UNIVERSITY OF LINCOLN

School of Computer Science

Name: Melissa Ho Hui Qi

Student ID: HOM 14581956

Assignment: CMP 3060M - Project Proposal

Project Title: Prediction of Human Emotions Through Robot Facial Recognition

Supervisor: Dr Heriberto Cuayahuitl

Due Date: 26th October 2017

Table of Contents

1. Introduction

Is it possible for future robots to recognise human emotions?

Emotion recognition for Artificial Intelligence(AI) is an important field yet to be discovered and implemented within a more human-like social robot (Kaliouby, 2016). In terms of human-like, it means that robots could not only be capable of thinking like a human but also responding to a situation analytically like humans do. For instance, a social robot should comfort people who feel sad or unhappy, which makes the robot a great companion to humans. For example, *Diagram1* shows an example of a toy robot named Cosmo which is able to display multiple emotions to show its moods while interacting with children. In the future, it is highly possible that this kind of social robot will be developed to accompany and take care of the older people on behalf of their children.

Diagram1: Cosmo Robot with Different Emotions (Macdonald, 2016)

1.1 Background

In this new era surrounded by technologies, the existence of robots has increasingly become more important in our daily life by taking over our daily tasks one-by-one. For instance, robots have been located within each field: from healthcare and homecare, to military use and emergency response (Hunter, 2013). Thus, an Alexithymia social robot could be an issue while interacting with humans. In simpler words, a robot without being able to read and interpret human emotions causing it to be no more than just a piece of working machine where there will be less social interaction between the user and the robot. Kaliouby, R. (2016) stated that it is important for robots to have Emotional Intelligence which enable them to read the user's emotions before reacting to it appropriately. This will result in a more likeable and yet successful robot within society. This could be supported by a movie named Ex-Machina which has foreseen the future in inventing a robot like Ava that is self-aware and able to read the expression a human gives before self-determining its following responses while communicating (Garland, 2015).

1.2 Reason for Proposing this Project

The topic of my proposed project is to program robots that are capable of reading and predicting human emotions through facial analysis while communicating with humans. As a student of BSc Computer Science, the knowledge learned from related modules like Artificial Intelligence, Image Processing, Machine Learning, Human-Computer Interaction, and Autonomous Mobile Robotics could also be implemented throughout this project. Hence, it is truly beneficial for my future as this is currently a popular topic in the field of robotics.

The following proposal starts by explaining the aim and objectives about how this project could be achieved. Later, it is followed by a relevant list of academic literature and how they relate to this project. After that, this proposal will continue to state the timeline plan for this project about how long it will take to achieve each step until the final product. Lastly, this proposal will list out the risk matrix about how likely for each risk to happen during the project.

2. Aim and Objectives of this Project

2.1 AIM

To create a robot[1] system able to predict human emotions through face detection and emotion recognition using data from a 2D camera

At real-time, the robot is able to view and interpret the current emotion of the human using the coded system, as shown in *Diagram2*. Convolutional Neural Network will be trained within the system as it has higher accuracy in predicting the correct result than other methods although it has slower efficiency.

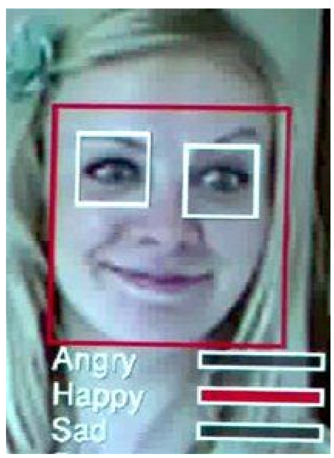

Diagram2: Example of robot analysing human emotions (Fenlon, 2011)

2.2 OBJECTIVES

1. ***The proposed robot is able to capture and select the images from its real-time scenario.***

[1] The Proposed Robot used for this project will be the Pepper: https://www.ald.softbankrobotics.com/en/robots/pepper

The robot is able to capture the image from its surroundings using a process called image acquisition.

2. ***The proposed robot is able to detect human faces among other objects.***

   The captured images are then being processed to detect the existence of human faces. Human faces are then cropped out while the rest of the image are being deleted to minimise the memory usage. At the same time, the cropped image are being filtered to be seemed more clearly.

3. ***The proposed robot is able to recognise the characteristics of human faces.***

   After that, the characteristics of a human face should now be programmed into the system to let the robot know what human face should be consisted of. For instance, the robot should recognise the location of eyebrows, eyes, nose, ears, and lip of a human.

4. ***The proposed robot is able to detect, recognise and differentiate some basic human emotions.***

   It is estimated to have around 3-4 human emotions (i.e. joy, neutral, sad, and(or) angry) within the system. All the difference characteristics of different human emotions are programmed separately within the system. The differences between human emotions against human neutral facial expression are compared within the system. Convolutional Neural Network will be used to train the system to predict the different emotions.

5. ***The proposed robot is able to predict human emotions although only partial of the face is captured.***

   The system recognises the characteristics of the partial face image captured, and analysed to obtain the final result of emotion displayed. For instance, the robot is able to analyse one's emotions even though he/she does not face the webcam of the robot directly.

<u>3. Academic Literature Review</u>

The following academic literature are related to my project which is to predict human emotions through facial recognition.

3.1 DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network

Dehghan, A., Ortiz, E., Shu, G. and Masood, S. (2017) created the Sighthound[2], an API that performs facial recognition to predict human emotion, age and gender. Sighthound implements Convolutional Neural Network(CNN) for a higher accuracy result as in *Table1*. *Diagram3* shows the process on how the final result is predicted from the input images. The system is trained with over 4,000,000,000 images for better prediction.

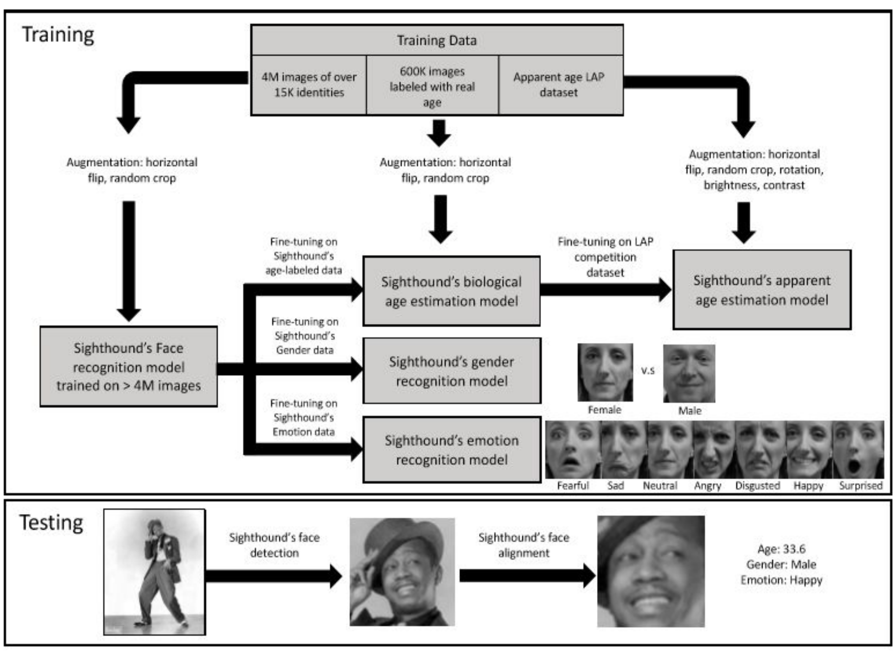

Diagram3: Proposed System Process

[2] Sighthound API available at https://www.sighthound.com/products/cloud

Methods	Accuracy
Sighthound	76.1%
Microsoft [13]	61.3%

Table1: Accuracy Result of Emotion Recognition

Hence, this article has given me a choice to use Sighthound as an alternative for my project.

3.2 Will AI ever understand human emotions?

Minku, L. (2017) states the detection of human emotions using Artificial Intelligence (AI) is used to identify most criminals through their hidden emotions. Besides facial recognition, this article gives me an idea that human emotions could be predicted through our voices.

3.3 Emotion Recognition in the Wild via Convolutional Neural Networks(CNN) and Mapped Binary Patterns.

Levi, G. and Hassner, T. (2015) helps me in understanding the process whereby the face region was first detected from the converted grayscale images. Later, each pixels of the images are coded with Local-Binary-Patterns(LBP) for Multi-Dimensional Scaling(MDS) and compared with its RGB code. The results of the predicted emotion are generated using CNN for higher accuracy rate in *Table2*.

			Validation			Test
	RGB	LBP_1, w.o. cyclic	LBP_1, cyclic	LBP_5, cyclic	LBP_{10}, cyclic	
Baseline (provided by the Challenge authors)			35.33%			39.13%
GoogleNet - single crop	41.68%	39.34%	41.45%	41.68%	40.28%	—
GoogleNet - Oversampling	41.21%	39.57%	40.98%	40.98%	41.45%	—
VGG_S - single crop	41.45%	39.34%	41.92%	41.92%	39.34%	—
VGG_S - Oversampling	40.04%	43.79%	42.38%	43.09%	41.21%	—
VGG_M-2048 - single crop	37.93%	38.17%	36.06%	40.98%	32.31%	—
VGG_M-2048 - Oversampling	40.74%	42.62%	36.76%	42.62%	33.72%	—
VGG_M-4096 - single crop	37.00%	41.92%	40.28%	40.28%	37.47%	—
VGG_M-4096 - Oversampling	37.93%	42.85%	44.73%	42.85%	40.98%	—
Ensemble	46.13	48.94	47.3	47.54	47.3	—
Ensemble of all methods			51.75%			54.56%

Table2: Accuracy of Emotion Detection using 4 Different Network Architecture

3.4 Supervised Committee of Convolutional Neural Networks(CNN) in Automated Facial Expression Analysis

Pons, G. and Masip, D. (2017) introduces 2 different types of CNN modelling: CNN-4L process (as shown in *Diagram4*), and VGG-16 process which are used to analyse the characteristics contained within the images before predicting the emotions.

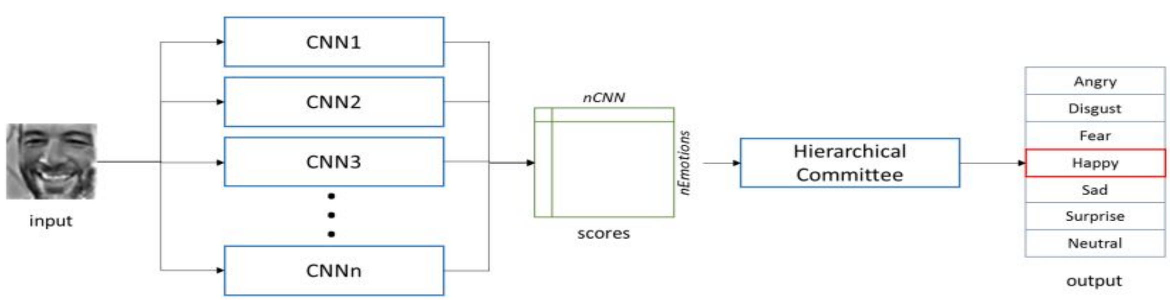

Diagram4: CNN-4L process

Method	CNN-4L networks	VGG-16 networks
Best individual CNN	37.8%	41.3%
Majority vote	35.0%	38.0%
Average	34.2%	40.8%
VAexpoWA	37.5%	41.2%
SVM	22.2%	19.0%
Neural Network	23.0%	19.9%
Proposed CNN	**39.3%**	**42.9%**

Table3: Accuracy Result

Hence, it is suggested to implement VGG-16 networks as the CNN within my project as it has higher accuracy rate (as shown in *Table3*).

3.5 Real-time Emotional State Detection from Facial Expression on Embedded Devices

Unlike the other projects using CNN, Turabzadeh, S., Meng, H., Swash, R., Pleva, M. and Juhar, J. (2017) implements Regression Modelling using K-Nearest Neighbour(K-NN) as shown in *Diagram5*. This modelling is more efficiency in capturing and analysing the images in real-time but has lower accuracy (47.44%) than CNN.

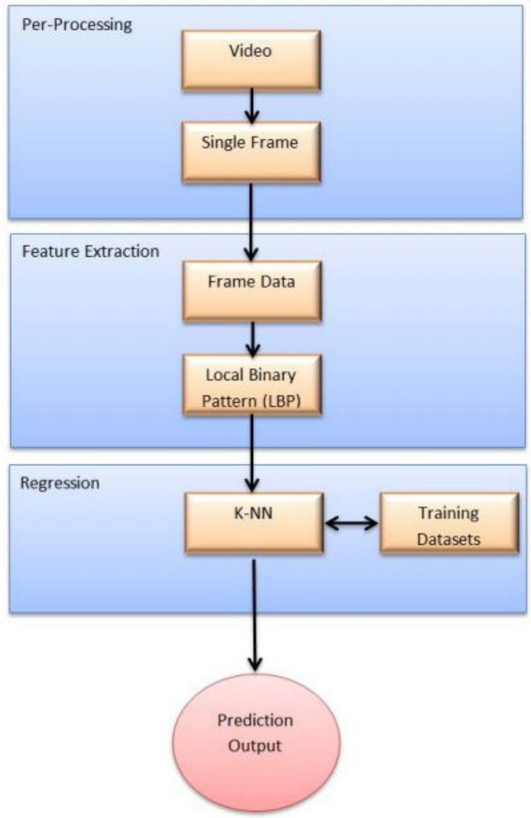

Diagram5: Process of Detecting Human Emotions

3.6 Summary

In sum, Minku, L. (2017) gives me an idea that human emotions could be predicted through body movement and voice recognition besides facial recognition whereas Dehghan, A., Ortiz, E., Shu, G. and Masood, S. (2017) compares the system of Sighthound with other commercial facial recognition API which gives me an idea how to improve my project. Lastly, there are several methods to perform facial recognition, such as through Regression Modelling (Turabzadeh, Meng, Swash, Pleva, and Juhar, 2017) or Convolutional Neural Network ((Levi, and Hassner, 2015), (Pons, and Masip, 2017)).

4. Project Plan

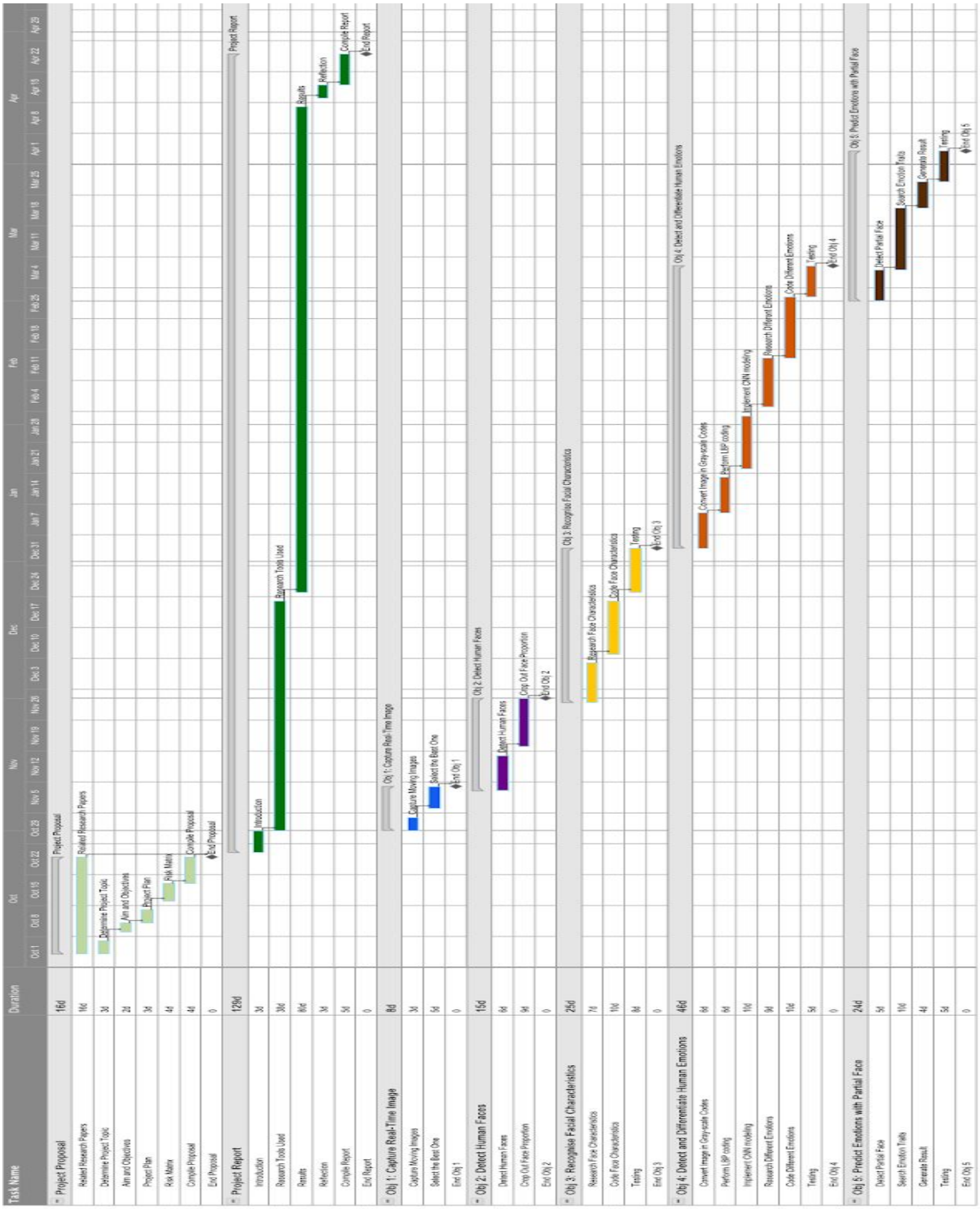

Diagram 5: Gantt Chart of the Overall Project Plan

Diagram 5 shows a Gantt Chart with timescale(days) and milestone for each task involved throughout this project. For this project, it consisted of 2 main written documentaries: Project Proposal, and Project Report.

4.1 Project Proposal

An estimation of 16 days is planned to complete the Proposal which consists of 6 subtasks: ***'Related Research Papers', 'Determine Project Topic', 'Aim and Objectives', 'Project Plan', 'Risk Matrix',*** and ***'Compile Proposal'***.

4.2 Project Report

An estimation of 129 days is planned to complete both Written Document and the Artefact.

4.2.1 Written Document

This involves writing the ***'Introduction'***, listing ***'Research Tools Used'***, recording ***'Results'***, concluding with ***'Reflection',*** and lastly ***'Compile Report'*** to check for any corrections.

4.2.2 Artefact

The Artefact will be produced through the following five objectives along with the help from the trainable robot system created by Cuayahuitl, H. (2017). The accuracy of Emotion Recognition will be measured by comparing the emotions that the subjects mimics the displayed slides and the predicted results by the system.

1. ***‘Capture Real-Time Images’*** requires 8 days to *‘Capture Moving Image’* using the webcam, and *‘Select the Best One’* chooses the best image out of several repetition photos per seconds. Alternatively, several images are selected and used to train data.

2. ***‘Detect Human Faces’*** requires 15 days to *‘Differentiate Human Faces’* which determines human faces against other objects within the selected images. An algorithm is used to categorise all images into 2 types: positive images (i.e. images with human faces), and negative images(i.e. images without human faces) (OpenCV, n.d.). Then *‘Crop Out Face Proportion’* allows the system to focus only on Human Faces.

3. ***‘Recognise Facial Characteristics’*** requires 25 days to *‘Research Face Characteristics’* where several facial characteristics (i.e. eyes, nose, and lips) will be selected for *‘Code Face Characteristics’* that will be implemented using Haar Classifiers to obtain more accurate results (Wilson, and Fernandez, 2006). Later, the result of the identified areas within the images will be observed and recorded under *‘Testing’*.

4. ***'Detect and Differentiate Human Emotions'*** requires 46 days to *'Convert Image into Grayscale Codes'* to reduce the risks caused by the colours before *'Perform LBP coding'* for Multidimensional Scaling (Levi and Hassner, 2015). Next, *'Implement CNN'* codes the proposed DQN modelling(Cuayahuitl,2017) into the system followed by *'Research Different Emotions'* which finds out the difference facial expressions for each human emotion (i.e. joy, neutral, sad, and(or) angry). Later, the research will be implemented under *'Code Different Emotions'*, and *'Testing'* is where the accuracy of emotions successfully detected will be recorded using the feedback from the subjects.

5. ***'Predict Emotions with Partial Face'*** is an advanced combination of all the previous objectives, and requires 24 days to *'Detect Partial Face'* is where the system to recognise part of the human face characteristics (i.e. eyes, lips, and etc.) without detecting the whole region of it. From the detected face, *'Search Emotion Traits'* is compared to the programmed emotions concepts and looks for similarities before *'Generating Result'*. Lastly, the results are recorded under 'Testing' to measure its accuracy through feedbacks.

5. Risk Matrix

Risk ID	Risk	Likelihood of Occurrence	Impact on the project	Mitigation
1	Low Battery	**Possible**	**High Impact.** The result is unable to be generated.	Keep track of the battery usage level. Charge it when necessary.
2	Electricity Shock	**Low**	**Low Impact.** It might cause physical harms to the users who are in contact with it.	Keep away from damp areas.
3	Fallen of the Robot	**Low**	**Low Impact.** It might cause physical injuries to the people nearby.	Be careful while moving the robot.
4	Webcam of the Robot Not Functioning Properly	**Possible**	**High Impact.** No images could be input into the system to proceed.	Test and check the webcam if it is in good condition before testing the system. Use an alternative webcam to capture the images.
5	The lighting of the environment for image capturing (i.e. too dark or too bright)	**Very Likely**	**High Impact.** Neither too dark nor bright causes the system unable to detect human faces within the image captured.	Perform the Testing indoors with light bulbs to reduce the impact caused of the sunlight.

6	The Face Movement is too fast to be captured and analysed at the same time	**Likely**	**High Impact.** The system might miss out the important key points that determine the person's emotions.	The system should try to capture as many frames per second as possible.
7	Error occurred within Convolutional Neural Network	**Rare**	**High Impact.** This may lead to system failed.	An alternative Neural Network modelling could be considered.
8	Robot being stolen or unavailable	**Unlikely**	**Low Impact**. This unable to prove that the system is compatible or functionable along with the robot.	The result can be obtained by using a laptop to carry out the process instead.

6. References

Cuayahuitl, H. (2017) Deep reinforcement learning for conversational robots playing games. IEEE. Available from http://eprints.lincoln.ac.uk/29060/1/hc-humanoids2017.pdf [accessed 21 October 2017].

Dehghan, A., Ortiz, E., Shu, G. and Masood, S. (2017) DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network. *arXiv preprint arXiv:1702.04280.* Available from https://arxiv.org/pdf/1702.04280.pdf [accessed 22 October 2017].

Fenlon, W. (2011) We Shouldn't Be Afraid of Facial Recognition and Other Amazing Technology - Tested.com. Available from http://www.tested.com/tech/2672-we-shouldnt-be-afraid-of-facial-recognition-and-other-amazing-technology/ [accessed 17 October 2017].

Garland, A. (dir.) (2015) *Ex-Machina* [DVD]. Universal Pictures, Film4 Productions, DNA Films.

Hunter, R. (2013) How Robots Are Revolutionizing Our World [Sponsored Post]. Available from https://www.popsci.com/trp-sponsored-article-analyst [accessed 9 October 2017].

Kaliouby, R. (2016) 5 Reasons Why Artificial Intelligence Needs Emotion. Available from http://blog.affectiva.com/5-reasons-why-artificial-intelligence-needs-emotion [accessed 9 October 2017].

Levi, G. and Hassner, T. (2015) Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns. In: *Proceeding ICMI '15 Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, Seattle, Washington, USA, 9-13 November. New York, NY, USA: ACM, 503-510. Available from https://dl.acm.org/citation.cfm?id=2830587 [accessed 14 October 2017].

Macdonald, C. (2016) Cozmo the robot has 'emotion engine' to bond with humans. Available from http://www.dailymail.co.uk/sciencetech/article-3662908/The-tiny-robot-just-wants-friend-Cozmo-develops-bond-owner-eyes-light-sees-them.html [accessed 9 October 2017].

Minku, L. (2017) Will AI ever understand human emotions?. Available from http://www.independent.co.uk/news/will-ai-ever-understand-human-emotions-a7521076.html#gallery [accessed 11 October 2017].

OpenCV (n.d.). Face Detection using Haar Cascades Available from https://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html [accessed 22 October 2017].

Pons, G. and Masip, D. (2017) Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis. *IEEE Transactions on Affective Computing*, PP(99) 1-1. Available from http://ieeexplore.ieee.org/abstract/document/8039231/ [accessed 16 October 2017].

Turabzadeh, S., Meng, H., Swash, R., Pleva, M. and Juhar, J. (2017) Real-time Emotional State Detection from Facial Expression on Embedded Devices. In: *Seventh International Conference on Innovative Computing Technology (INTECH 2017)*, Luton, UK, 16-18 August. Slovakia: IEEE. Available from http://v-scheiner.brunel.ac.uk/handle/2438/14961?mode=simple [accessed 17 October 2017].

Wilson, P. and Fernandez, J. (2006) FACIAL FEATURE DETECTION USING HAAR CLASSIFIERS. *Journal of Computing Sciences in Colleges*, 21(4) 127-133.