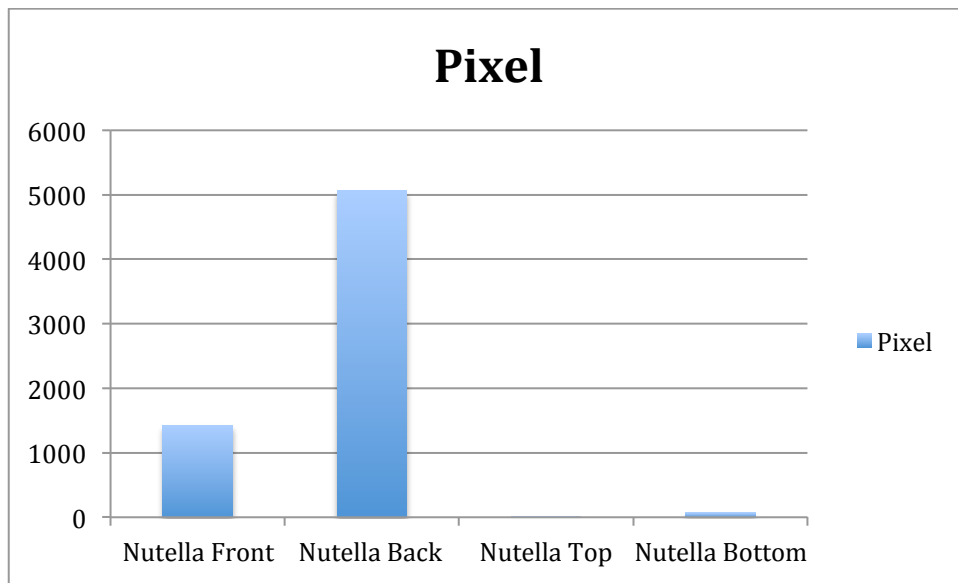
### 3.2.2 Effect of surface pattern



Figure 10: The effect of patterns on the surface to the creation of marker.

Here is a comparison of several picture of Nutella container showing how important the patterns on the surface are for creating a marker. As it can be seen, on the top and bottom do not have any pattern on it, which mean the program cannot create any detectable point (which is shown with the yellow crosses in the previous pictures) on the object therefore some part of the object become unaugmentable.

Figure 7, Figure 8, and Figure 9 show the same result; the peak values of detectable point are on the back while the lowest are on the top. In Figure 10, it shows that even though on the front side and the back side of the Nutella Container both have pattern on it but on the back side, (5061pixels) the detectable point are way more on the front (1426 pixels). This is because text patterns are more detectable than picture patterns as seen in the picture (figure 10) that on the backside they are various description of the Nutella, which can be used to create detectable point.



**Figure 11: Comparison of number of pixel between four sides of Nutella**

This Figure 11 is showing the max value of detectable pixels which is 5061 pixel in the back, in the front is 1426 pixels, in the bottom 76 pixels and the least amount of detectable pixel is the top which is only 15 pixels.

## 3.3 Statue

### 3.3.1 High focal Point



**Figure 12: Picture of Fukuzawa bronze from various angles with high focal point**

Here are the pictures of the Fukuzawa bronze statue taken from various sides in a similar fashion to the Nutella container, eight pictures are taken with average of 45 degree difference. These pictures were taken around the late afternoon during the autumn season by a Fuji Film camera with high focal point of 12.7, shutter speed of 2000ms and with ISO of 6400. Neither secondary lighting nor artificial lighting was used for taking these photos; the only light source was from the afternoon sun. Later these sets of image would be used and converted to augmented image via Vuforia the same means as the Nutella container. The converted image would then be used to count the yellow pixels and calculate the detectable points as well.

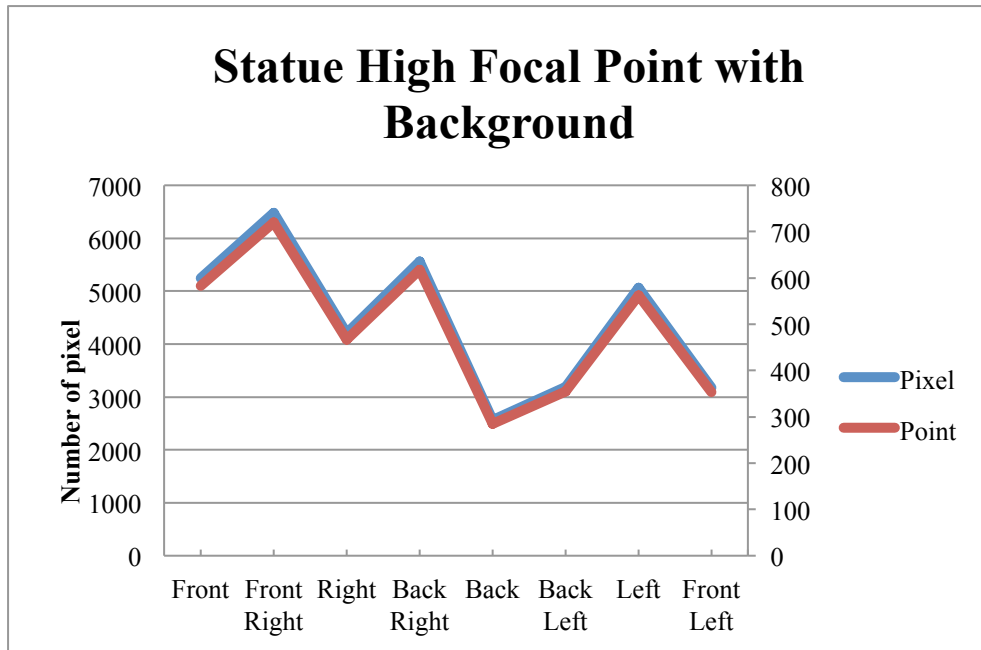


Figure 13: The number of pixel and point on each side of Fukuzawa bronze statue with background and high focal point

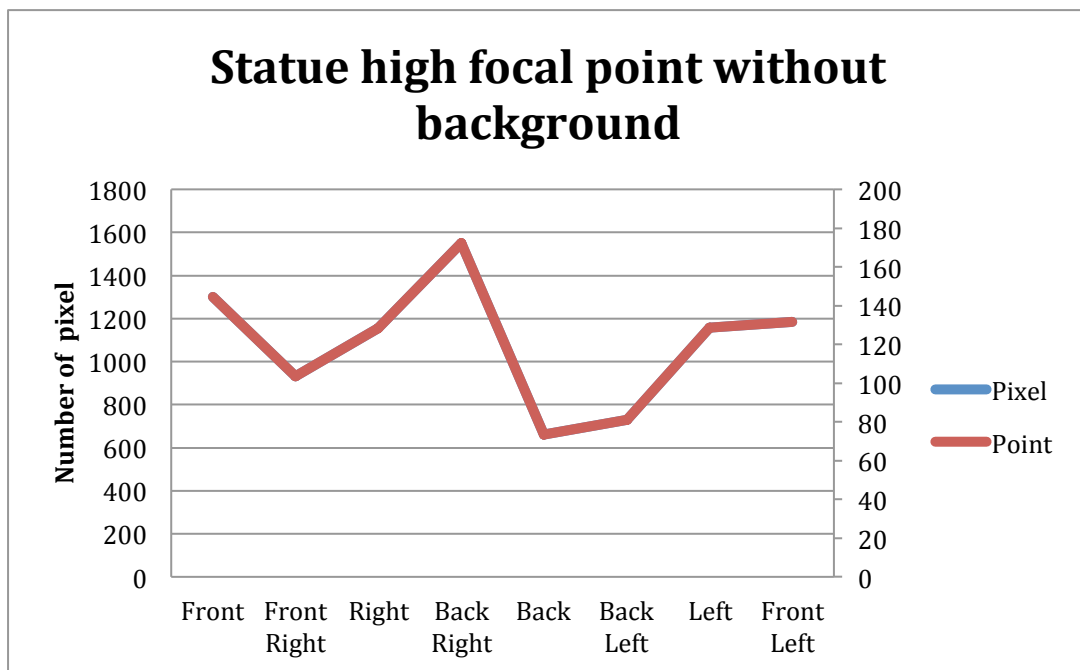


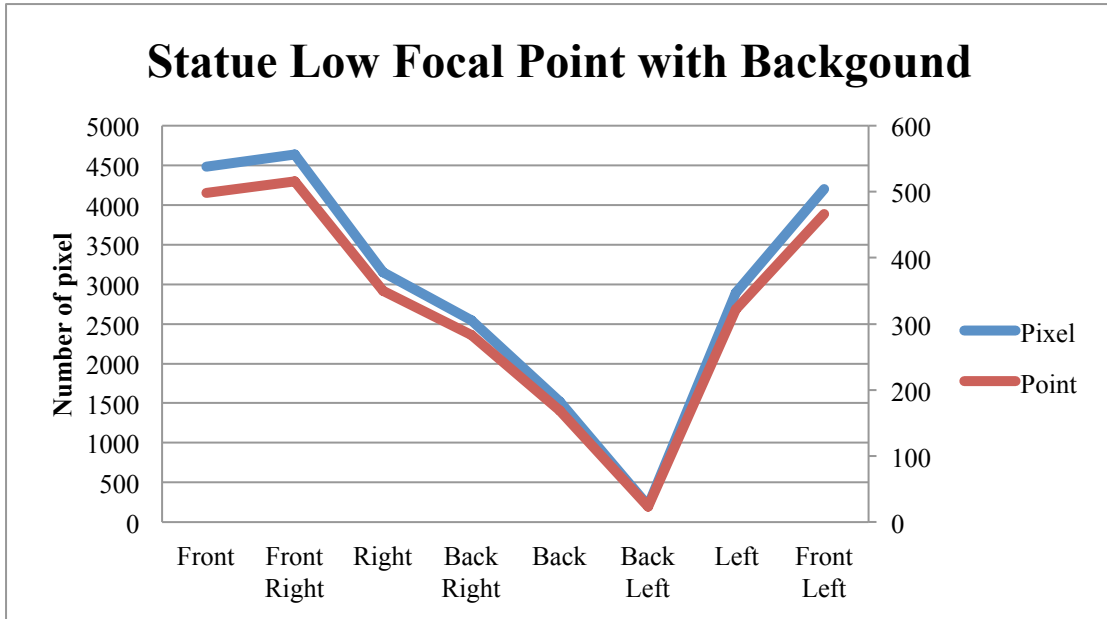
Figure 14: The number of pixel and point on each side of Fukuzawa Bornze statue without background in high focal point.

### 3.3.2 Low Focal Point

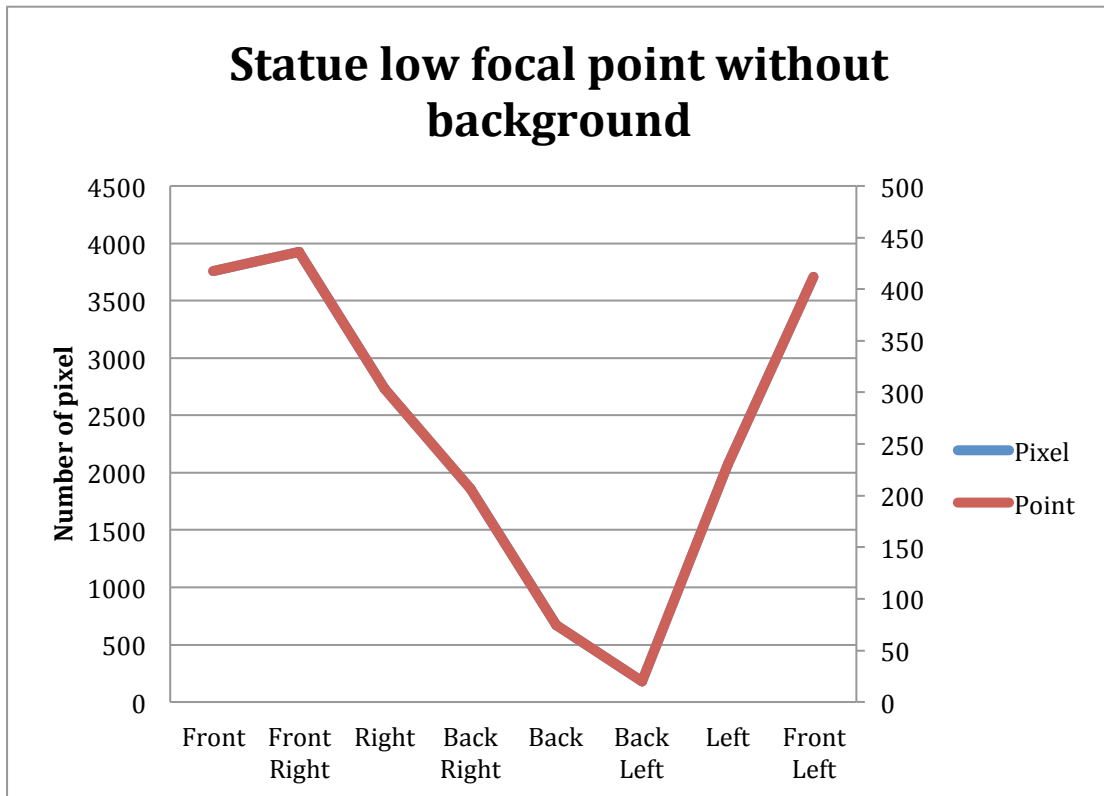


**Figure 15: Picture of Fukuzawa bronze from various angles with low focal point.**

Here are the pictures of Fukuzawa bronze statue taken from various sides similar with the Nutella container 8 pictures are taken with average of 45 degree apart. Similar with the previous set, but this time taken with different setting, the pictures were taken around the early afternoon during the autumn season by a Fuji Film camera with high focal point of 2, shutter speed of 2000ms and with ISO of 6400. During the taking of these images no secondary lighting such as flash, nor artificial light such as lamp or streetlight were used, the only light source is from the afternoon sun. The timing is different from when the high focal length picture was side, at this set of images looks brighter in comparison. But in this set of pictures, as seen in Figure 15 at the back left picture of the statue, the pictures were taken against the light and therefore the images look blurry and foggy. These images will pose a problem during the processing and it will be explained later in this paper.



**Figure 16:** The number of pixels and points on each side of Fukuzawa Bronze statue with background in low focal point.



**Figure 17:** The number of pixels and points on each side of Fukuzawa bronze statue without background in low focal point.



### 3.4 Effect of Background

#### 3.4.1 Effect of Background on High focal point



Figure 18: Processed images of Fukuzawa bronze statue with high focal length from various sides.



Figure 19: Processed images of Fukuzawa bronze statue with high focal length from various sides without background.

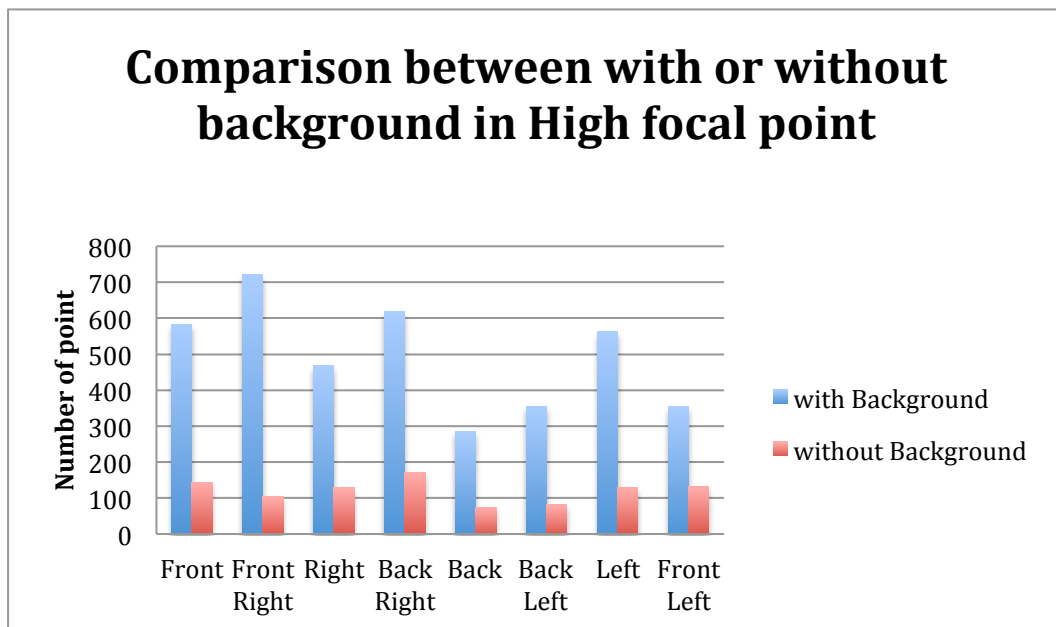
Figure 18 are Fukuzawa images that were processed by Vuforia. As seen from the images, a lot of detected point are scattered all around the image. By the look of it, we would have thought that these pictures that have a lot of yellow dot on the picture would have high detectability but it is does not. In this image, most of the detectable point (Yellow dots) is mostly on the background of the image. This mean during the scan, in order for the program to recognize the image and produce an augmented reality image, the background must be consistently one color, which is highly improbable in real life situations. Therefore, the true detectable points of the side of the statue would be the detectable points on the statue where the true value will be after subtracting the detectable point from the background.

From Table 2, with background the highest value of detectable points is the picture taken from front right with 6482 pixels followed by the back right which is 5556 pixels. The lowest value of detectable point is the back side which is 2564 pixels and the back left side is 3191 pixels. Comparing these values by removing the background, the highest value without background is the back right side with 1549 pixels and front side with 1301 pixels. The lowest is on the back side at 660 pixels and back left side at 728 pixels. In contrast, between the width background and without background, on average the without background's number of detectable point is diminished by about 74.6%. With the maximum of 85.6% on front right and the minimum of 62.8% on the front left.

This means that all the pictures' actual detectability average around 25% of its actual value. The detail comparisons between the two for all the sides of the statue are shown on Figure 20. This value will not be enough to produce a suitable marker.

**Table 2: The number of pixels and points on each side of Fukuzawa Bronze statue: high focal point with background in comparison to without background.**

Statue High Focal Point				
Side	With Background		Without Background	
	Pixel	Point	Pixel	Point
Front	5244	582.6666667	1301	144.5555556
Front Right	6482	720.2222222	931	103.4444444
Right	4206	467.3333333	1156	128.4444444
Back Right	5556	617.3333333	1549	172.1111111
Back	2564	284.8888889	660	73.33333333
Back Left	3191	354.5555556	728	80.88888889
Left	5060	562.2222222	1159	128.7777778
Front Left	3182	353.5555556	1184	131.5555556



**Figure 20: The comparison of the detectable points of high focal point between with background and without background.**

### 3.4.2 Effect of Background on Low Focal Point



Figure 21: Processed images of Fukuzawa bronze statue with low focal length from various sides.



Figure 22: Processed images of Fukuzawa bronze statue with high focal length from various sides without background.

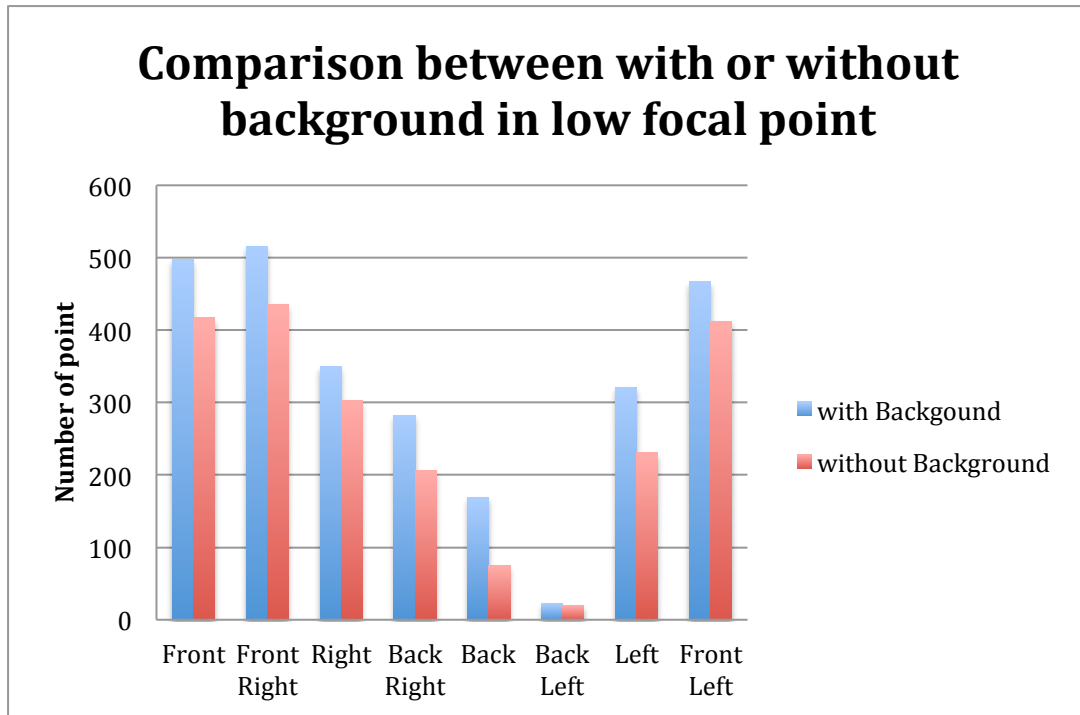
Figure 21's images of Fukuzawa are quite similar with figure 18 excluding just difference in the focal point. The easiest way to tell the main difference between these two figures is the amount of objects on the background in this set of images (figure 21). They have lesser significantly detectable point in the background than the previous set of images (figure 18). Furthermore, other differences include more detectable point in the statue while in low focal point. As mentioned before, in the low focal point section regarding the back left picture which was taken against the light, these picture have shown a drastic decrease in detectable point.

Table 3 shows that with background, the highest value of detectable point are front right side with 4640 pixels follow by the front side which is 4481 pixels. The lowest value of detectable point is the back left side which is 207 pixels and back side which is 1521 pixels. Comparing these values with the removed background, the highest value is front right side with 3924 pixels and front with 3757 pixels. The lowest are on the back left side: that is 180 pixels and back side at is 670 pixels. In contrast, between with background and without background on average, without background the number of detectable point is diminished by about 22.6%. With a maximum 55.9% for the back side and the minimum of 13.0% on the back left.

This means that all these picture's actual detectability are on average of only 78% of its true value. The difference between with background and without background value is actually quite small, therefore in the low focal the impact of the background are low. The detail comparisons between the two for all sides of the statue are shown on Figure23. These values are a lot more reassuring when compared with high focal point. The reason of this will be discussed in the next section: Effect of focal point.

**Table 3: The number of pixels and points on each side of Fukuzawa bronze statue in Low focal point with background and without background.**

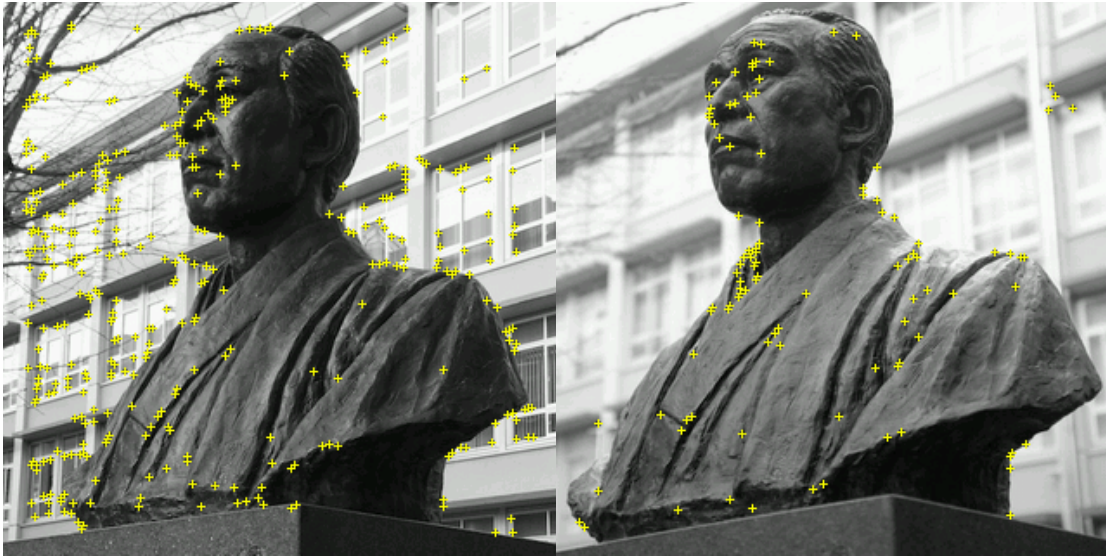
<b>Statue Low Focal Point</b>				
<b>Statue Low Focal Point</b>				
	<b>With Background</b>		<b>Without Background</b>	
<b>Side</b>	<b>Pixel</b>	<b>Point</b>	<b>Pixel</b>	<b>Point</b>
<b>Front</b>	4481	497.8888889	3757	417.4444444
<b>Front Right</b>	4640	515.5555556	3924	436
<b>Right</b>	3149	349.8888889	2729	303.2222222
<b>Back Right</b>	2545	282.7777778	1861	206.7777778
<b>Back</b>	1521	169	670	74.44444444
<b>Back Left</b>	207	23	180	20
<b>Left</b>	2893	321.4444444	2073	230.3333333
<b>Front Left</b>	4199	466.5555556	3706	411.7777778



**Figure 23: The comparisons of the detectable point in low focal point between with background and without background.**

What is not desirable is to have any detectable point in the background, because it will cause complication during the processing phase of the program. Furthermore, if they are detectable points in the background, it will cause confusion to the Vuforia program. Since the Vuforia compares the amount and position of detectable points, whether it is the same as the recorded image or not. If it is not the same, the augmented image cannot be shown to the users. In this case, it is impossible to keep the background from changing due to various reasons such as weather, time and season. All these reasons occur naturally, which cannot be control by any means. Another option to prevent these backgrounds from changing is to have the statue store indoor, this method can be highly inconvenient dependent on the object in question.

### 3.5 Effect of focal point



**Figure 24: The effect of background High focal point on the left and Low focal point on the right.**

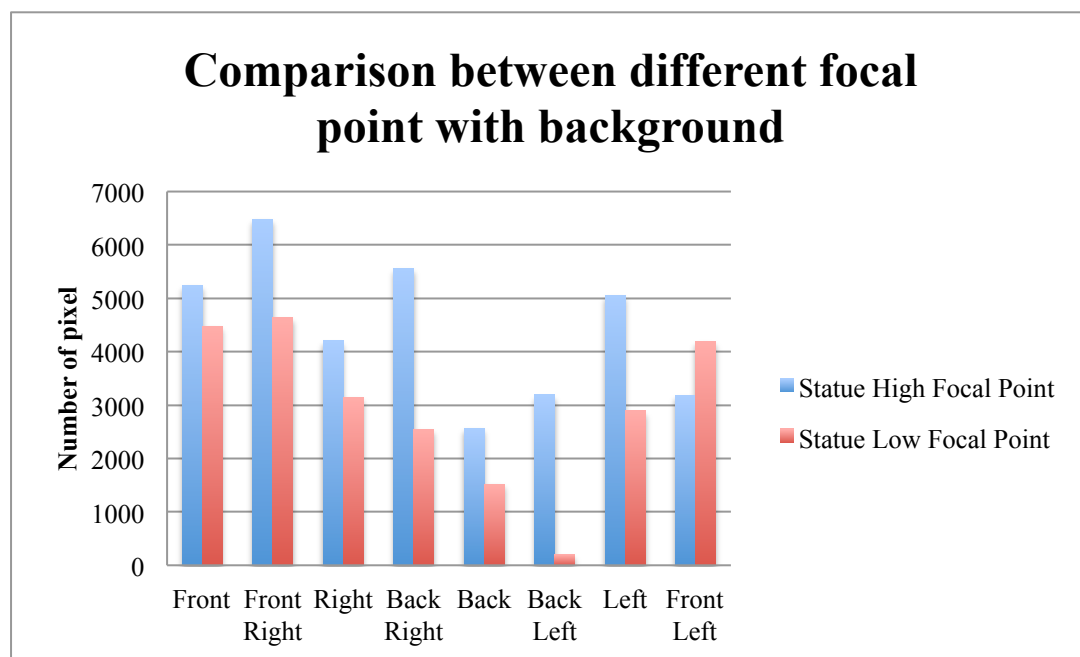
By reducing the focal length of the lens in the camera, the background of the picture is blurred out (out of focus), so the program cannot use the background to create detectable points (yellow dots). Now all the detectable points will fall within the focus point only (in this picture means on the statue only). As shown in Figure 24, there is a greatly reduced number of detectable points in the background with the help of changing the focal length of the lens.

The basic concept of how the programs work has been proven in the previous section regarding the effect of background. It also can be seen in Figure 20 and Figure 23 showing that the number of detectable points on the background has been greatly reduced simply by reducing the focal length of the lens. In the enlarged left image in Figure 24, where the image is more in focus, thus sharper, and more detectable points are recognized on unrelated objects. If compared to the right side image, the background of the image is blurred out.

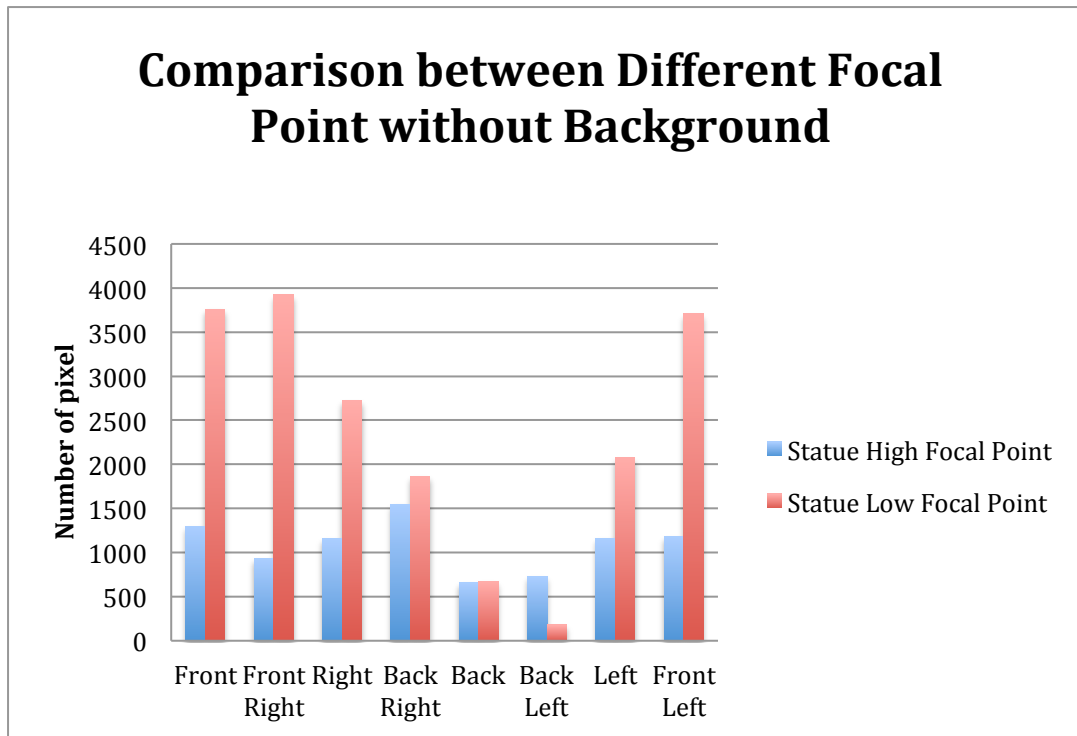


There are two methods to remove this; one is to use a large single colored screen to cover the background which is very trivial task, while the other way is to use a low focal point lens in order to blur out the background. The latter method is used in this experiment and it is the main method used in this paper.

In Figure 25, it can be noticed that the number of detectable points in high focal length pictures in all sides are more than low focal point except one; the Front left side. The highest value is the back right 3011 pixels followed by the back left side 2984 pixels and the lowest are the front side 763 pixels follows by the back side 1043 pixels. The lowest value does not include the front left which is a negative value of -1017 pixels which means that in the front left side, the detectable point of low focal length are more than high focal length. Lastly the average difference is 1481 pixels.



**Figure 25: The comparison of the detectable point statue picture with background**



**Figure 26: The comparisons of detectable point statue picture without background between low and high focal point.**

In Figure 26 the number of detectable point in low focal length picture in all sides are greater than the high focal point excluding the back left side. The highest value is the front right side 2993 pixels follows by the front left side 2522 pixels and the lowest values are the back side 10 pixels follows by the back right side 312 pixels the lowest values does not include the back left side which is a negative value of -548 pixels, meaning that in the back left side the number detectable point of high focal length are greater than low focal length. The average difference is about 1279 pixels.

The comparison between focal points is from Figure 25 and Figure 26. First, in the Figure 25 shows that detectable point of high focal length are far greater than low focal length in contrast with Figure 26 after the background have been removed where it becomes the opposite, this shows that in low focal point, the most of the detectable point are on the statue and not on the background, which was the most desirable results. Not only that most of the detectable point falling with in the statue,

but more detectable points are formed within the statue when comparing with high focal point as well, even though the highest amount of detectable point that can be obtained from the picture is from high focal length at 6482 pixels rather than the low focal length at 4640 pixels. This is because in high focal point, the picture is focused on all the objects in the picture equally. Therefore in every part of the image, it is crystal clear, which in the case of low focal point where the focus of the picture on the statue left the surrounding or the background blurred. As mentioned before, the program uses sharp edges of an image to create detectable point, hence the reason why the highest number of detectable points can be obtained from high focal point. Figure 18 and 21 also prove that in high focal point, pictures lose about 74.5% of its detectable points. If the background is removed while in low focal point it lose only 22% on average, that is about 50% different.

From the result of these graphs, it can be concluded that low focal length is better and give more detectable points as well as reducing the detectable points created in the background. Excluding for the back left side image, in this back left side image, it shows completely different result from all other sides, because the picture was taken against the sunlight. This type of pictures are called Backlit Images which have a negative effect to the image processing program as well as the Vuforia program, which was used in this experiment.

### 3.6 Effect of backlit image



**Figure 27: The normal backlit image (right) as well as the augmented backlit image (left)**

Backlit image are images where the light source and the camera are on the opposite of each other. In this situation (Figure 27), the light source is the sun. Again, as mentioned before, the function of the program is to uses the shape edges on the image to create detectable points. If the image is taken against the light, it will be more often than not creating a foggy or blurry layer on the picture. Even though the image is sharp, but due to this layer of blur, it will obstruct the sharpness. Therefore, the Vuforia program cannot recognize it in order to create detectable point and reducing the number of detectable points. The problem is in this image; even though it is taken with a low focal point lens, still has a little number of detectable points even inside the statue where the image should be focused.

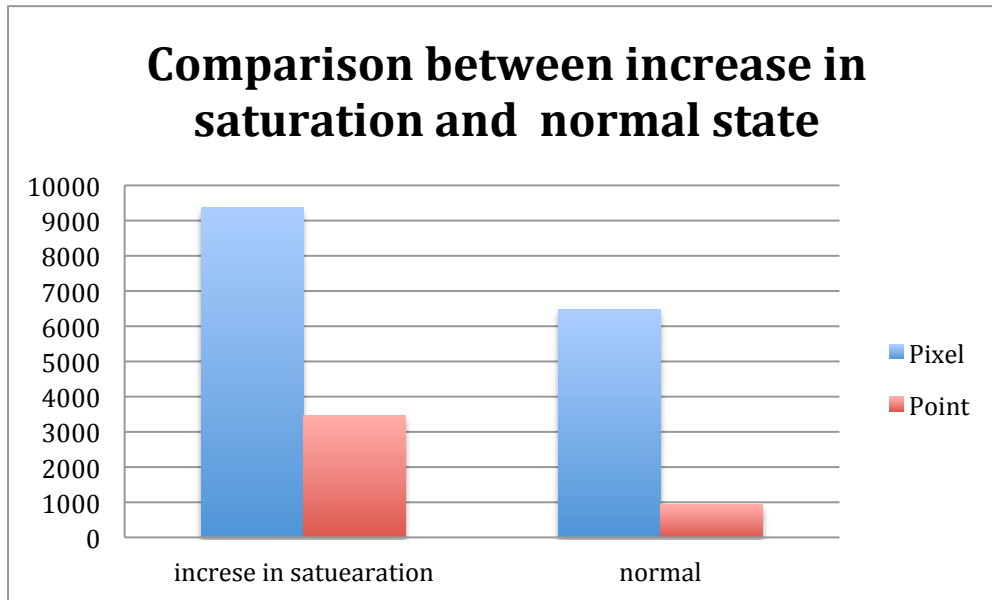
### 3.7 Effect of Saturation



**Figure 28: The effect of saturation.**

This is another possible way to produce more detectable point on the picture incase the number of detectable point is not enough to be used as a marker. An increase in saturation as well as an increase in exposure with this method the image will looks brighter and sharper, thus it will greatly boost the number of detectable point (yellow dots) as demonstrated in Figure28. This is an post image processing which is difficult to know what percentage of saturation should be increased or how much exposure should be increased in order to optimize the result. Because in different pictures, it need different amounts of saturation as well as exposure in order to obtain the maximum number of detectable points. As mentioned before, it is a post

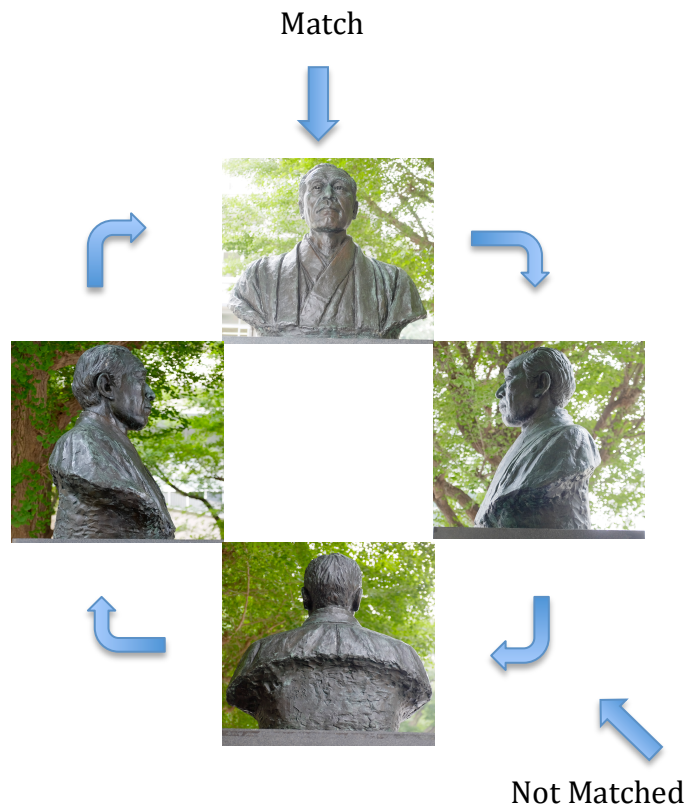
image processing, which is unwanted, because the image should be able to use directly from the camera of the mobile phone or camera. Even though the numbers of detectable points increase through this method, the picture would not be able to match due to the difference of saturation.



**Figure 29: The comparison of number of pixels and detectable points between normal image and increased in saturation image.**

It still greatly increases the number of detectable point on the image, regardless Figure 29 increasing. Saturation images have 9362 pixels and normal image have 6482 pixels, which is calculated to about 30.7% more detectable point than normal images. These calculations are calculated with all the detectable point including background. Now the without background, increase in saturation image have 3458 pixels and normal image have 931 pixels, which mean that it greatly increase by 73% comparing to the normal picture, a very significant amount. This method can be used as a last resort in order to increase the number of detectable points on the image.

### 3.8 Problem of the method

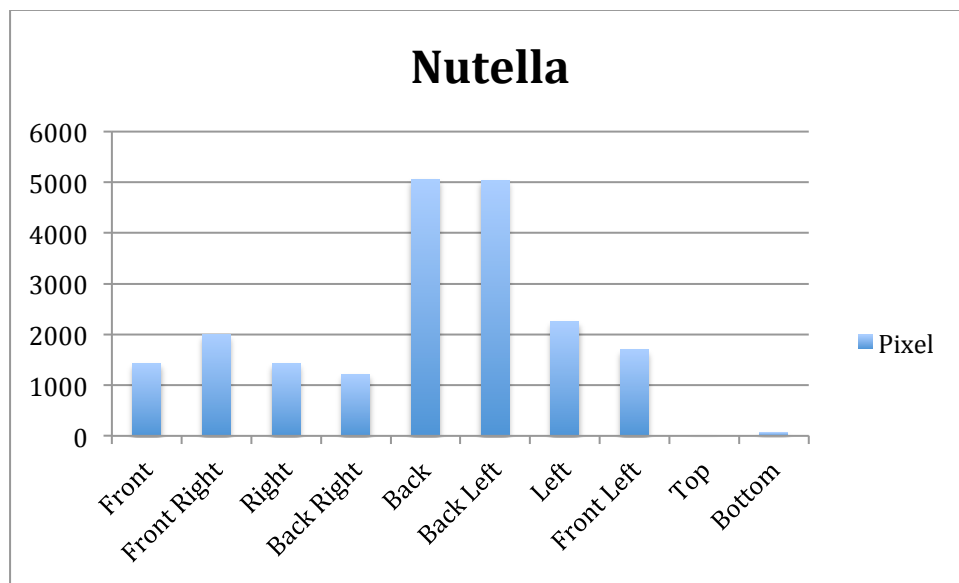


**Figure 30: The problem of Vuforia**

There are several problems with Vuforia method, for example: assuming that there are only four images stored in Vuforia as shown in the above figure 30. If users scan the image on the diagonal, of course it will be a no match because they are no image from that side or direction. This is a simple example explaining the flaw of using multiply picture marker program such as Vuforia. It can only recognize the image store prior to the scanning. Even though more picture were used to create the 3D marker, it still will not be enough since it cannot be knew that which direction the user will scan it from, even if all pictures were really taken from all direction, the size of the data will be too large to comprehend.

### 3.9 Discussion

As for verification for image matching method, an experiment was conducted to check the recognition rate for image matching method. The aim of this experiment is to analyze whether recognition rate is high enough for practical uses or not. There are two object used for this experiment, the Nutella container and Fukuzawa Bronze statue.



**Figure 31: The number of detectable pixel on each side of Nutella**

As for Nutella, the learning image was taken in a control environment and the scanning phrase was also conducted in a control environment. The element that is controlled in this control environment is lighting. This experiment was conducted in a well-lit, constant lighting throughout the room. As for the result the Nutella container is able to be recognized by the image matching method program from various direction up to about 2 meter distance with the exception of the top and the bottom of the Nutella container which have zero recognition at any distance. Referred to the figure above (Figure 31) it shows that from all side except top and bottom have high number of detectable pixel, as mention before the more the detectable pixel the better



the detectability. In the top and bottom the number of the detectable pixel were too low for the program to recognize, this render these two side unaugment able.

As for Fukuzawa bronze statue, it cannot be moved, so it is impossible to conduct the learning phrase in the control environment. Consequently, the learning picture was taken outside with various uncontrollable variables. From experiment with Fukuzawa bronze statue a result was obtained regarding the best method to increase the number of detectable pixel in order to improve the overall detection capability of the program. Referring to Figure 28 in section 3.5 Effect of Focal Point, it shows that the number of pixel was increase by more than 50% on average. As a result due to various uncontrollable factors such as amount of light and direction of the light, the image matching method program cannot recognize the Fukuzawa bronze statue. From any direction and distance, the image matching method program has zero recognition rates for Fukuzawa bronze statue even though they have high number of detectable point from all side.

The experiment result is that the recognition rates are too low for practical uses if the object is placed outside, but if the object is located indoor, where the environment can be controlled, the image matching method program can work perfectly.

# Chapter 4 Machine Learning Method

## 4 Visual Recognition Bluemix

### 4.1 What is Visual Recognition Bluemix (IBM Watson, 2016)

The IBM Watson™ Visual Recognition service uses deep learning algorithms to analyze images (.jpg, .gif, or .png) for scenes, objects, faces, text, and other content, and return keywords that provide information about that content. In classify calls these keywords are called **classes**, and there are hundreds prebuilt class in this service and available for use immediately. You can use these classes to identify relationships between images and create applications that can utilize those relationships. When you classify an image, each identified class receives a **score**. Scores can range from 0 - 1, with higher scores indicating greater correlation.

You can also train a group of new classes to create your own custom classifier with the Visual Recognition service. During training, you can supply images to create two or more classes that are trained against each other. These classes are grouped into a **classifier** and maintain their relationship later, when you use the custom classifier to analyze an image. This allows you to create a multi-faceted classifier that is specific to your needs.

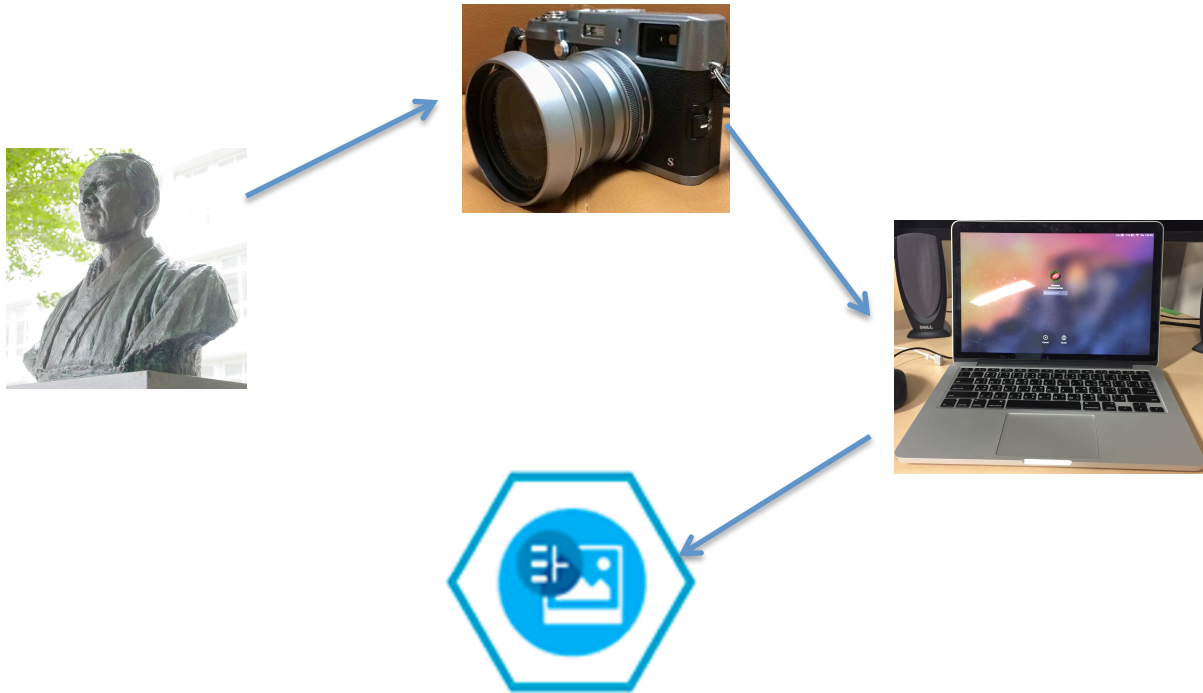
## 4.2 The purpose of using Visual Recognition Bluemix

The purpose of using the service of IBM Bluemix is to take advantage of its ability, the Image Analyzing ability. This ability can recognize an image of a scene, objects, faces, and many other things and its able to return certain information of those things. The returned information usually consists of two types of information which can include the name of its class and the systems confidence score. These two pieces of information are the important information we need in order to develop this markerless augmented system.

The other purpose of using this visual recognition software is new classes can be taught to the system or create a custom classifier for the service to recognize our object of interest. Lastly, this service is offered via cloud, one of the greatest advantages of using this service, since this deep learning or this machine learning process need a lot of computing power which normal mobile phone cannot provide or it will take too much time to compute for everyday practical uses. This cloud system allow the users to upload the picture into the cloud system and the cloud system will compute, process the image and send the result back to the mobile phone or the application. This means that it can be used in any device that can connect to the Internet.

## 4.3 System diagram

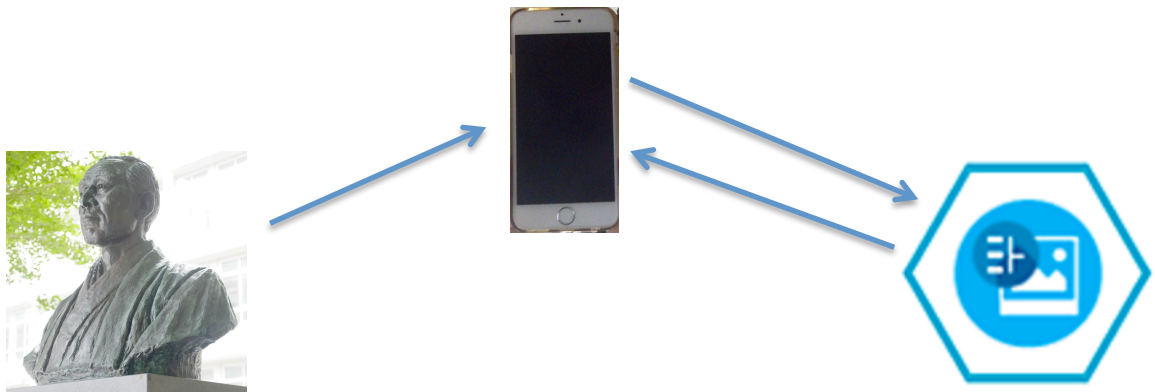
### 4.3.1 Learning Phrase



**Figure 32: The system diagram of Learning Phrase**

The picture is taken around the object and store in the computer. Then from the computer, it was uploaded to the cloud system from IBM Bluemix for training.

### 4.3.2 Scanning Phase



**Figure 33: The system diagram of Scanning Phase**

The image is taken from the mobile phone, then that image is sent to the cloud system allowing the IBM Bluemix program to analyze the picture. After that, the result is then sent back to the mobile phone, which then produces an augmented reality image on the mobile phone.

## 4.4 Confidence score

Confidence score is a score given by the program ranging from 0 to 1, which represents the certainty of recognition of the object.

### 4.4.1 Learning phase

For the learning phase of the program, the Bluemix Visual Recognition program requirement is to have 50 positive images and 50 negative images with a total size of less than 100MBs. Positive images are the images of the object and negative image are the images that have similar features to the object but differentiates from the object in other ways. For this experiment, 2,100 image was used, 2,000 positive images and 100 negative images. The positive 2,000 images consisted of three types of images, two types are from a camera and one type from phone camera. For the camera images, 500 images were taken with normal lens and 500 images were taken with a zoom lens. The rest of the 1000 image were from an Iphone camera. These three types of image all consisted of different type of image. It consisted of the images of the Fukuzawa bronze statue, which are taken at different times throughout the day from 6:00 to 20:00; and during the night. Two set of image are taken: one with flash, and another one without flash. At the top of each hour, the images are taken from various directions of the Fukuzawa bronze statue with different type of camera and lens.

### 4.4.2 Testing

16 sets of images are used similarly with the learning phase, each set consisting of eight images from the front, front left, left, back left, back, back right,

right, and front right sides. These 16 sets are taken from 6:00 to 20:00; and at 20:00 two sets are used, one set uses flash and the other does not.

## 4.5 Variable and Formula

Average represents the average of confidence score of all the image by time or by side

$$\textit{Average} = \textit{confidence score of all the picture} \div \textit{Total number of picture}$$

Subaverage represents the average of confidence score of a detectable image only and neglect the non detectable image, in simple word it mean any image with zero confidence score is ignore in the calculation.

$$\begin{aligned} \textit{Subaverage} &= \textit{confidence score of all detected image} \\ &\div \textit{The number of detected image} \end{aligned}$$

Maximum represents the maximum confidence score of photos taken in relation to a specific time or side of the statue.

*Maximum(for side)*

*= Highest confidence score from a single side at that particular time*

*Maximum(for time)*

*= Highest confidence score from a single time at that particular side*

Minimum represents the minimum confidence score of photos taken in relation to a specific time or side of the statue.

*Minimum(for side)*

*= Lowest confidence score from a single side at that particular time*

*Minimum(for time)*

*= Lowest confidence score from a single time at that particular side*



Minimum represents the minimum confidence score of photos taken in relation to a specific time or side of the statue, but ignoring the non-detected image.

*Minimum(for side)*

= *Lowest confidence score of a detected image from a single side at that particular time*

*Minimum(for time)*

= *Lowest confidence score of a detected image from a single time at that particular side*

Count represents the number of non-detected image, which also mean the images that have zero confidence score.

*Error Count = Number of nondetected image*

Recognition percentage represent the percentage which the program able to recognize the image.

$$\text{Recognition percentage} = \frac{\text{Total number of image} - \text{Count}}{\text{Total number of image}} \times 100$$

## 4.7 Statue

### 4.7.1 Times Affects Confidence score

In this section, the effect of time on the confidence score will be the topic of discussion. Time is one of the variables that affected the confidence score, yet it did not affect the confident score directly but indirectly, since this experiment was conducted in the outdoors without artificial lighting of the Fukuzawa Bronze statue. What is actually affecting confidence score is the position of the sun, which further affects the amount of light as well as the direction of the light on the statue. There are many trees in the area surrounding the Fukuzawa bronze statue, these trees' shadow positions as well as how sunlight passes through each leaf and falls on the statue also change according to the time. These shadows and the sunlight, in turn, affected the confidence score either.

In order to decrease the complexity of the data presented in this paper, the confidence score on each side of the Fukuzawa Bronze statue was averaged out and the data focusing on time, type of camera used, and confidence score only. In depth comparisons of each side for each time graph can be found in the appendix.

From 8PM onwards, an artificial light was used; in this case a flash from the camera and the mobile phone. In order to understand how lighting would help improving confidence score during nighttime, the flash function had to be used. In this experiment a flash was used with the camera and the phone excluding the zoom lens, due to hardware incompatibility reason, therefore there was no results regarding the confidence score for 8PM with flash using the zoom lens.

#### 4.7.1.1 Camera

Figure 34 shows that the times, which the confidence score are peaked are around 7:00, 14:00, and 17:00. The highest is at 14:00 with a value of 0.6169, followed by 7:00 with a value of 0.616. From 9:00 to 10:00 the confidence score remained constant. The times, which confidence score are at it is trough, are 8:00, 12:00, and 18:00. From 10:00 to 12:00 the confidence score drops from 0.53 to 0.39 and from 12:00 to 14:00 the confidence score rises from 0.39 to 0.6169. During the other times the confidence score fluctuate throughout the day.

Figure 35 shows the Maximum, Minimum, and the Minimum Detect. The Maximum value does not fluctuate much throughout the day, only a 0.02 differences between consecutive times. The Maximum deviations of the Maximum Confidence score are at after 17:00 from 0.68 to 0.65, which accounted to 0.03 differences. The Minimum deviations are from 14:00 to 15:00 with 0.663 to 0.666, which accounts to only a 0.003 difference. Compared to the Maximum, the Minimum Detect fluctuates quite a bit more, up to the maximum of 0.1 confidence score; while the minimum fluctuations are only up to 0.001. The red line is the minimum line for all the times, excluding 7:00 and 14:00 the rest are zero. At 7:00 and 14:00 the confidence score are 0.544 and 0.516 respectively.

Figure 34 and Figure 35 show that the average line and the subaverage line meet at 7:00 and 14:00. This means that from all sides during those times, all images were able to be recognized by the program; hence why average subaverage is equal. Similarly, in Figure 35, the Minimum and Maximum Detect lines meet at the same time period, therefore this graph confirms the statement above.

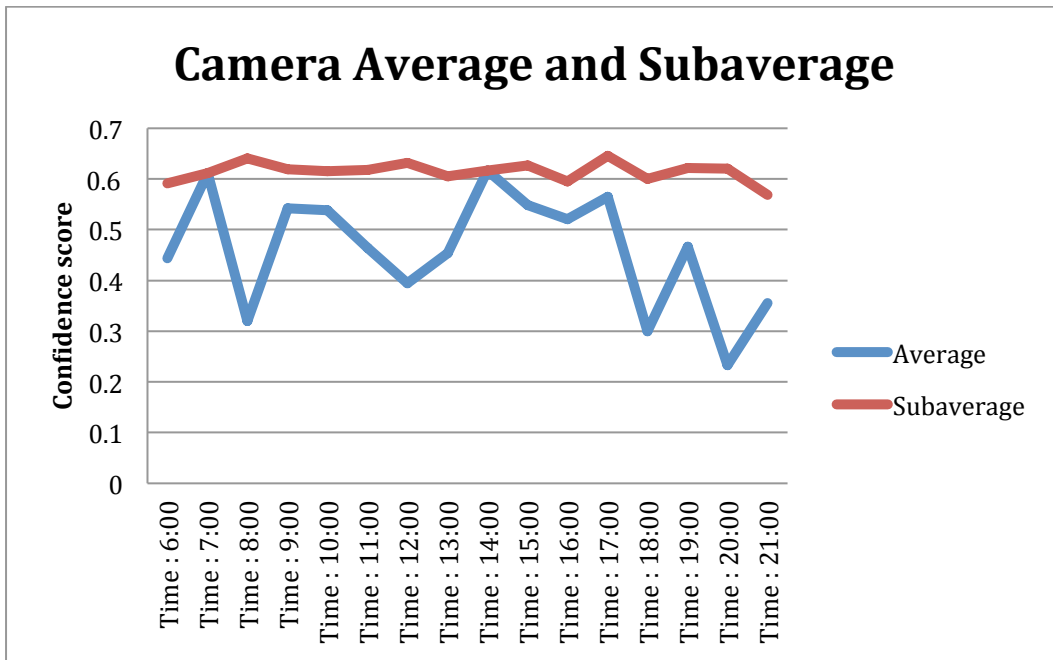


Figure 34: Comparing between the Average and Subaverage confidence scores for the camera.

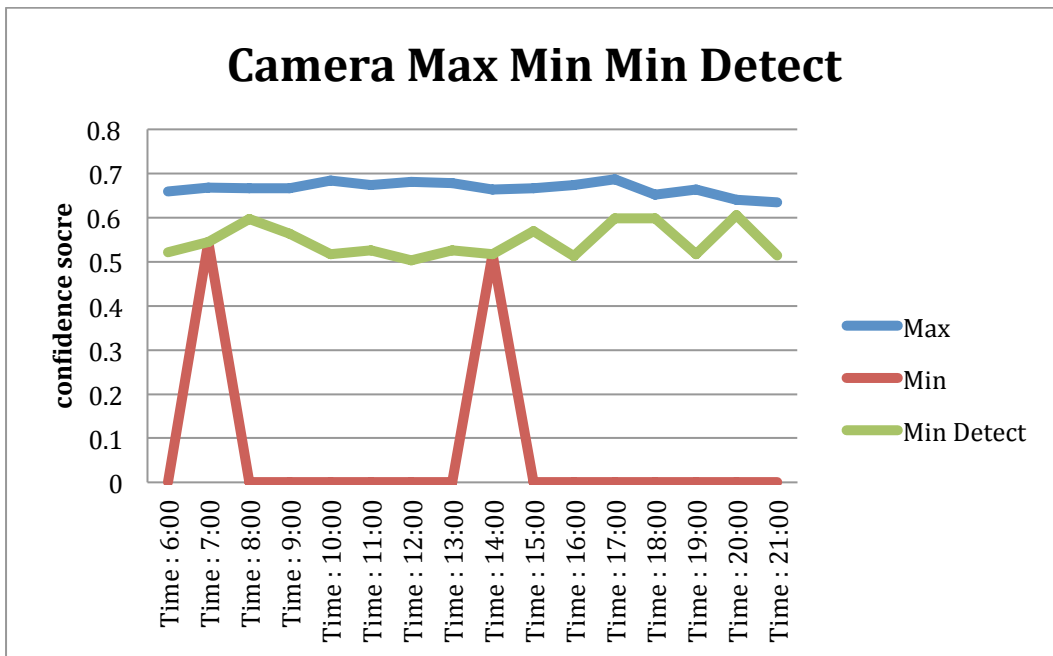


Figure 35: Comparing the Minimum, Maximum and Minimum Detected image confidence score for the camera

**Table 4: The confidence score at different times as well as the average, subaverage, Maximum, Minimum and Minimum Detect for Camera.**

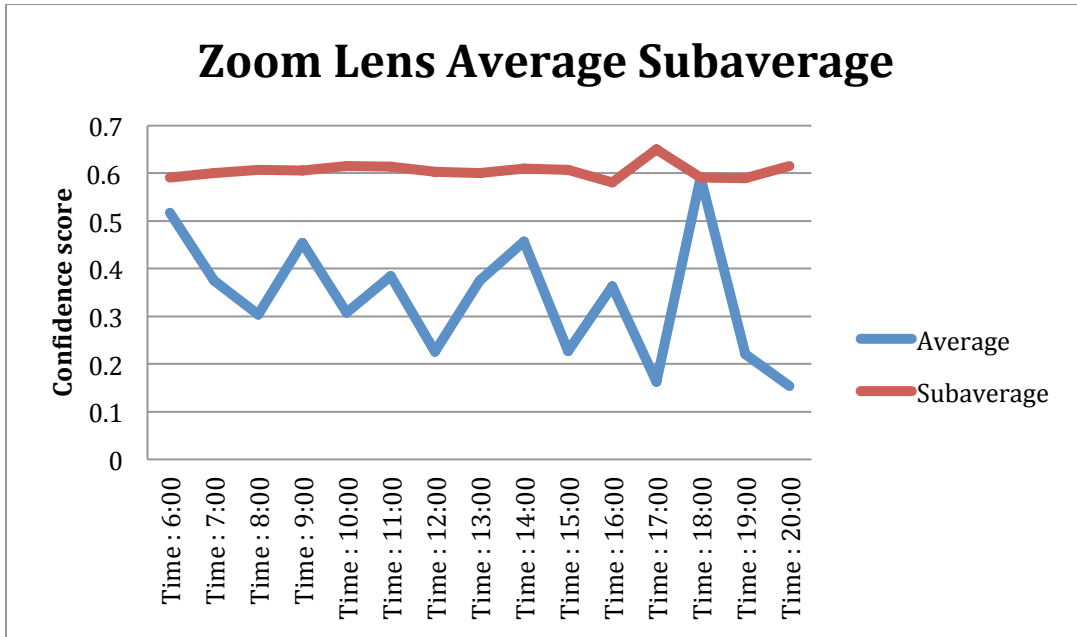
<b>Camera</b>	<b>Average</b>	<b>Subaverage</b>	<b>Max</b>	<b>Min</b>	<b>Min Detect</b>
<b>Time : 6:00</b>	0.443371125	0.5911615	0.658923	0	0.521687
<b>Time : 7:00</b>	0.61162375	0.61162375	0.668124	0.544783	0.544783
<b>Time : 8:00</b>	0.320511375	0.64102275	0.666184	0	0.596643
<b>Time : 9:00</b>	0.541896	0.619309714	0.665919	0	0.563159
<b>Time : 10:00</b>	0.538089875	0.614959857	0.683541	0	0.517203
<b>Time : 11:00</b>	0.463362	0.617816	0.673838	0	0.525534
<b>Time : 12:00</b>	0.394783625	0.6316538	0.680714	0	0.503145
<b>Time : 13:00</b>	0.454349375	0.605799167	0.678001	0	0.526098
<b>Time : 14:00</b>	0.616973375	0.616973375	0.663159	0.516993	0.516993
<b>Time : 15:00</b>	0.548484	0.626838857	0.666998	0	0.569535
<b>Time : 16:00</b>	0.52059275	0.594963143	0.673054	0	0.512741
<b>Time : 17:00</b>	0.564589125	0.645244714	0.68623	0	0.59877
<b>Time : 18:00</b>	0.30039375	0.6007875	0.652649	0	0.59877
<b>Time : 19:00</b>	0.465977	0.621302667	0.66293	0	0.517686
<b>Time : 20:00</b>	0.2327405	0.620641333	0.63983	0	0.604924
<b>Time : 21:00</b>	0.355148	0.5682368	0.634148	0	0.513637

#### 4.7.1.2 Zoom Lens

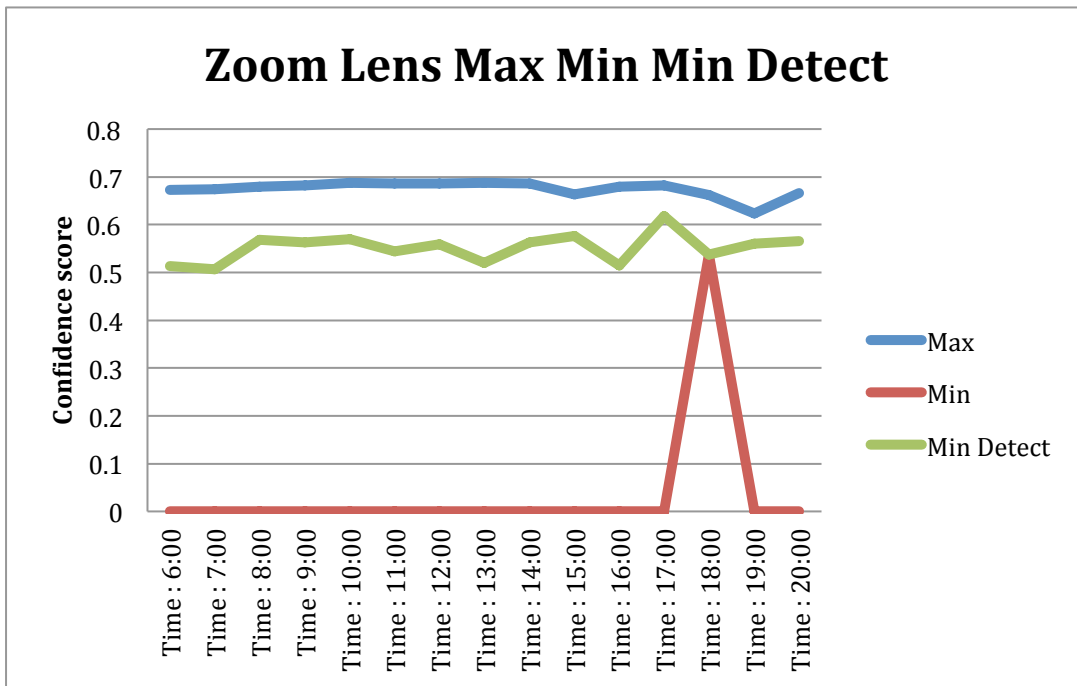
Figure 36 shows that the times, which the confidence score are peaked, are around 6:00, 8:00, 13:00, and 18:00. The highest is at 18:00 with the value of 0.5916 followed by 6:00 with the value of 0.51. The times which confidence score are at its trough are 12:00, 15:00, and 17:00. From 6:00 to 8:00 the confidence score drops from 0.51 to 0.303 and from 12:00 to 14:00 the confidence score rises from 0.22 to 0.4572. During the other times the confidence score fluctuate throughout the day.

Figure 37 shows the Maximum, Minimum, and the Minimum Detect. The Maximum value does not fluctuate much throughout the day, only a 0.004 differences between consecutive times. The Maximum deviation of Maximum Confidence score are of from 16:00 to 17:00 from 0.51 to 0.61, which accounted to a 0.1 difference. The Minimum deviations are of 8:00 to 9:00 from 0.567 to 0.563 in which there is only a 0.003 difference. Compared to the Maximum, the Minimum Detect fluctuates quite a bit more, up to the maximum of a 0.1 confidence score and the minimum fluctuations are around 0.005. The last red line that is the minimum line for all the time, excluding 18:00 the rest are zero. At 18:00 the confidence score are 0.5369.

Similarly, with the previous graph for camera, Figure 36 and Figure 37 show that the average line and the subaverage line meet at 18:00. This means that from all side at that time, all the images were able to recognized by the program, which is how the average and the subaverage can be equal. Similarly, in Figure 37 the Minimum line and Minimum Detect line meet at the same time further confirming the statement above.



**Figure 36: Comparison between the average and subaverage confidence score for Zoom lens**



**Figure 37: Comparison the Maximum, Minimum and Minimum Detect confidence score for Zoom lens**

**Table 5: The confidence score at different time as well as the average, Subaverage, Maximum, Minimum and Minimum Detect for Zoom lens.**

<b>Zoom lens</b>	<b>Average</b>	<b>Subaverage</b>	<b>Max</b>	<b>Min</b>	<b>Min Detect</b>
<b>Time : 6:00</b>	0.517509125	0.591439	0.672693	0	0.512745
<b>Time : 7:00</b>	0.375104875	0.6001678	0.674068	0	0.506502
<b>Time : 8:00</b>	0.303684375	0.60736875	0.679334	0	0.567663
<b>Time : 9:00</b>	0.4546045	0.606139333	0.682641	0	0.563488
<b>Time : 10:00</b>	0.307916	0.615832	0.687954	0	0.569167
<b>Time : 11:00</b>	0.383886875	0.614219	0.685873	0	0.543674
<b>Time : 12:00</b>	0.2260425	0.60278	0.686309	0	0.559093
<b>Time : 13:00</b>	0.3756555	0.6010488	0.687913	0	0.519926
<b>Time : 14:00</b>	0.457276125	0.6097015	0.686178	0	0.562783
<b>Time : 15:00</b>	0.2279615	0.607897333	0.663494	0	0.576694
<b>Time : 16:00</b>	0.363143125	0.581029	0.679518	0	0.514563
<b>Time : 17:00</b>	0.162581625	0.6503265	0.682681	0	0.617972
<b>Time : 18:00</b>	0.591670625	0.591670625	0.661661	0.536892	0.536892
<b>Time : 19:00</b>	0.221082375	0.589553	0.623404	0	0.560573
<b>Time : 20:00</b>	0.153965	0.61586	0.666339	0	0.565381



#### *4.7.1.3 Phone Camera*

Figure 38 shows that the times, which the confidence score are peaked, are around 8:00 and 12:00. The highest is at 11:00 with the value of 0.279 and followed by 12:00 with the value of 0.294. From 13:00 to 15:00 the confidence score remain constant. The times, which confidence score are at its trough, are 9:00 and 13:00. From 12:00 to 13:00 the confidence score drops from 0.29 to 0.07 and from 9:00 to 12:00 the confidence score rises from 0.14 to 0.29. During the other times the confidence score fluctuate throughout the day.

Figure 39 shows the Maximum, Minimum, and the Minimum Detect. The Maximum value does not fluctuate much through out the day only with around 0.02 differences between consecutive times. The Maximum Deviations of Maximum Confidence score are at 8:00 to 9:00 from 0.2 to 0.14, which accounted to 0.06 different. The Minimum deviations are from 16:00 to 17:00 at 0.613 to 0.616 with only 0.003 differences. Compared to the Maximum, the Minimum Detect fluctuates a quite a bit more, up to the maximum of 0.1 confidence score and the minimum fluctuation are 0.001. The last red line that is the minimum line, all the rest are zero.

Figure 38 and figure 39 show that the Max line and the Min average line meet at 6:00, 13:00, 15:00 and 19:00 it can be mistaken that it means that from all sides at that time, all the images were able to be recognized by the program like the previous 4 graphs but this time differently. It lacks confirmation from average and sub-average, therefore it is not exactly in the same way as before. In this graph the Maximum and Minimum Detect meet each other meaning that there is only a single value to calculate. This implies that only one side was recognized by the program, making it have only value able to be calculated.

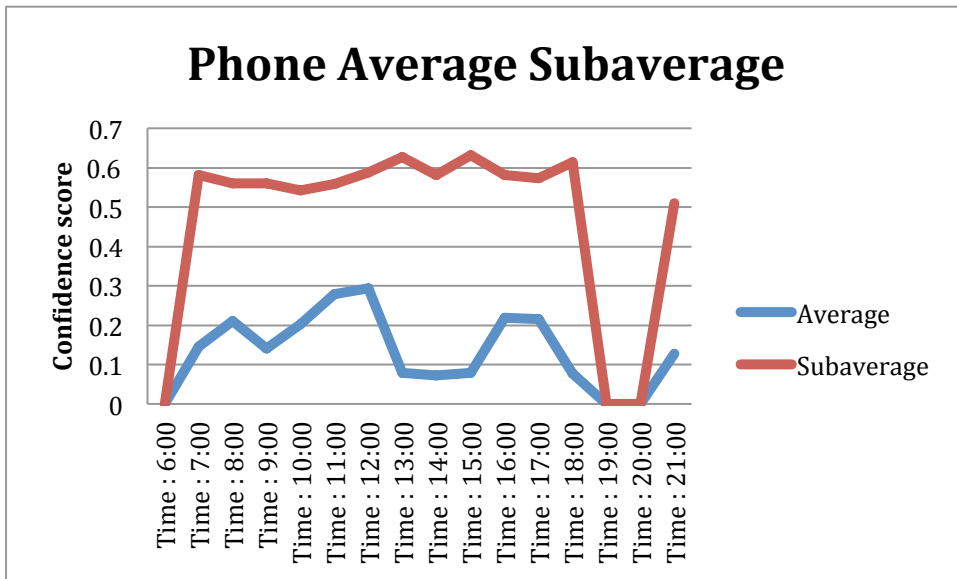


Figure 38: Comparison between the average and subaverage confidence score for Phone

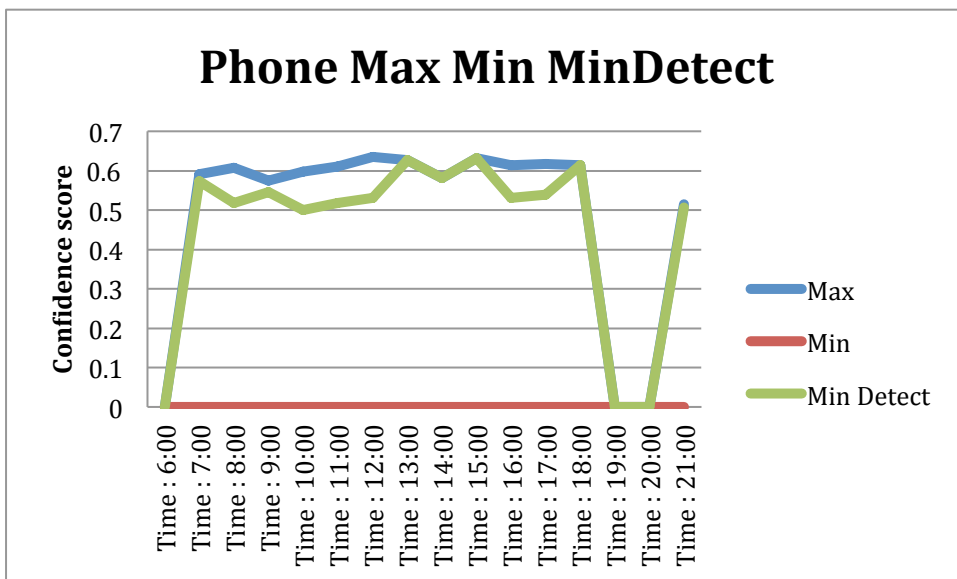


Figure 39: Comparison between the Maximum, Minimum and Minimum Detect confidence score for phone.

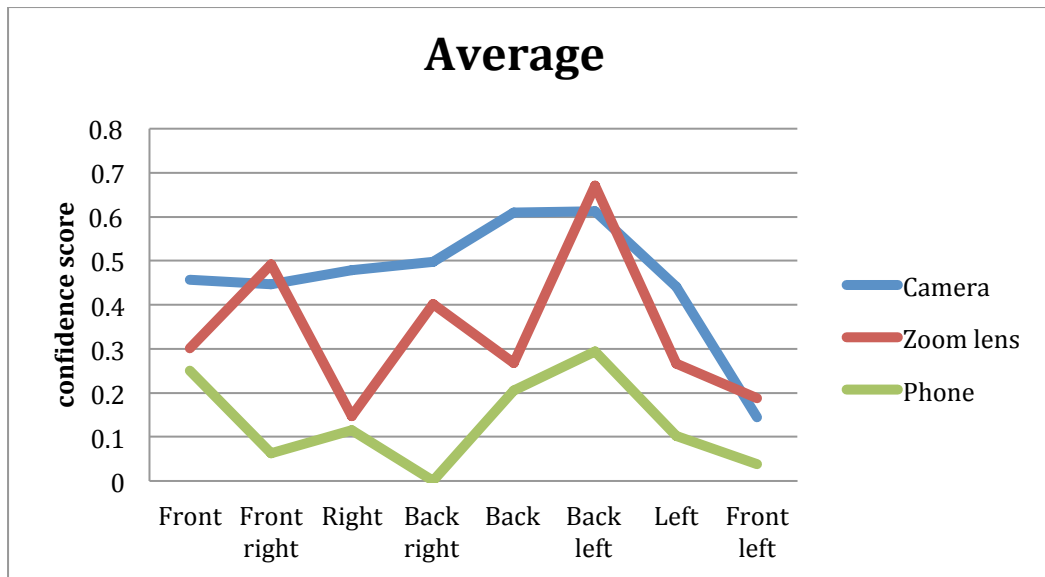
**Table 6: The confidence score at different time as well as the Average, Subaverage, Maximum, Minimum, and Minimum Detect for Phone.**

<b>Phone</b>	<b>Average</b>	<b>Subaverage</b>	<b>Max</b>	<b>Min</b>	<b>Min Detect</b>
<b>Time : 6:00</b>	0	0	0	0	0
<b>Time : 7:00</b>	0.145442375	0.5817695	0.590679	0	0.57286
<b>Time : 8:00</b>	0.20992525	0.559800667	0.607055	0	0.517347
<b>Time : 9:00</b>	0.140188625	0.5607545	0.57529	0	0.546219
<b>Time : 10:00</b>	0.20321425	0.541904667	0.598373	0	0.500623
<b>Time : 11:00</b>	0.279327125	0.55865425	0.610206	0	0.517918
<b>Time : 12:00</b>	0.2942975	0.588595	0.634185	0	0.530334
<b>Time : 13:00</b>	0.078445375	0.627563	0.627563	0	0.627563
<b>Time : 14:00</b>	0.07258	0.58064	0.58064	0	0.58064
<b>Time : 15:00</b>	0.07891175	0.631294	0.631294	0	0.631294
<b>Time : 16:00</b>	0.218235875	0.581962333	0.613689	0	0.53053
<b>Time : 17:00</b>	0.21503775	0.573434	0.616924	0	0.539612
<b>Time : 18:00</b>	0.076706875	0.613655	0.613655	0	0.613655
<b>Time : 19:00</b>	0	0	0	0	0
<b>Time : 20:00</b>	0	0	0	0	0
<b>Time : 21:00</b>	0.12757875	0.510315	0.51425	0	0.50638

### 4.7.2 Side Effect score

Side is also one of the variables, which affects confidence score. In this case, on each side of the Fukuzawa bronze statue, there is a different pattern. In this section, the discussion will center around how the Side variable can affect confidence score. Even though during the learning phase of the program, each side equal number of image is used, as a result, the outcome of the confidence score tended to be different; some of these ended up different to as far of an extent as for some of the sides to have zero percent recognition. Front sides have the most patterns, including the face and the wrinkle on the shirt of the statue. In theory, the front sides would have the most confidence scores and the backside would have the least confidence score due to the low amount of pattern on the backside. The right and left side would have equal or nearly equal confident score since the statue is nearly symmetric.

The result is quite different from what was theorized; the front side that should have the highest confidence, and the back left side have the highest score didn't quite match results as predicted. The lowest confidences score has become the front left instead of the backside. All the images taken in the back right side taken at any time cannot be recognized by the program. In contrast, all the images taken in the back left side from zoomlens from anytime and frames can be recognized by the program. The results of the back right side came up with a zero percent recognition rate; while the back left, however lacking in differences it was, came up with a 100% recognition rate.



**Figure 40: Comparison between average confidence score from different sides**

**Table 7: Average confidence score from different side**

<b>Time: Average</b>	<b>Camera</b>	<b>Zoom lens</b>	<b>Phone</b>
<b>Front</b>	0.456905938	0.3013948	0.250879625
<b>Front right</b>	0.446260063	0.491284867	0.062937688
<b>Right</b>	0.478810688	0.1467296	0.11522075
<b>Back right</b>	0.497167625	0.401048267	0
<b>Back</b>	0.609924313	0.268076933	0.20586525
<b>Back left</b>	0.612442	0.669972667	0.293936875
<b>Left</b>	0.440249688	0.2659486	0.102752125
<b>Front left</b>	0.1446825	0.187322467	0.038353438

In Figure 40, peak confidence score for all the three types of devices are at the same side, the back left side, this can confirm that back left side has the highest average score during all times of the day. The lowest confidence score, for the camera and zoom lens tended to coincide with each other, and with the front left, but for the Phone, it was the back right, which had zero recognition. Zoom lens performed better than normal lens for front right and back left sides. Overall, the order of the best performance from best to worst ranks as camera, zoom lens, and phone.

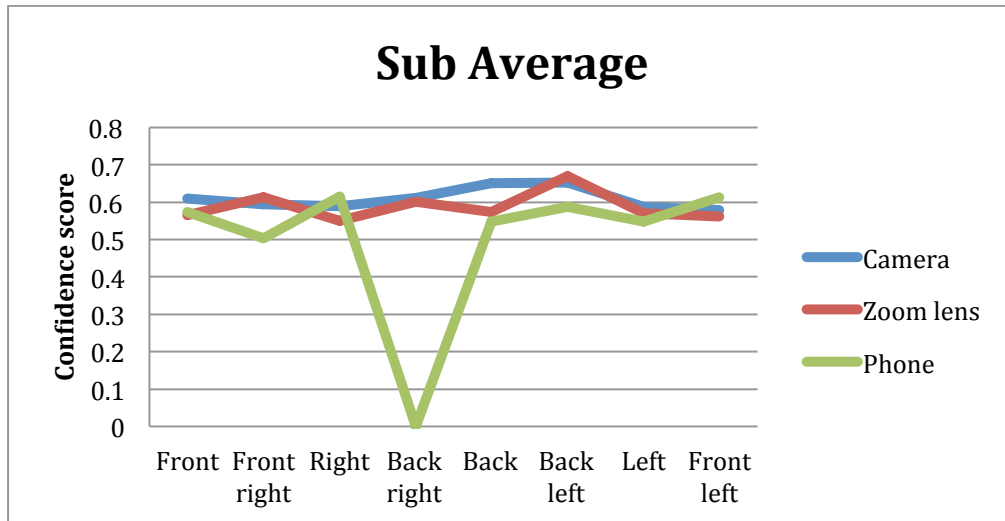
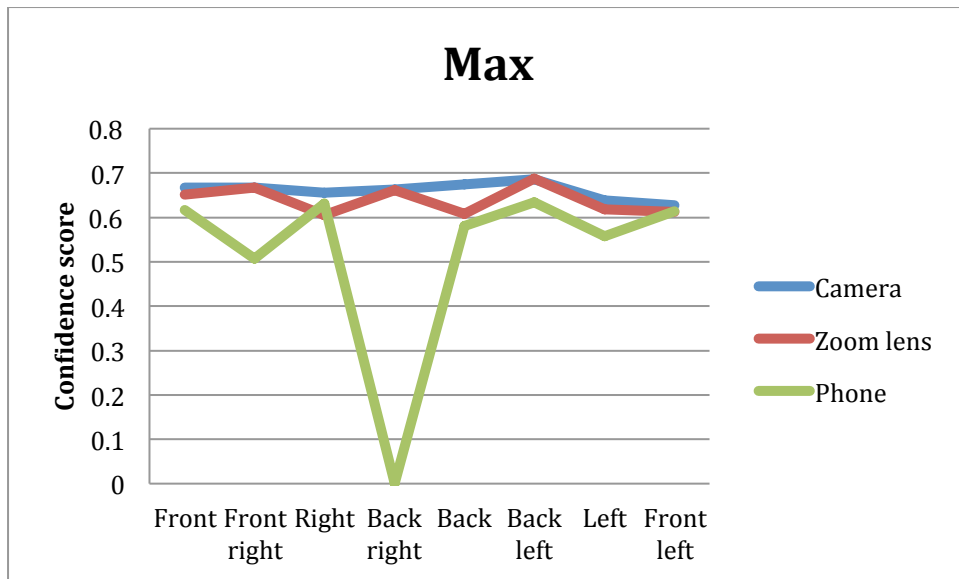


Figure 41: Comparison between subaverage confidence score from different side

Table 8: The Subaverage confidence score from different side

Time:	Camera	Zoom lens	Phone
<b>subAverage</b>			
<b>Front</b>	0.609207917	0.56511525	0.573439143
<b>Front right</b>	0.595013417	0.614106083	0.5035015
<b>Right</b>	0.589305462	0.550236	0.614510667
<b>Back right</b>	0.611898615	0.6015724	0
<b>Back</b>	0.650585933	0.574450571	0.548974
<b>Back left</b>	0.653271467	0.669972667	0.58787375
<b>Left</b>	0.586999583	0.569889857	0.548011333
<b>Front left</b>	0.57873	0.5619674	0.613655

From figure 41, by not calculating the non recognize image, they are a great improvement for phone but overall camera and zoom lens are still better. For camera and zoom lens, in the back left side, the confidence score remain the highest, but for phone, right is the highest changing from back left in figure 37. Though from this graph phone confidence score in the front left seem to over take the other two but it is not reliable due to low sample size.



**Figure 42: Comparison of maximum confidence score from different side**

**Table 9: The Maximum confidence score from different side.**

Time: MAX	Camera	Zoom lens	Phone
<b>Front</b>	0.666998	0.651088	0.616924
<b>Front right</b>	0.66792	0.667469	0.50638
<b>Right</b>	0.655608	0.604576	0.631658
<b>Back right</b>	0.66293	0.661661	0
<b>Back</b>	0.673838	0.607947	0.58064
<b>Back left</b>	0.68623	0.687954	0.634185
<b>Left</b>	0.639103	0.617972	0.558203
<b>Front left</b>	0.627438	0.612165	0.613655

Figure 42 for Maximum confidence score from nearly all side camera devices thrives except for back left with which zoom lens is better. As usual, phone ranked last in most areas. Phone's back right side has zero recognition, therefore maximum confidence score for phone is zero in the back right. The highest confidence score is 0.6879 from back left side via zoom lens.

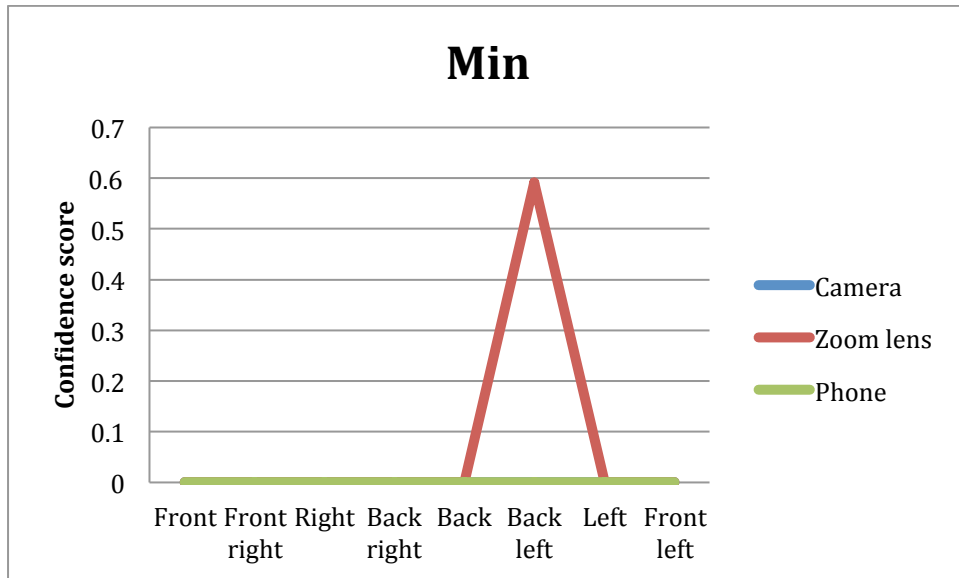


Figure 43: Comparison of minimum confidence score from different side

Table 10: The Minimum confidence score from different side.

Time: MIN	Camera	Zoom lens	Phone
Front	0	0	0
Front right	0	0	0
Right	0	0	0
Back right	0	0	0
Back	0	0	0
Back left	0	0.591191	0
Left	0	0	0
Front left	0	0	0

Figure 43 noted in this graph, these are the Minimum and the absolute Minimum counting the non-recognize value as well that is zero. That is why most of the sizes are zero, because in one of the time interval from 6:00 to 20:00, they are a single time period where the image was not recognize, therefore the Minimum value show zero except for back left in zoom lens which during any interval of time the image was recognized by the program.





Figure 44: Comparison of minimum detect confidence score from different side

Table 11: The Minimum Detect confident score from different side

Time: MIN	Camera	Zoom lens	Phone
<b>Detect</b>			
<b>Front</b>	0.517646	0.512745	0.51425
<b>Front right</b>	0.513637	0.559093	0.500623
<b>Right</b>	0.503145	0.515869	0.610206
<b>Back right</b>	0.544783	0.557374	0
<b>Back</b>	0.527338	0.54784	0.517347
<b>Back left</b>	0.561535	0.591191	0.517918
<b>Left</b>	0.517203	0.514563	0.539612
<b>Front left</b>	0.516993	0.506502	0.613655

Figure 44 for minimum detect, it mean that it will take into account only the value of detected which therefore all the non detected image will not be calculate into the minimum value. Similarly with the subaverage the phone confidence score increase to match the value of the other two devices.

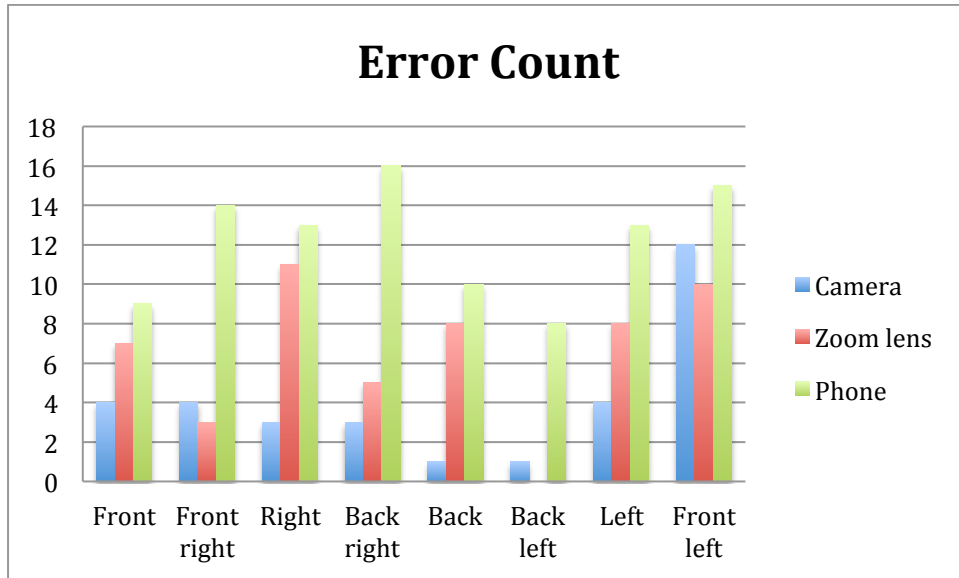


Figure 45: Comparison of number of non-detect image from different side

Table 12: The number of non-detect image from different side.

Time: Error Count	Camera	Zoom lens	Phone	
Front		4	7	9
Front right		4	3	14
Right		3	11	13
Back right		3	5	16
Back		1	8	10
Back left		1	0	8
Left		4	8	13
Front left		12	10	15

Figure 45 shows the number of non-recognize image on each side from 6:00 to 20:00, which consist of 16 images for each side. The lesser the number of count the better the recognition. From the table above, the least count is zoom lens at back left that have 100percent recognition followed by camera back and back left which have only one time frame that the program does not recognize. As for phone on the back right that have 16 count means it have zero recognition from the program.

### 4.7.3 Effect of lens on confidence score

Different type of image taking device will have some effect on the confidence score as well even though the ratio of the image uses in the learning phrase are equal. During the learning phrase 2000 image were uses 1000 from camera which include 500 from normal lens and 500 from zoom lens, and 1000 image from phone. Which accounted to 1:1 ratio of camera to phone image.

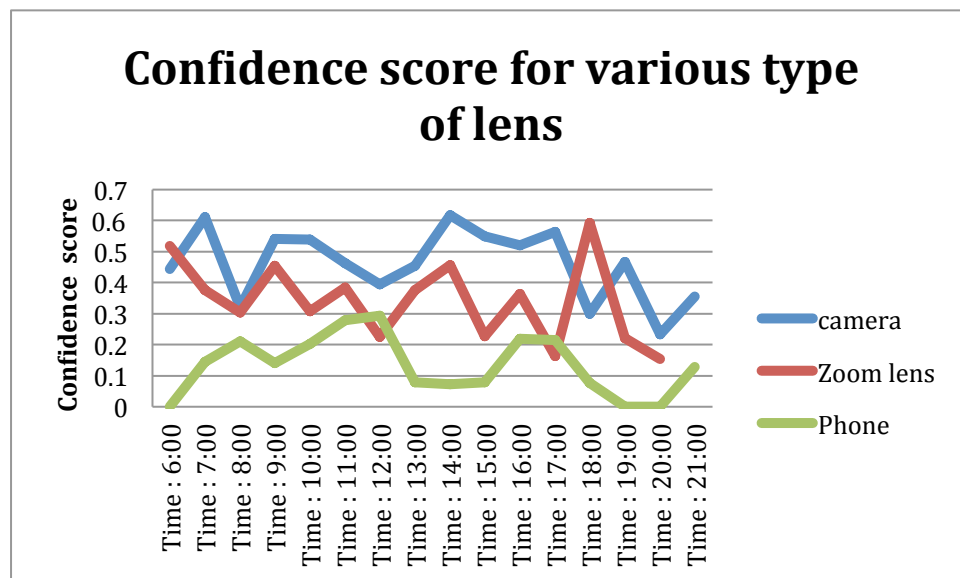


Figure 46: Comparison of confidence score at different time from different lens

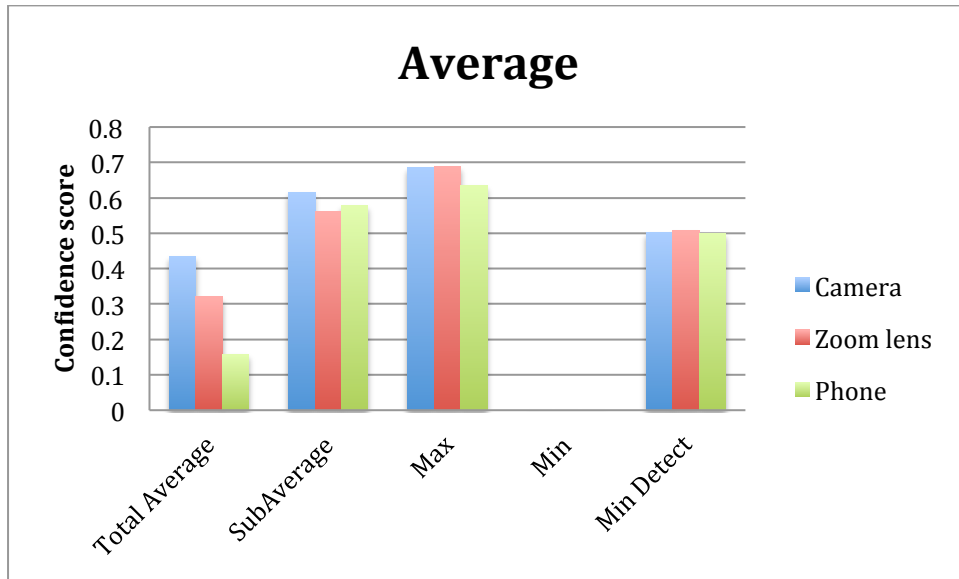
From figure 46, it shows the difference in confidence score from three types of device. In average, the order from the highest to lowest confidence score is camera, zoom lens and phone. Sometime zoom lens outclass camera confidence score during various time such as 6:00 and 18:00 similarly with phone that overtake zoom lens at 12:00 and 17:00. For all line, the fluctuation in confidence score is quite high ranging from 0.02 to 0.4 the highest fluctuation is in zoom lens from 17:00 to 18:00.

The reason why the phone camera image have low confidence score as well as low recognition rate are because phone camera does not have manual control function such as change in focal length, exposure, shutter speed and ISO. These function help

to increase the brightness of the picture and improve the lighting of the image. These functions are essentials without these function it decrease the overall amount of light thus further decrease the amount of detectable point. The other reason for this is that the amount of picture use during the learning phrase is not enough.

**Table 13: The confidence score at different time from different type of device.**

	<b>camera</b>	<b>Zoom lens</b>	<b>Phone</b>
<b>Time</b>			
<b>Time : 6:00</b>	0.443371125	0.517509125	0
<b>Time : 7:00</b>	0.61162375	0.375104875	0.145442375
<b>Time : 8:00</b>	0.320511375	0.303684375	0.20992525
<b>Time : 9:00</b>	0.541896	0.4546045	0.140188625
<b>Time : 10:00</b>	0.538089875	0.307916	0.20321425
<b>Time : 11:00</b>	0.463362	0.383886875	0.279327125
<b>Time : 12:00</b>	0.394783625	0.2260425	0.2942975
<b>Time : 13:00</b>	0.454349375	0.3756555	0.078445375
<b>Time : 14:00</b>	0.616973375	0.457276125	0.07258
<b>Time : 15:00</b>	0.548484	0.2279615	0.07891175
<b>Time : 16:00</b>	0.52059275	0.363143125	0.218235875
<b>Time : 17:00</b>	0.564589125	0.162581625	0.21503775
<b>Time : 18:00</b>	0.30039375	0.591670625	0.076706875
<b>Time : 19:00</b>	0.465977	0.221082375	0
<b>Time : 20:00</b>	0.2327405	0.153965	0
<b>Time : 21:00</b>	0.355148		0.12757875



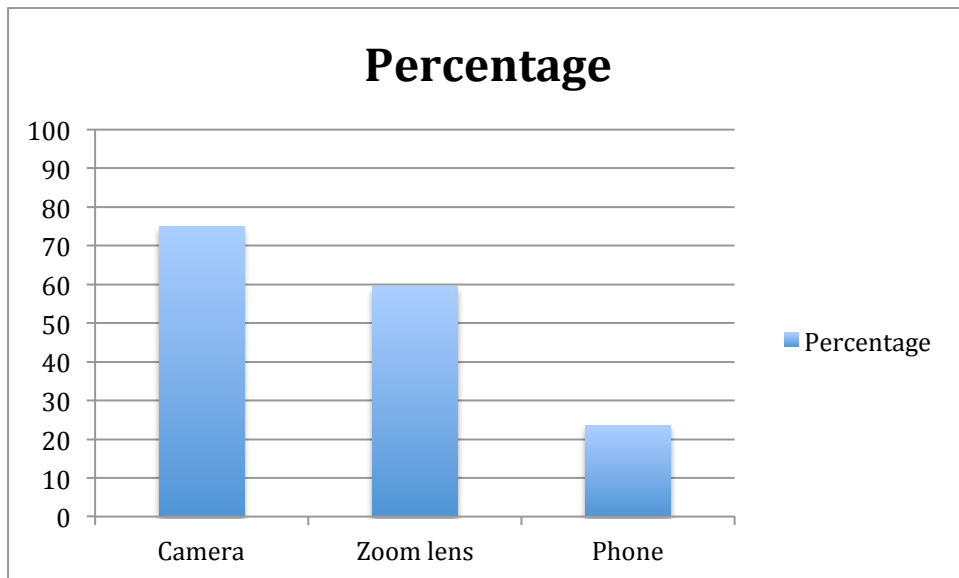
**Figure 47: Comparison of average, subaverage, maximum, minimum, minimum detect confidence score from different lens.**

**Table 14: The confidence score from different type of device as well as the average, subaverage, Maximum, Minimum and Minimum Detect for Phone.**

	Camera	Zoom lens	Phone
<b>Total Average</b>	0.433699154	0.320130258	0.15718368
<b>SubAverage</b>	0.614270933	0.560222431	0.577718609
<b>Max</b>	0.68623	0.687954	0.634185
<b>Min</b>	0	0	0
<b>Min Detect</b>	0.503145	0.506502	0.500623

From figure 47 and table above, these are the total average value where all of the confidence scores of each side and at each different time is average out into one value in order to simplified and compare with 3 different types of device uses in this experiment. In total average camera have the highest confidence score with 0.433 followed zoom lens with 0.32 and the last phone with only 0.157. For max the ranking changes with camera on the top followed phone and last is the zoom lens.

Minimum detect nearly equal for all the three device which is around 0.5, that is because if it goes below 0.5, the program will treat the image not recognize. From figure 48, it is showing the percentage of the recognize image. In total, 16 times interval were taken from 6:00 to 20:00 and 8 sides were test, therefore in total of 128 were test for each device. This graph below shows which type of device image that have the highest recognition with regardless of its confidence score. For AR application the performance of phone camera must be improve to the point that it should have recognition rate of more than 90%.



**Figure 48: Comparison of detection percentage from different type of lens**

## 4.8 Nutella

Table 15: The confidence score of Nutella with different background

	Background 1	Background 2	Background 3
	Confidence score	Confidence score	Confidence score
<b>Front</b>	0.577954	0.561497	0.592187
<b>Front right</b>	0.584419	0.554627	0.641013
<b>Right</b>	0.607886	0.570355	0.646246
<b>Back right</b>	0.553251	0.598092	0.61534
<b>Back</b>	0.509026	0.576737	0.604808
<b>Back left</b>	0.564763	0.585729	0.694142
<b>Left</b>	0.571368	0.573386	0.5984
<b>Front left</b>	0.556165	0.608951	0.618928
<b>Top</b>	0.981735	0.99043	0.957696
<b>Bottom</b>	0.969403	0.978231	0.958322

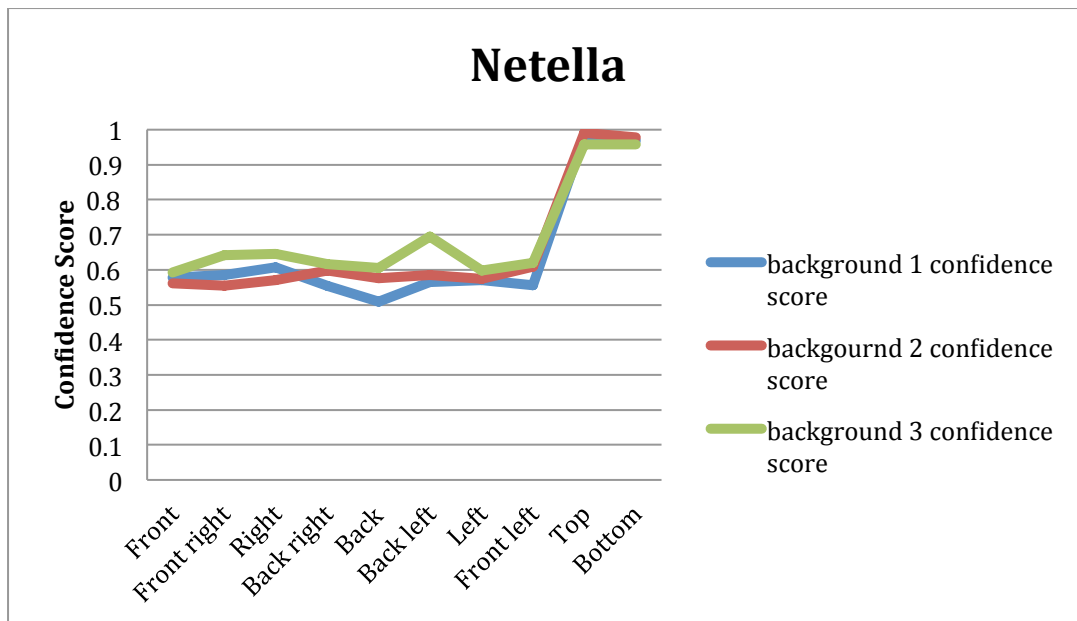


Figure 49: Comparison of detection percentage from different type of lens

For testing of Nutella, three type of background are used, first type is white background, second type is dark background that the Nutella is also placed under the table to create a illusion of low lighting room, and lastly background with high noise means that many object are placed in the background to trick the program to miss recognize the object. The highest confidence score are on the top and the bottom of the Nutella and the lowest confidence score are on the back. Around the side of the Nutella bottle, confidence score are ranging from 0.5 to 0.7 and the top and bottom are ranging from 0.9 to 1. Table 15 notices that from all side and all type of background, all images can be recognized by the program, but the recognition rate is not actually 100%. If anything blocking between the Nutella and the camera, the program will not be able to recognize the Nutella bottle. Out of hundred images of Nutella are tested only a few images, which cannot be recognized, therefore the recognition rate are over 90%. This is the highest recognition rate ever achieved from this experiment. The reason for the high score in the top and the bottom of Nutella container are probably because of the algorithm of uses by machine learning from IBM Bluemix. This algorithm understand the structure of the object not just remember the image data only. Since it understand the basic structure of the object the top and bottom of the Nutella container are the most simple looking object, the top is a circle and the bottom is an oval therefore it is very easy for the program to understand these simple structure therefore this is the reason why the program are more confident identifying the top and bottom of Nutella container more than the side.



## 4.9 Discussion

As for verification for Visual recognition Bluemix similarly with Vuforia, an experiment was conducted to check the recognition rate for Visual recognition Bluemix. The aim of this experiment is to analyze whether recognition rate is high enough for practical uses or not. There are two object uses for this experiment, which are the Nutella container and Fukuzawa Bronze statue.

As for Nutella the learning image was taken in a control environment and the scanning phrase was also conducted in a control environment. The element that is control in this control environment is lighting, in total 500 images of Nutella were used. This experiment was conducted in a well-lit room, constant lighting throughout the room. As for the result the Nutella container can be recognized by the Visual recognition Bluemix program from various direction including the top and the bottom of the Nutella container. The recognition rate for this Nutella bottle are over 90%, which is enough for practical uses.

As for Fukuzawa bronze statue, it cannot be moved, so it is impossible to conduct the learning phrase in the control environment therefore the learning picture was taken outside with various uncontrollable variable, in total of 2100 image were used for the learning phrase 500 images from camera with normal lens, 500 images from camera with zoom lens, and 1000 image from Iphone camera. Although there are various uncontrollable factors such as amount of light and direction of the light, The Visual recognition Bluemix program still can be recognize Fukuzawa bronze statue. Three types of images were used in this experiment to check the recognition rate for each type of images. The three types of images consist of normal camera images, zoomed camera images and Iphone images. These three types of image have different recognition rate, the recognition rate are as followed 75% for normal

camera, 60% for zoomed images and only 25% for Iphone images. As for practicality, it is too low to be used in the real world.

In order for AR application to perform better phone camera performance must be improve. Since AR application are relying on mobile phone camera. To improve the confidence score as well as recognition rate the amount of phone picture used during the learning should been increases these increasing picture will help the machine learning program to adapt and learning to recognize picture from mobile phone better.

The experiment result is that the recognition rates are too low for practical uses, but as overall the program does its job, able to recognize the object as well as able to show augmented reality on the screen as a description of the object.

# Chapter 5 Discussions

## 5.1 Prototype System

This program was created to test out the functionality of the API in the real world environment as well as to be used in the experiment

### 5.1.1 Webprogram

#### 5.1.1.1 learning

#### Classifiers

	classifier_id	name	status	description	
1.	<a href="#">fukuzawa3202100_1145672675</a>	fukuzawa 320 2100	ready	it is a Fukuzawa bronze	<a href="#">edit</a>   <a href="#">delete</a>   <a href="#">test</a>
2.	<a href="#">lipbalmcurel_236326994</a>	lip balm curel	ready	Curel lip balm price 1000 yen	<a href="#">edit</a>   <a href="#">delete</a>   <a href="#">test</a>

[New](#)

**Figure 50: The first page of the web program.**

This first page allows the user to choose pre existing classifier or create new one. The user can also choose this pre-existing classifier to test out the result as well as change the description of the classifier.

#### New Classifiers

Name:

##### Positive

Class1:  File (.zip):  No file chosen

Class2:  File (.zip):  No file chosen

Class3:  File (.zip):  No file chosen

Class4:  File (.zip):  No file chosen

Class5:  File (.zip):  No file chosen

##### Negative

- File (.zip):  No file chosen

**Figure 51: The Creating interface of the web program.**

This is new classifier page, which allows the user to create their own classifiers, the requirements for creating are Name of the classifier, positive image and negative image.

### 5.1.1.2 Testing

## Testing : lipbalmcurel\_236326994

Image:  No file chosen

with default Classifiers:

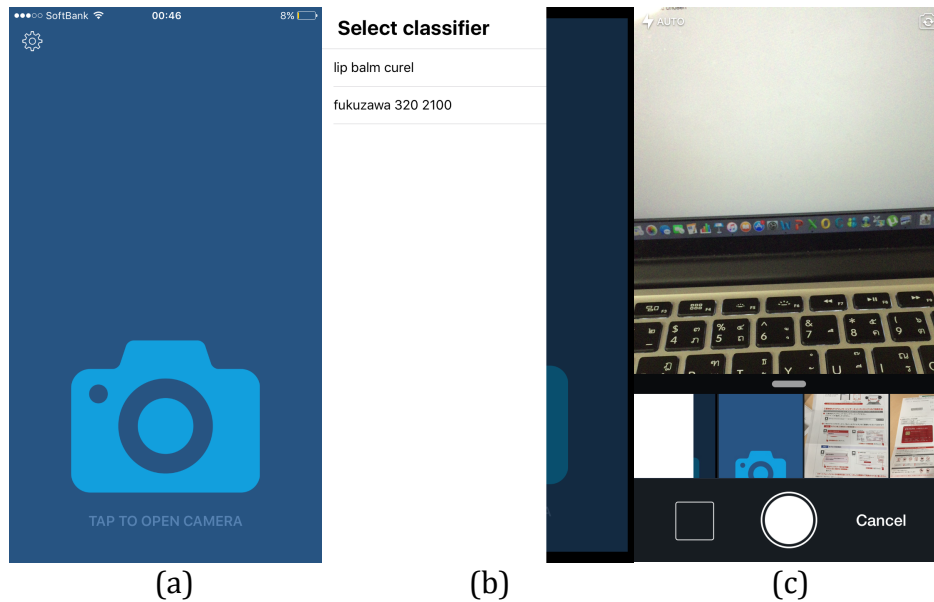
Figure 52: The testing interface.

As for testing, after users have selected a classifier, the program will direct the user to this page. In this page the user can choose any picture file and upload to the cloud system. After that, the picture of the result will be shown on the page as in figure 53 below.

```
3
4 {
5   "images": [
6     {
7       "classifiers": [
8         {
9           "classes": [
10            {
11              "class": "holding",
12              "score": 0.994637
13            }
14          ],
15          "classifier_id": "lipbalmcurel_236326994",
16          "name": "lip balm curel"
17        }
18      ],
19      "image": "image_1467647527.jpg"
20    }
21  ],
22  "images_processed": 0
23 }
```

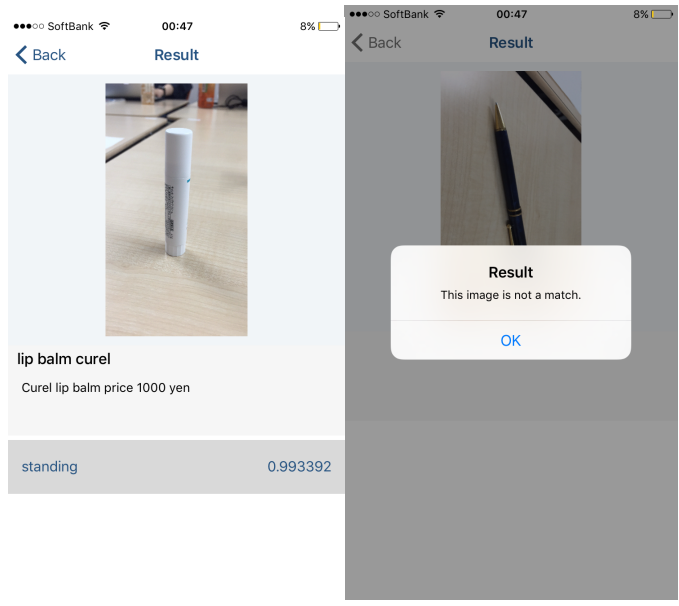
Figure 53: The result of the test.

## 5.1.2 Mobile App



**Figure 54:** These pictures show different interface of the mobile app program.

Figure 54 shows the interface of the mobile app. Once the user click on the app icon, figure 54(a) will appear. The user can touch on the gear icon on the top left hand side to choose the classifier that user have created before in the web application as shown in figure 54(b). Figure 54(c), this page shown that the program is ready for the user to take the picture for the scanning or the user can pick any picture from the storage. If the picture is positive (matches with the data base), it will show the confidence score as well as the description of the object as shown in figure 55(a) if not the app will display no match as a result shown in figure 55(b)



(a)

(b)

**Figure 55: The positive and negative result**

## 5.2 Limitation and Future Work

As the machine learning is quite new area in the programming arena, it still has many limitations. The imitation found in this research is that the application requires a load of images in order to recognize the interested objects. Otherwise, the accuracy is relatively low. Moreover, in the case of actual practice, uncontrollable factors, particularly when the application is used outdoor, need to be considered, for examples; lighting, the background, and the angle deviation. These factors may affect the application on mistakenly object detection. Since this research employed the Fukuzawa Bronze Statue as an only sample, there is a possibility that the application cannot be used for the other exhibits due to the variation of background, angle, natural lighting or other factors.

Although the machine learning has not been greatly developed, and some area is still complicated, this area will be evolved somehow and machine learning will be more practical than nowadays. Moreover, the utilization will vary, not only for museum usage, but also for other purposes. To illustrate, it can be adapted to tracing and tracking application used in the supermarket, factory or other situation when the AR markers cannot be installed. However, it is expected that the future research can contribute from this study, and the human lifestyle will be easier and more convenient due to that development.

## 5.3 Validation

### 5.3.1 Description

2016/__/__	
Graduate School of System Design and Management Keio University	
<u>Experiment Consent</u>	
Experiment Title:	Validation of Markerless Augmented Reality app
Experiment Supervisor:	Prof. Ogi Tetsuro
Experiment Content:	
The experiment is conducted to validate this markerless augmented reality App. The participate will be using this application to take a picture of an object from any direction in any condition suitable for the users and the program will shows that it recognize the object or not. If it recognizes the program will shows the user information about the object as well as the confidence score. The user will then asked to fill a questionnaire about their experience about the program	
Date:	2016/__/__
Handling of data: The obtained data is use only for academic purpose. Data recorded from the experiment contain gender, age, and measured data in questionnaire where participant's name is not specified.	
I understand content above and agree to participate the experiment.	
	Date: 2016/__/__
	Signature
	_____



### 5.3.2 Questionnaire

#### Participant's detail

Have you know about Augmented reality before: Yes / No

Age: \_\_\_\_\_ Gender: Male / Female

**Please tick the following boxes with to rate the following statements from 1-5**

**base on your opinion.** (1=strongly disagree, 2= disagree, 3=I think so, 4=agree & 5=strongly agree)

Statement	Points				
	1	2	3	4	5
1. The program easy to uses					
2. Font size is easy to read					
3. It is easy to change classifier					
4. It is easy to access image storage in your mobile phone					
5. The program can identified the object from all direction					
6. The program identified the object correctly					
7. The program can identified the images quickly					
8. The confidence score is high					
9. This program improve how you identified thing					
10. This program give you more information about the object more than the package information					
11. This program help you identified thing faster					
12. How many picture you have taken in total (Write down the number)					

**Comments/Opinions:**

---



---

### 5.3.3 Questionnaire Result

Regarding to the problems mentioned in the second chapter, a demo application is presented as an alternative solution for museums usage when the AR markers cannot be used. In order to obtain the validation, a survey was employed in Keio University. Overall 10 participants used the application ( $n = 10$ ), and tested if it can literally work and satisfy them or not. The measurement is the questionnaire employing Likert scale. The following table shows the overview of means divided by gender.

**Table 16: Result from ten participant**

Report											
Gender	1	2	3	4	5	6	7	8	9	10	
Male	Mean	4.6667	4.1667	4.8333	4.5000	4.6667	5.0000	4.1667	3.3333	3.5000	3.5000
	N	6	6	6	6	6	6	6	6	6	6
	Std. Deviation	.51640	.75277	.40825	.54772	.51640	.00000	.40825	1.03280	.54772	.83666
Female	Mean	4.2500	4.7500	4.7500	4.2500	4.7500	4.5000	5.0000	3.0000	3.2500	3.7500
	N	4	4	4	4	4	4	4	4	4	4
	Std. Deviation	.50000	.50000	.50000	.50000	.50000	.57735	.00000	.81650	1.25831	.95743
Total	Mean	4.5000	4.4000	4.8000	4.4000	4.7000	4.8000	4.5000	3.2000	3.4000	3.6000
	N	10	10	10	10	10	10	10	10	10	10
	Std. Deviation	.52705	.69921	.42164	.51640	.48305	.42164	.52705	.91894	.84327	.84327

According to the table, there are 10 questions asking about the respondents' satisfaction to the application, ranging from 1 point for strongly disagree to 5 points for strongly agree. The results of questionnaires reveal that the average mean is 4.78 ( $M=4.78$ ). This value is statistically acceptable. Thus, the proposed application is valid.

As for question eight, nine and ten the overall score is quite low comparing to the other question. Question eight the question is the confidence score is high, since

the experiment was conducted on the nutella bottle which have low confidence score on the side which is maximum of 0.69 therefore most of the user felt that the confidence score is quite low.

As for question nine the question state that this program improve how you identified thing some people feel that they it does not change how they identified things this feeling is quite subjective from person to person therefore this score varies.

Lastly as for question ten the question is this program gives you more information about the object more than the package information since the information is this program currently is quite low therefore some user thinks that the label give them more information than the application program.

## 5.4 Difference Between Image Matching Method and Machine Learning Method

The difference between these two methods is how the program matches things. The image matching is a tool which matches the scanned image to the nearly similar image from the storage data. The learning in image matching is to store multiple images. The Machine Learning method works differently, this method uses to understand the structure of the object by multiple images given to the program prior to scanning. Learning for machine learning is to use multiple pictures from various directions to understand the basic structure of that object; therefore, the machine learning does not need to store those images but it stores and learns the structure of the object in the image instead. So the image scan by this program is then matched to the nearest similar structure instead of images like how image matching works. To put it in simple terms, image matching matches picture to picture but machine learning matches picture to structure.

## **5.5 Benefits and Advantages**

### **5.5.1 Image Matching Method**

The benefit for this image matching is that it has high accuracy as well as recognition for object with pattern on it. It works perfectly under control environment. The other advantage of this method is that it able to track the orientation of object therefore it is able to project 3D image on the surface of the object.

So the best condition for using this image matching is for indoor environment such as indoor museum, super market and mall. The best object that is suitable for image matching are objects that have lot of pattern on them such as painting, food packing and boxes with label on it.

### **5.5.2 Machine learning Method**

The benefit for this machine learning is that it can be use on any object with no retraction. It also works with object without pattern and it also work with outdoors object where environment cannot be control. It also required lesser space in the data base due to it does not need to store the image of the object but it learn the basics structure of the object from the image and then store it therefore the picture can be remove after the learning.

So the best condition for using this machine learning method is for outdoor environment where they are a lots of uncontrollable variable such as outdoor museum, exhibition and park. As for the best object would be complex object with single color scheme such as statue and antique.

## Chapter 6 Conclusion

### 6 Conclusion

In order for the Vuforia program to obtain the best results in processing, there are three methods to increase the overall detectability of an object. First, a low focal lens should be used to take photo in order to create picture marker for augmented reality, since it will be able to focus all the detectable points into the object only. This helps reduce the background noise as mentioned in the Effect of background section, low focal point lens without background have more detectable points when compared with high focal points. Secondly, the properties of the object should have a lot of pattern on it if recognition will be successful. This pattern can help produce more detectable points. Lastly, regarding backlit images, it is not efficient to use backlit image to create picture marker, since the foggy layer will not allow any detectable points to be formed on the image. The result from using Vuforia program to create markerless AR system was a failure, because no matter how many pictures were used to create the 3D marker, it was never enough since it cannot know which direction the user will scan it from. Even if pictures were really taken from all directions, the size of the data will be too large to comprehend for the device and program. Despite the failures, the findings from the Vuforia experiment were a great help for this thesis regarding the background as well as the focal point lens. These findings from Vuforia were built on and used in the second experiment as well.

As for the second method, the machine learning was a success even though IBM Bluemix has changed the algorithm of the visual recognition API during when the experiment was conducted. This change in algorithm has greatly decreased the confidence score as well as recognition percentage. These changes have affected the results but still the overall result is acceptable. The results obtained from this second

method are that the best time with the highest confidence score is around 14:00 for this Fukuzawa Bronze statue, and the best side is the back right which have the highest confidence score as well as 100% recognition rate. Both during the learning phrase as well as the scanning phrase in the night, artificial light should be used in order to improve both the confidence score and recognition rate. The overall recognition rate from camera, zoom lens, and phone are as follows: 75%, 60%, and 25% respectively. As for camera and zoom lens both of these recognition rate is still acceptable, but phone is too low for practical uses. Ignoring these recognition rates, the uses of machine learning to create markerless augmented reality system is considered a success. To improve this recognition rate, more pictures should be taken of the Fukuzawa Bronze statue. The properties of these new pictures should vary, for example; take pictures of Fukuzawa Bronze statue at different weather and season. The more the varieties of the image, the easier it is for the machine learning to understand them. This will give more types of information for the machine learning, which will further improve both the confidence score as well as recognition rate. Similarly to the proviouse experiment with Vuforia, backlit images should be ignored as well. This type of image does not contribute to the learning as well as these backlit image cannot be recognized during the scanning phrase as well. Low recognition rate on the phone aside overall experiment to create markerless AR systems was a success.

Form museum application since in museum both outdoor and indoor most of the antiques are fragile and weak therefore it is nearly impossible to trample those object with AR marker. The place where AR marker cannot be uses this markerless system can solve that problem. By using this markerless augmented reality system it will make the augmented reality possible for obeject where AR marker cannot be

place such as old antique. This machine learning method can also provide augmented reality for complex single scheme color object such as statue and metallic utensile as well. It also works for outdoor object where they are lots of uncontrollable variables such as amount of sunlight and direction of the sunlight.

Lastly about the difference between two methods are how they compare the image obtained to the data stored in data based. The image matching method is that it matches the scan image with the nearest image similar image in the data based but for machine learning it matches the scan image with the nearest similar structure object in the data based.



# Chapter 7 Reference and Acknowledgement

## 7.1 Reference

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., & Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675-678). ACM.

Koch, C., Neges, M., König, M., & Abramovici, M. (2014). Performance study on natural marker detection for augmented reality supported facility maintenance. In *Australasian Journal of Construction Economics and Building-Conference Series, Australia*.

Weng, E. N. G., Khan, R. U., Adruce, S. A. Z., & Bee, O. Y. (2013). Objects tracking from natural features in mobile augmented reality. *Procedia-Social and Behavioral Sciences*, 97, 753-760.

Srivastava, N., & Salakhutdinov, R. R. (2012). Multimodal learning with deep boltzmann machines. In *Advances in neural information processing systems*(pp. 2222-2230).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Ciresan, D., Meier, U., & Schmidhuber, J. Multi-column Deep Neural Networks for Image Classification Supplementary Online Material. *Descente du gradient*,12, 13-18.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1912-1920).

Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE Computational Intelligence Magazine*, 5(4), 13-18.

Koch, C., Neges, M., König, M., & Abramovici, M. (2014). Natural markers for augmented reality-based indoor navigation and facility maintenance. *Automation in Construction*, 48, 18-30.

IBM Watson (2016), Overview of the Visual Recognition Service (n.d), Retrieved April 27, 2016 from <http://www.ibm.com/watson/developercloud/doc/visual-recognition/>

Koch, C., Neges, M., König, M., & Abramovici, M. (2014). Natural markers for augmented reality-based indoor navigation and facility maintenance. *Automation in Construction*, 48, 18-30.

LeCun, Y., Huang, F. J., & Bottou, L. (2004, June). Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (Vol. 2, pp. II-97). IEEE.

Torralla, A., Fergus, R., & Weiss, Y. (2008, June). Small codes and large image databases for recognition. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.

Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE transactions on pattern analysis and machine intelligence*, 32(1), 105-119.

Chapelle, O., Haffner, P., & Vapnik, V. N. (1999). Support vector machines for histogram-based image classification. *IEEE transactions on Neural Networks*, 10(5), 1055-1064.

Lazebnik, S., Schmid, C., & Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. 2, pp. 2169-2178). IEEE.

Zhang, L., Lin, F., & Zhang, B. (2001, October). Support vector machine learning for image retrieval. In *Image Processing, 2001. Proceedings. 2001 International Conference on* (Vol. 2, pp. 721-724). IEEE.

Joutou, T., & Yanai, K. (2009, November). A food image recognition system with multiple kernel learning. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 285-288). IEEE.

Russell, B. C., Freeman, W. T., Efros, A. A., Sivic, J., & Zisserman, A. (2006). Using multiple segmentations to discover objects and their extent in image collections. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. 2, pp. 1605-1614). IEEE.

Fergus, R., Perona, P., & Zisserman, A. (2003, June). Object class recognition by unsupervised scale-invariant learning. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (Vol. 2, pp. II-264). IEEE.

## 7.2 Acknowledgement

This Thesis become a reality by the support of many individuals, I would like to express my sincere gratitude to all of them.

Foremost I would like to express my deepest gratitude to my supervisor Professor Ogi Tetsuro for his full support, understanding, expert guidance and encouragement throughout my study and research. Without his incredible patience, time, wisdom, advice and counsel my thesis would have been an impossible task. I also like to thanks my secondary advisor Assistance Professor Kohtake Naohiko for giving advise on my work.

I would also like to thanks Assistant Professor Dr. Manasawee Kaenampornpan and her student Nuttawat Jurdpongsaton and Surasak Srisawangwong for guiding and help me with the creation of my program.

My thanks also go to my fellow graduate student at System Design and Management Department and staff of Keio University Japan. Special thanks goes to all of my friends who supported me and encouraged throughout this 2 years academic exploration. I also like to thanks the Keio University who provide this place, the sacred temple of learning.

Finally, I would like to thanks my parents, brother and sister for their unconditional love, support as well as paying my living and tuition fee. I would not have been able to complete this thesis without their love and encouragement.

# Chapter 8 Appendix

## 8.1 Based On time

### 8.1.1 Time 6:00

Table 17: Showing the confidence score of three type of lens and device at 6:00.

Time : 6:00	Camera	Zoom lens	Phone
Front	0.521687	0.512745	0
Front right	0	0.648553	0
Right	0.624098	0.515869	0
Back right	0	0.585698	0
Back	0.658923	0.59235	0
Back left	0.561535	0.672693	0
Left	0.632519	0	0
Front left	0.548207	0.612165	0

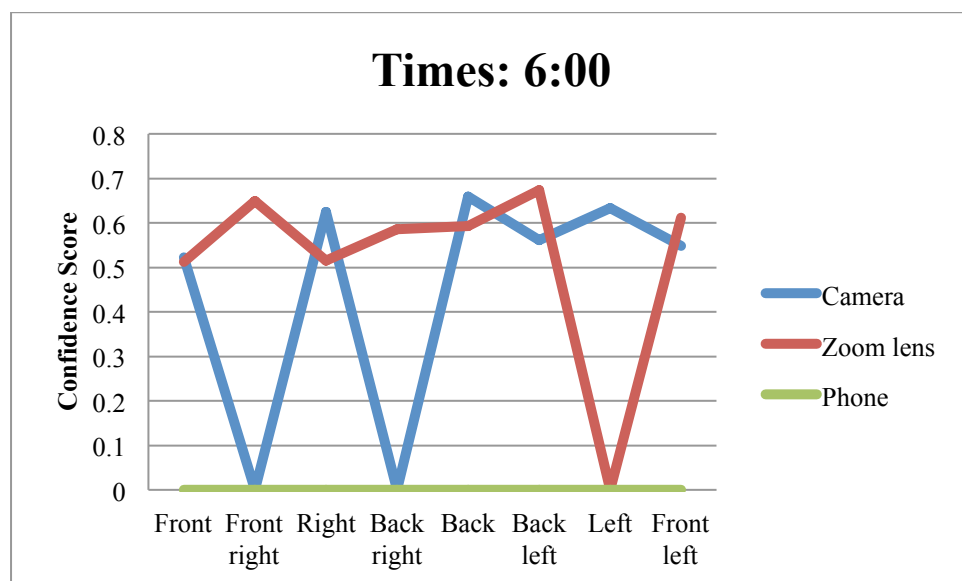
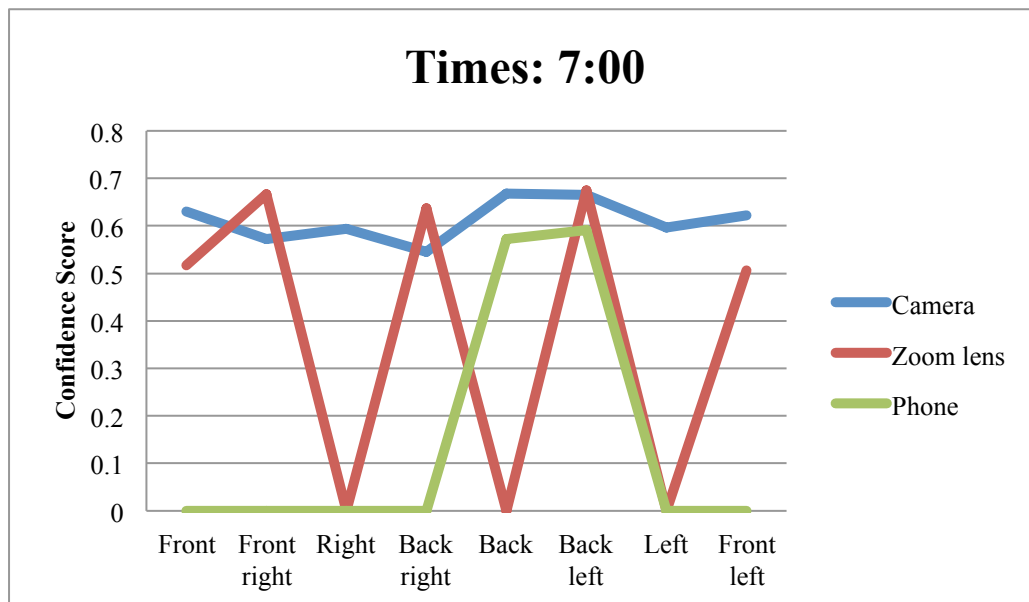


Figure 56: Showing the Confidence score at 6:00

### 8.1.2 Time 7:00

**Table 18: Showing the confidence score of three type of lens and device at 7:00.**

Time : 7:00	Camera	Zoom lens	Phone
Front	0.629894	0.517283	0
Front right	0.572755	0.665917	0
Right	0.593219	0	0
Back right	0.544783	0.637069	0
Back	0.668124	0	0.57286
Back left	0.665146	0.674068	0.590679
Left	0.596787	0	0
Front left	0.622282	0.506502	0
Average	0.61162375	0.375104875	0.145442375
Subaverage	0.61162375	0.6001678	0.5817695

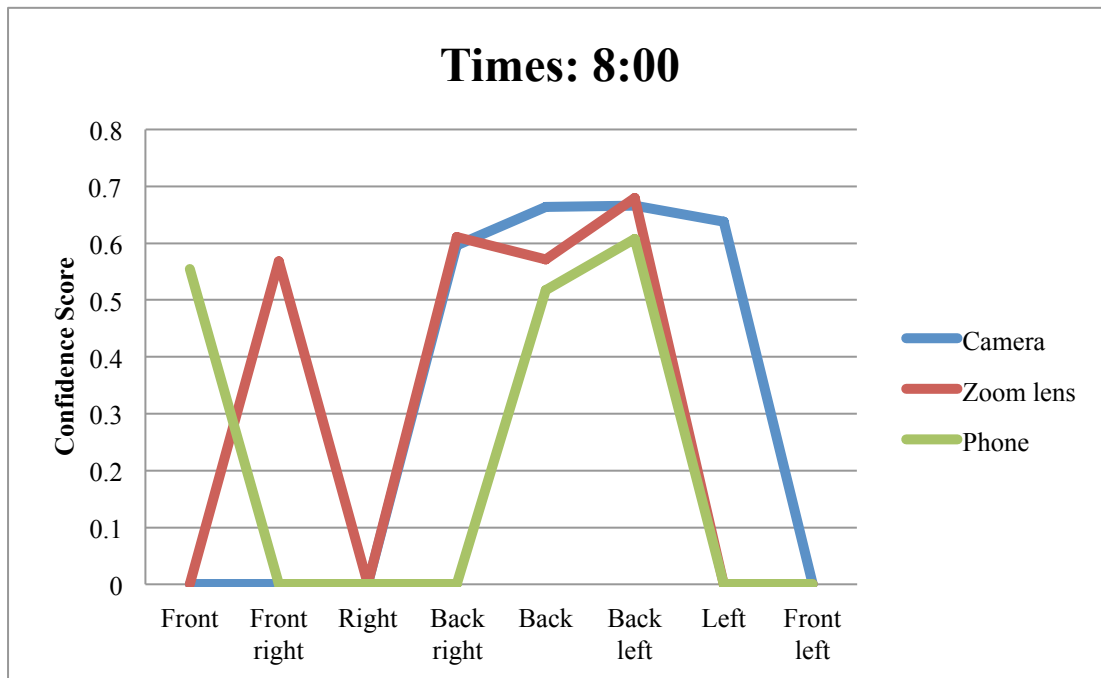


**Figure 57: Showing the Confidence score at 7:00**

### 8.1.3 Time 8:00

**Table 19: Showing the confidence score of three type of lens and device at 8:00.**

Time : 7:00	Camera	Zoom lens	Phone
Front	0.629894	0.517283	0
Front right	0.572755	0.665917	0
Right	0.593219	0	0
Back right	0.544783	0.637069	0
Back	0.668124	0	0.57286
Back left	0.665146	0.674068	0.590679
Left	0.596787	0	0
Front left	0.622282	0.506502	0
Average	0.61162375	0.375104875	0.145442375
Subaverage	0.61162375	0.6001678	0.5817695



**Figure 58: Showing the Confidence score at 8:00**

### 8.1.4 Time 9:00

Table 20: Showing the confidence score of three type of lens and device at 9:00.

Time : 9:00	Camera	Zoom lens	Phone
Front	0.59847	0.595706	0.57529
Front right	0.563159	0.57645	0
Right	0	0	0
Back right	0.623227	0	0
Back	0.665919	0.607947	0
Back left	0.653153	0.682641	0
Left	0.603802	0.563488	0.546219
Front left	0.627438	0.610604	0
Average	0.541896	0.4546045	0.140188625
Subaverage	0.619309714	0.606139333	0.5607545

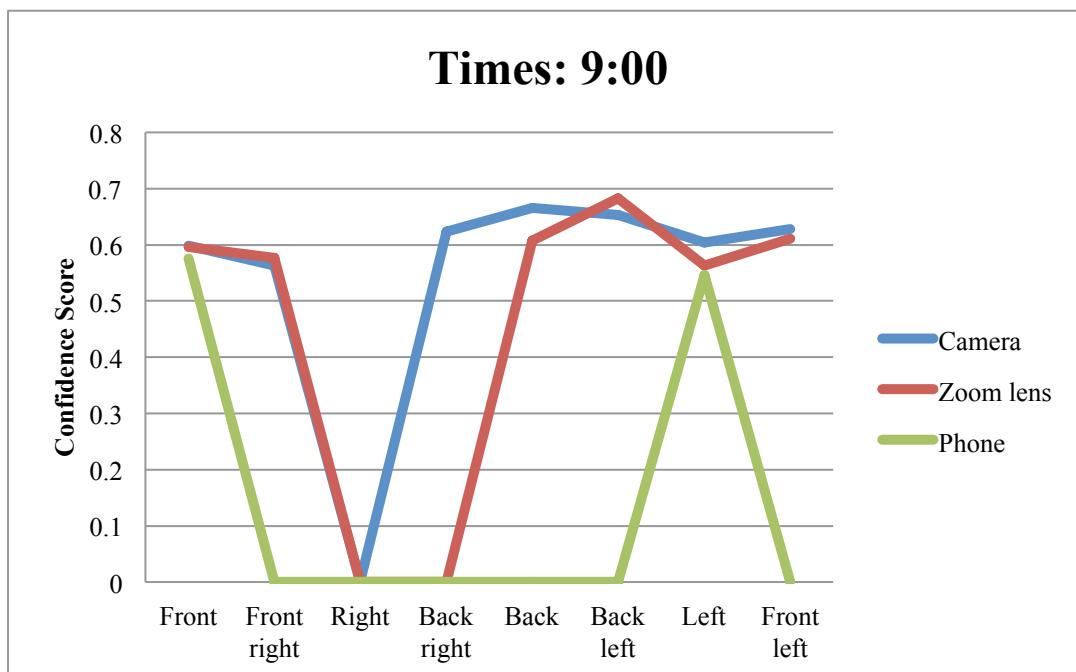
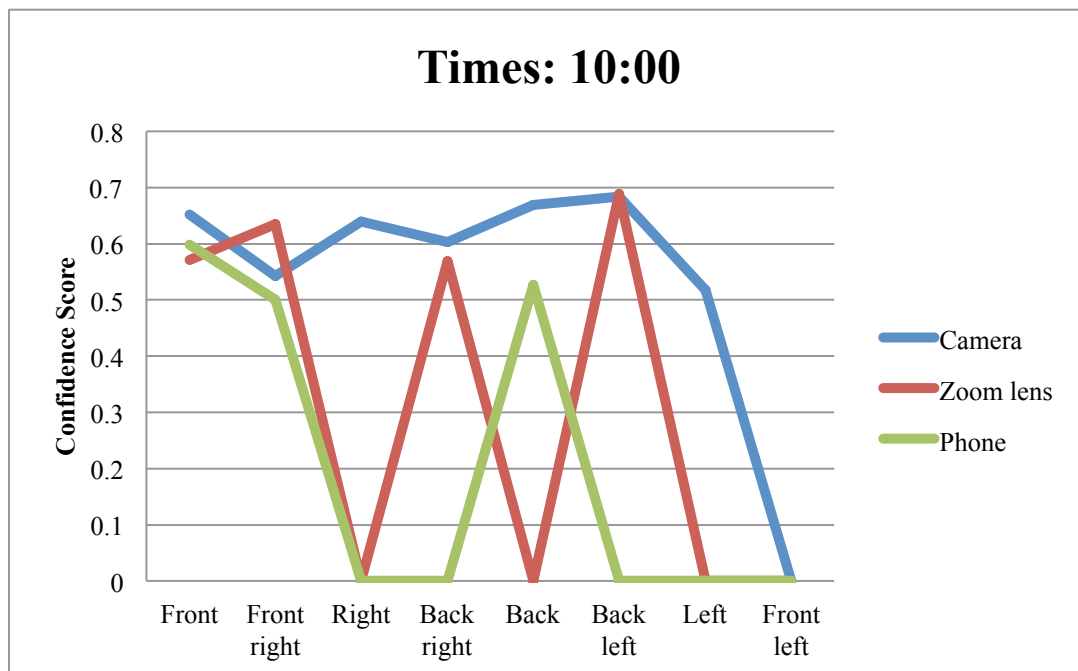


Figure 59: Showing the Confidence score at 9:00

### 8.1.5 Time 10:00

**Table 21: Showing the confidence score of three type of lens and device at 10:00.**

Time : 10:00	Camera	Zoom lens	Phone
Front	0.651695	0.571016	0.598373
Front right	0.542119	0.635191	0.500623
Right	0.639023	0	0
Back right	0.602398	0.569167	0
Back	0.66874	0	0.526718
Back left	0.683541	0.687954	0
Left	0.517203	0	0
Front left	0	0	0
Average	0.538089875	0.307916	0.20321425
Subaverage	0.614959857	0.615832	0.541904667



**Figure 60: Showing the Confidence score at 10:00**



### 8.1.6 Time 11:00

Table 22: Showing the confidence score of three type of lens and device at 11:00.

Time : 11:00	Camera	Zoom lens	Phone
Front	0.647524	0	0.540548
Front right	0.525534	0.6346	0
Right	0.612068	0	0.610206
Back right	0.613867	0.644105	0
Back	0.673838	0.562843	0.565945
Back left	0.634065	0.685873	0.517918
Left	0	0	0
Front left	0	0.543674	0
Average	0.463362	0.383886875	0.279327125
Subaverage	0.617816	0.614219	0.55865425

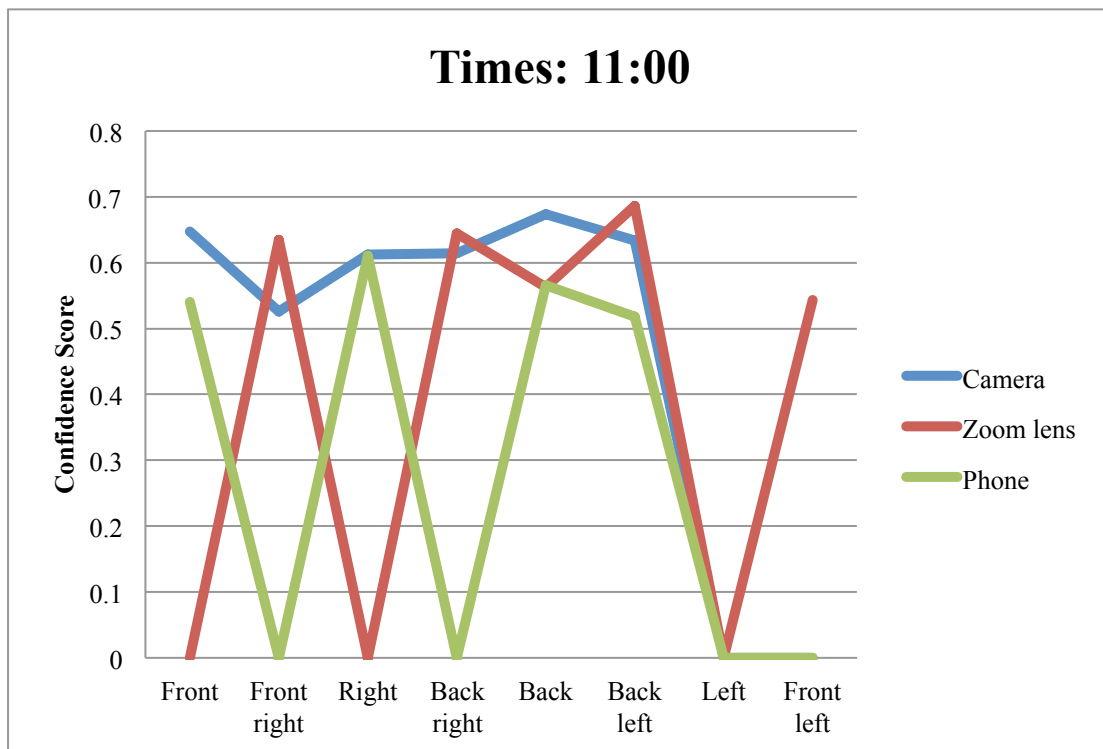
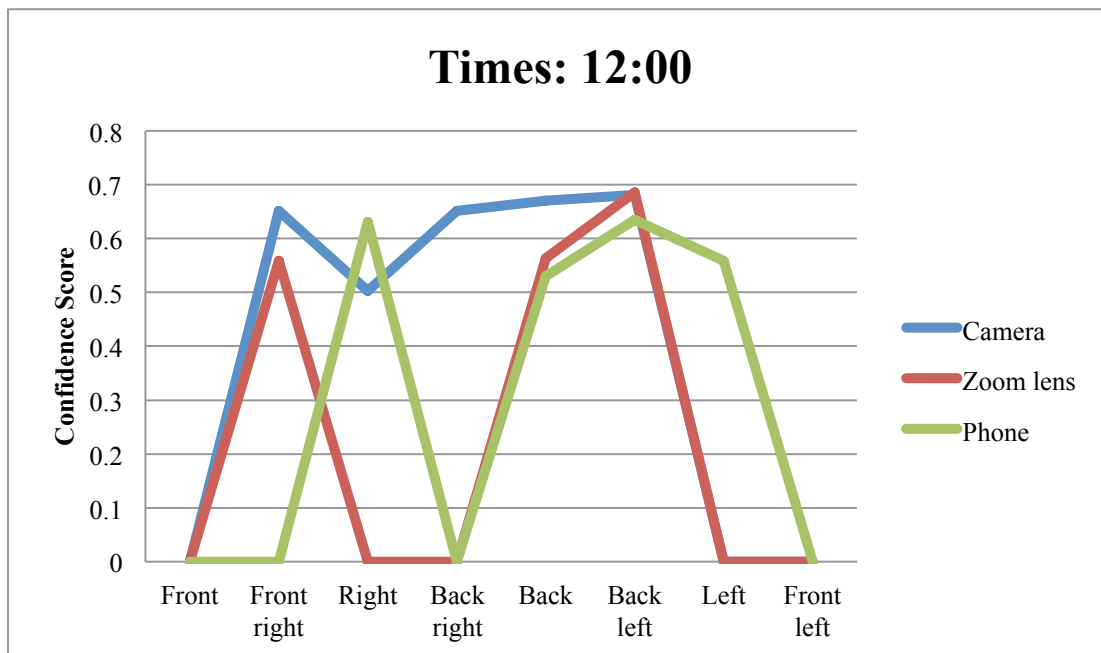


Figure 61: Showing the Confidence score at 11:00

### 8.1.7 Time 12:00

**Table 23: Showing the confidence score of three type of lens and device at 12:00.**

Time : 12:00	Camera	Zoom lens	Phone
Front	0	0	0
Front right	0.651985	0.559093	0
Right	0.503145	0	0.631658
Back right	0.651885	0	0
Back	0.67054	0.562938	0.530334
Back left	0.680714	0.686309	0.634185
Left	0	0	0.558203
Front left	0	0	0
Average	0.394783625	0.2260425	0.2942975
Subaverage	0.6316538	0.60278	0.588595



**Figure 62: Showing the Confidence score at 12:00**

### 8.1.8 Time 13:00

Table 24: Showing the confidence score of three type of lens and device at 13:00.

Time : 13:00	Camera	Zoom lens	Phone
Front	0.620734	0.589544	0
Front right	0	0.625812	0
Right	0.554079	0.519926	0
Back right	0.598861	0.582049	0
Back	0.657022	0	0
Back left	0.678001	0.687913	0.627563
Left	0.526098	0	0
Front left	0	0	0
Average	0.454349375	0.3756555	0.078445375
Subaverage	0.605799167	0.6010488	0.627563

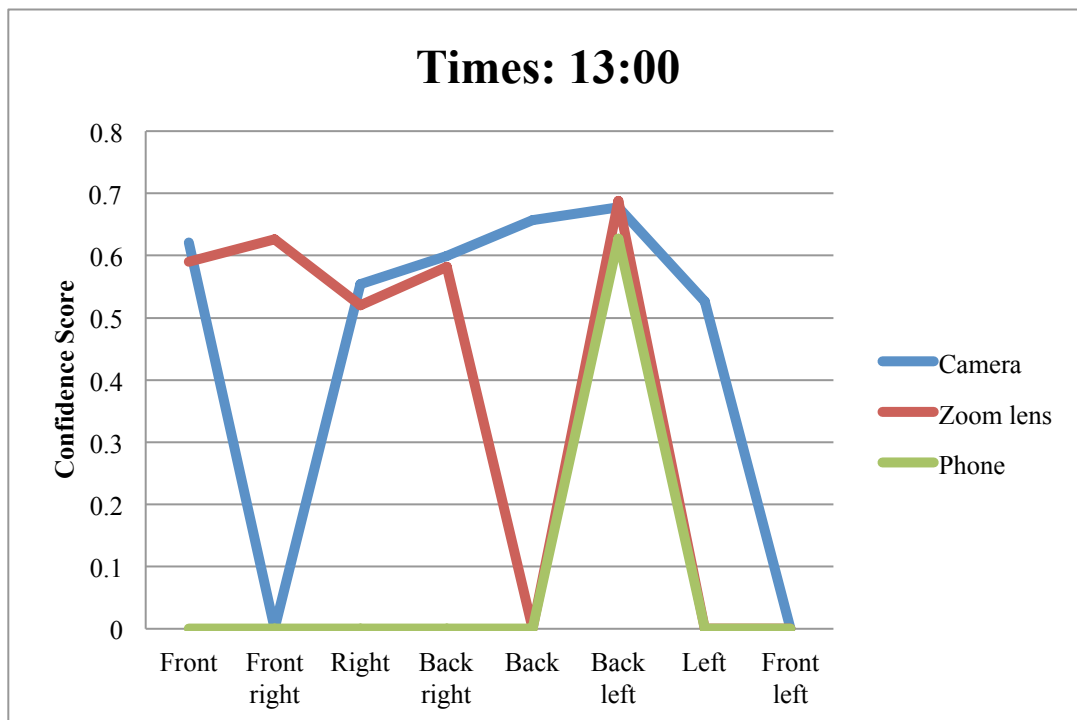
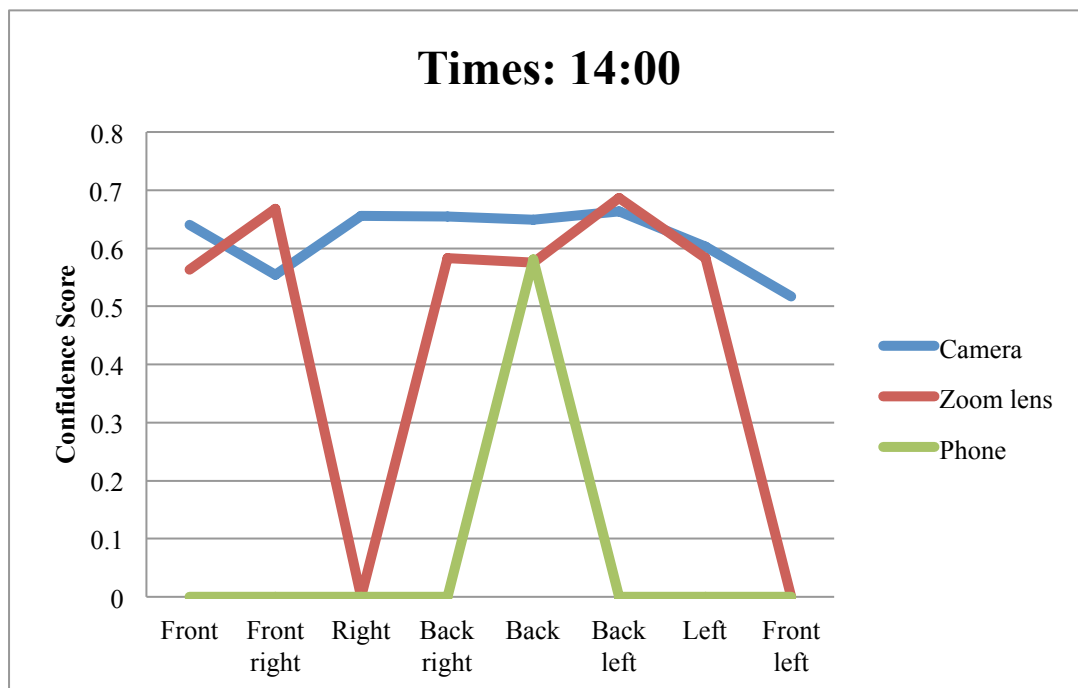


Figure 63: Showing the Confidence score at 13:00

### 8.1.9 Time 14:00

**Table 25: Showing the confidence score of three type of lens and device at 14:00.**

Time : 14:00	Camera	Zoom lens	Phone
Front	0.640304	0.562783	0
Front right	0.554028	0.667469	0
Right	0.655608	0	0
Back right	0.654371	0.58331	0
Back	0.648931	0.575367	0.58064
Back left	0.663159	0.686178	0
Left	0.602393	0.583102	0
Front left	0.516993	0	0
Average	0.616973375	0.457276125	0.07258
Subaverage	0.616973375	0.6097015	0.58064

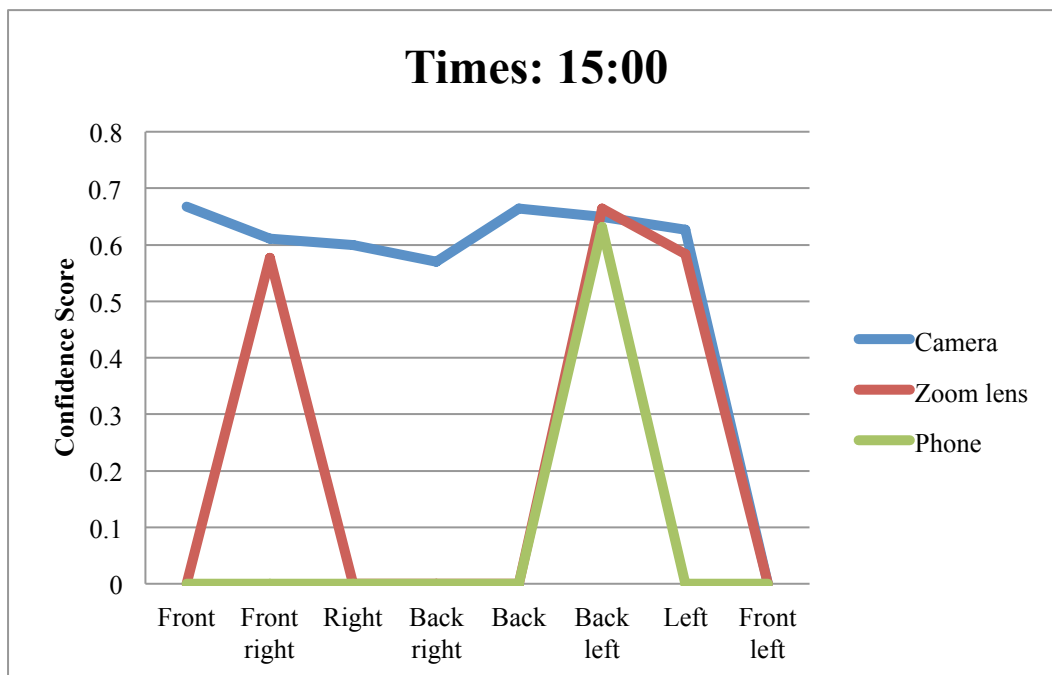


**Figure 64: Showing the Confidence score at 14:00**

### 8.1.10 Time 15:00

**Table 26: Showing the confidence score of three type of lens and device at 15:00.**

Time : 15:00	Camera	Zoom lens	Phone
Front	0.666998	0	0
Front right	0.611167	0.576694	0
Right	0.599706	0	0
Back right	0.569535	0	0
Back	0.663521	0	0
Back left	0.649742	0.663494	0.631294
Left	0.627203	0.583504	0
Front left	0	0	0
Average	0.548484	0.2279615	0.07891175
Subaverage	0.626838857	0.607897333	0.631294

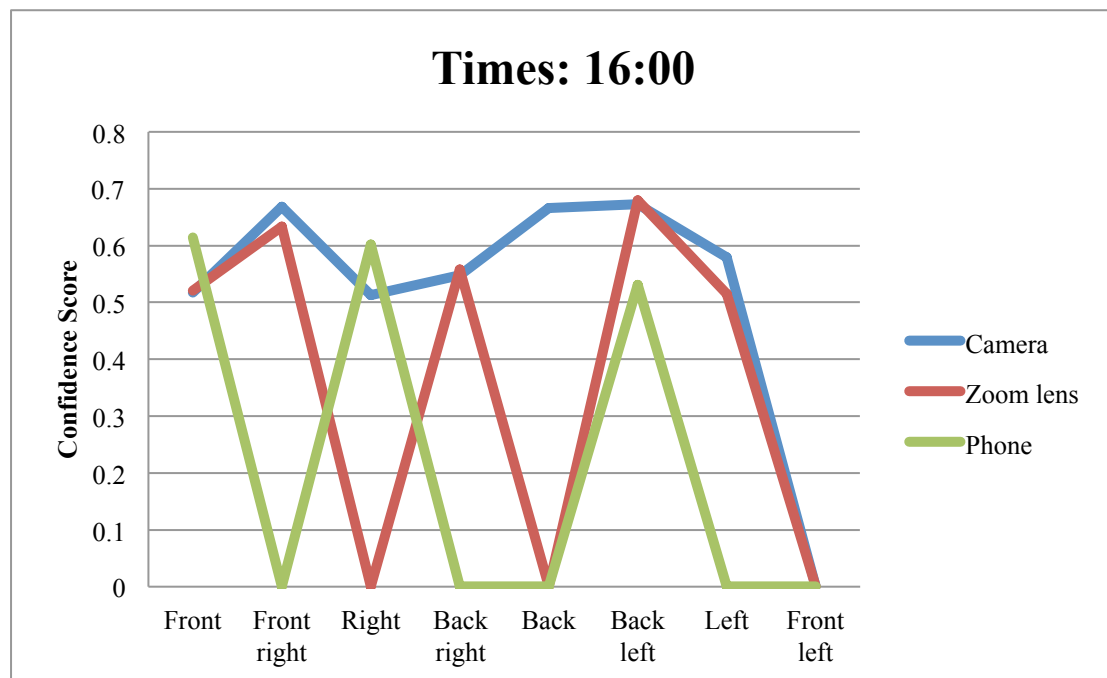


**Figure 65: Showing the Confidence score at 15:00**

### 8.1.11 Time 16:00

**Table 27: Showing the confidence score of three type of lens and device at 16:00.**

Time : 16:00	Camera	Zoom lens	Phone
Front	0.517646	0.520757	0.613689
Front right	0.66792	0.632933	0
Right	0.512741	0	0.601668
Back right	0.547549	0.557374	0
Back	0.666418	0	0
Back left	0.673054	0.679518	0.53053
Left	0.579414	0.514563	0
Front left	0	0	0
Average	0.52059275	0.363143125	0.218235875
Subaverage	0.594963143	0.581029	0.581962333



**Figure 66: Showing the Confidence score at 16:00**

### 8.1.12 Time 17:00

**Table 28: Showing the confidence score of three type of lens and device at 17:00.**

Time : 17:00	Camera	Zoom lens	Phone
Front	0.59877	0	0.616924
Front right	0.641421	0	0
Right	0.633976	0	0
Back right	0.648803	0	0
Back	0.66841	0	0
Back left	0.68623	0.682681	0.563766
Left	0.639103	0.617972	0.539612
Front left	0	0	0
Average	0.564589125	0.162581625	0.21503775
Subaverage	0.645244714	0.6503265	0.573434

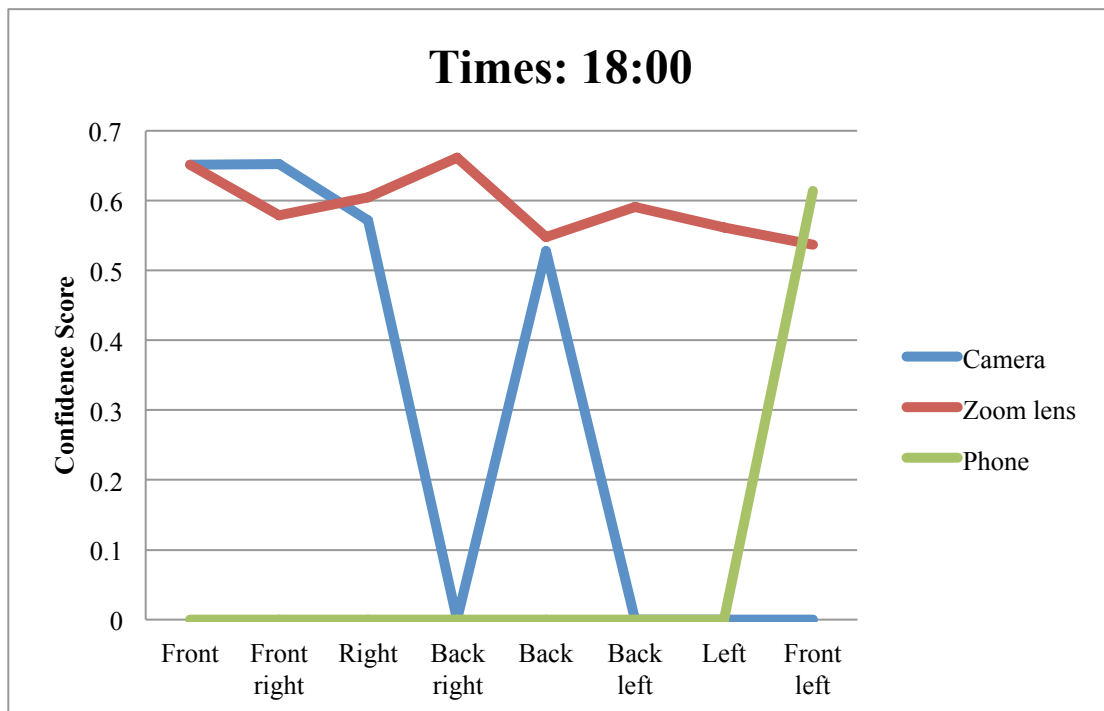


**Figure 67: Showing the Confidence score at 17:00**

### 8.1.13 Time 18:00

**Table 29: Showing the confidence score of three type of lens and device at 18:00.**

Time : 18:00	Camera	Zoom lens	Phone
Front	0.65182	0.651088	0
Front right	0.652649	0.578898	0
Right	0.571343	0.604576	0
Back right	0	0.661661	0
Back	0.527338	0.54784	0
Back left	0	0.591191	0
Left	0	0.561219	0
Front left	0	0.536892	0.613655
Average	0.30039375	0.591670625	0.076706875
Subaverage	0.6007875	0.591670625	0.613655



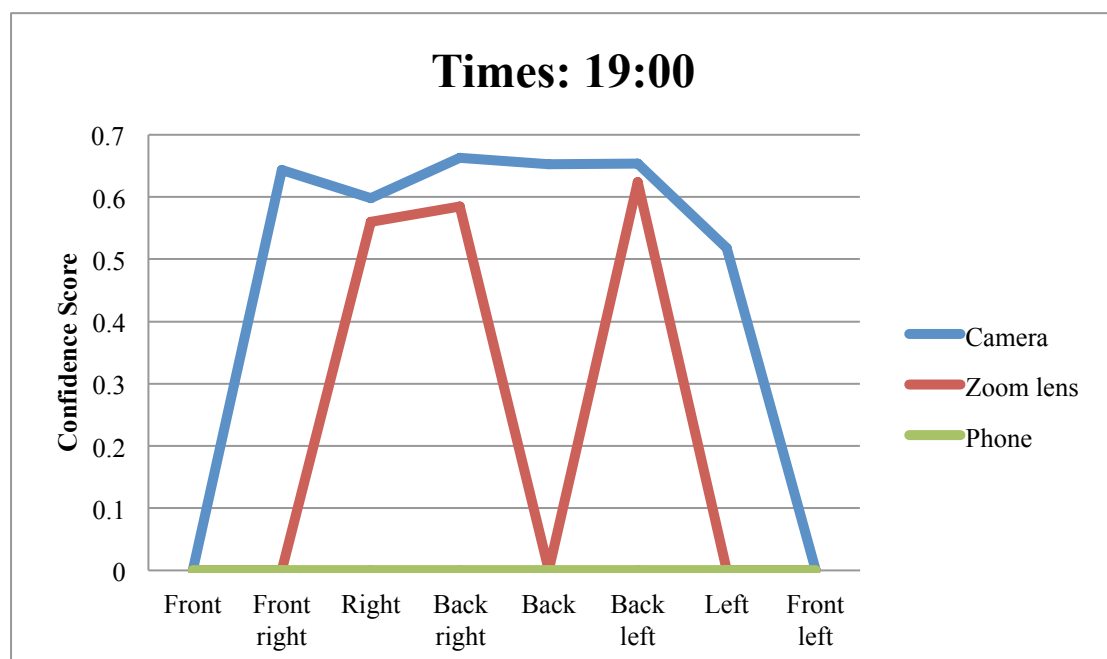
**Figure 68: Showing the Confidence score at 18:00**



### 8.1.14 Time 19:00

**Table 30: Showing the confidence score of three type of lens and device at 19:00.**

Time : 19:00	Camera	Zoom lens	Phone
Front	0	0	0
Front right	0.643787	0	0
Right	0.597742	0.560573	0
Back right	0.66293	0.584682	0
Back	0.652441	0	0
Back left	0.65323	0.623404	0
Left	0.517686	0	0
Front left	0	0	0
Average	0.465977	0.221082375	0
Subaverage	0.621302667	0.589553	0

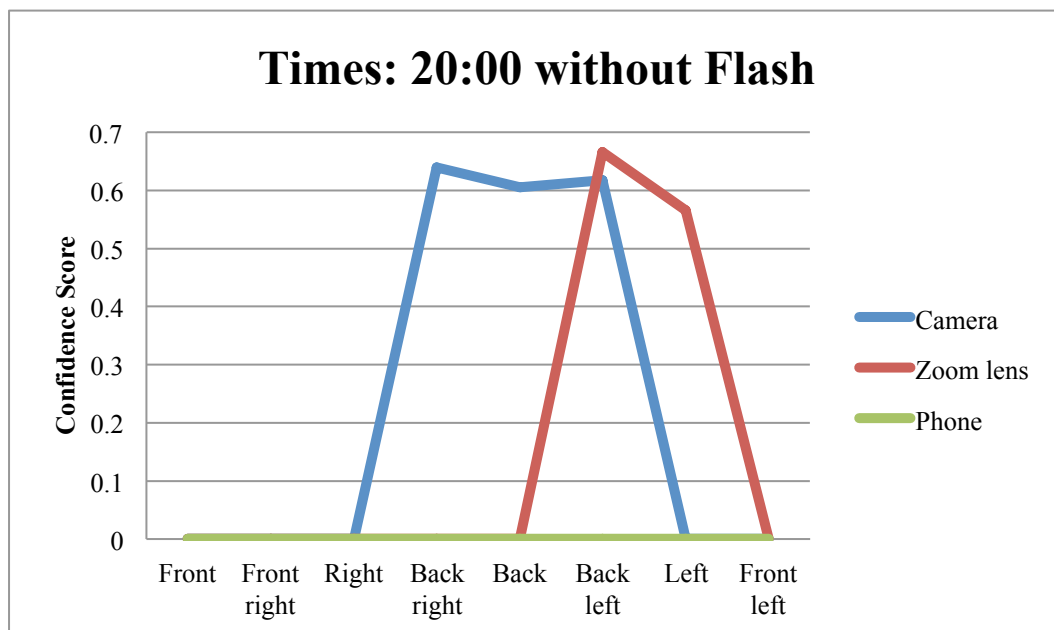


**Figure 69: Showing the Confidence score at 19:00**

### 8.1.15 Time 20:00 without flash

**Table 31: Showing the confidence score of three type of lens and device at 20:00 without flash**

without	flash		
	Camera	Zoom lens	Phone
<b>Time : 20:00</b>			
<b>Front</b>		0	0
<b>Front right</b>		0	0
<b>Right</b>		0	0
<b>Back right</b>	0.63983		0
<b>Back</b>	0.604924		0
<b>Back left</b>	0.61717	0.666339	0
<b>Left</b>		0	0.565381
<b>Front left</b>		0	0
<b>Average</b>	0.2327405	0.153965	0
<b>Subaverage</b>	0.620641333	0.61586	0

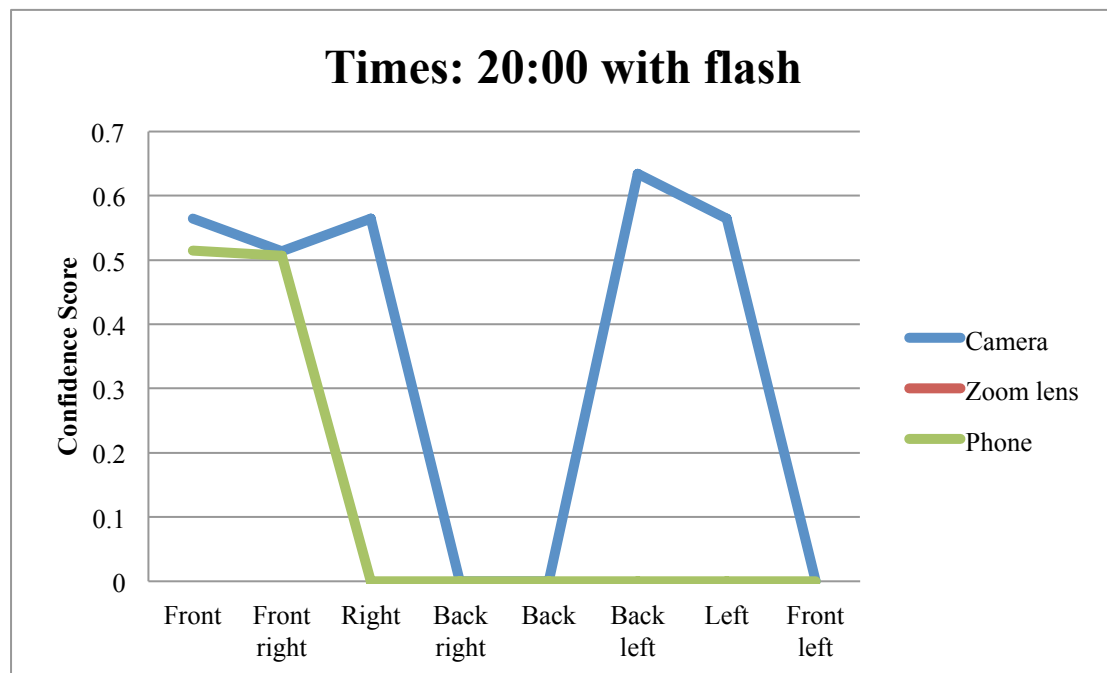


**Figure 70: Showing the Confidence score at 20:00 without flash**

### 8.1.16 Time 20:00 with flash

**Table 32: Showing the confidence score of three type of lens and device at 6:00 with flash.**

with flash			
Time : 20:00	Camera	Zoom lens	Phone
Front	0.564953		0.51425
Front right	0.513637		0.50638
Right	0.564223		0
Back right		0	0
Back		0	0
Back left	0.634148		0
Left	0.564223		0
Front left		0	0
Average	0.355148		0.12757875
Subaverage	0.5682368		0.510315



**Figure 71: Showing the Confidence score at 20:00 with flash**