Modeling and Design Analysis of Facial Expressions of Humanoid Social Robots Using Deep Learning Techniques

by

Shweta Murthy

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2017 by the
Graduate Supervisory Committee:

Ashraf Gaffar, Chair
Javier Gonzalez Sanchez
Arbi Ghazarian

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

A lot of research can be seen in the field of social robotics that majorly concentrate on various aspects of social robots including design of mechanical parts and their movement, cognitive speech and face recognition capabilities. Several robots have been developed with the intention of being social, like humans, without much emphasis on how human-like they actually look, in terms of expressions and behavior. Furthermore, a substantial disparity can be seen in the success of results of any research involving "humanizing" the robots' behavior, or making it behave more human-like as opposed to research into biped movement, movement of individual body parts like arms, fingers, eyeballs, or human-like appearance itself. The research in this paper involves understanding why the research on facial expressions of social humanoid robots fails where it is not accepted completely in the current society owing to the uncanny valley theory. This paper identifies the problem with the current facial expression research as information retrieval problem. This paper identifies the current research method in the design of facial expressions of social robots, followed by using deep learning as similarity evaluation technique to measure the humanness of the facial expressions developed from the current technique and further suggests a novel solution to the facial expression design of humanoids using deep learning.

DEDICATION

To my husband and my parents

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr.Ashraf Gaffar, for guiding me through my research and giving me an opportunity to work on something so interesting. I thank my committee for their kind reviews and suggestions.

I would also like to thank my husband, who has not only been my best friend, but also the greatest mentor and critic I have ever had. It is because of him that I have been able to do the things that I have done.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

Robotics in the current world resonates with automation. Robotics has been making an appearance everywhere, starting from simple robotic arms in factories, to intricate medical instruments, to complex social robots. Automation not only helps with efficiency but also serves as an alternative or sometimes the only option in situations where other help is difficult. This would include scenarios of nuclear disasters where self-maneuvering robots can be capable of controlling the situation from worsening or in environments of care and companionship like elderly homes. "Humanoid" robots haven't been the most successful part of robotics research. Humanoids are robots that have human like appearance and can perform cognitive actions and display of emotions like humans. Since human behavior, expressions and actions are so diverse and complicated, the lack of successful social robots that both look like humans and display behavioral expressions and characteristics is obvious.

Even if there has been considerable research into facial design of humanoids most of the research concentrates on the appearance of the face being as human-like as possible but not so much on how human-like the expressions displayed by the face look. But even after these considerations about the design features and characteristics that are most identified as human Sinha *et al.* (2006), their static expressions or behavior with changing expressions add to uncanny valley theory, making these expressions appear more scary and less human-like.

The research on facial features and behavior of humans has been going on since decades. The research can be thought of as being pursued in three different timelines or generations. First being research on human faces and their expressions and understanding points of importance in the human expression that differentiate every expression from the other, while also identifying different faces. This was followed by a second generation of research, where the available results and analysis of its outcomes were applied in movies and animation along with research into artificial intelligence software such as face recognition and prediction. The last generation of research, which can be thought of as the present era of research, is the research on social robots, specifically humanoids.

Initially, during the “first generation” of research on human faces, emphasis was on understanding facial features that uniquely make up the human expressions, along with understanding human cognitive behavior which helps in the recognition of the expressions by humans themselves. The results from these experiments were extensively used in animation Animation (1998) as well as primarily in the introduction and improvement of artificial facial recognition software and engines DiSalvo *et al.* (2002), Hizem *et al.* (2006). One crucial requirement in animation is modeling the facial expressions accurately. The animation software require and study data obtained from human facial expressions Hirose and Ogawa (2007). This era of initial research can be thought of as being complete, since we can see relevant results from such experiments that are being extensively used in movies and predictive and classification photo applications.This kind of research has also been used in medical research for cosmetic and healing studies. This can be thought of as another era of research that was successfully conducted and concluded. This resulted in a large amount of data to pertaining to human behavior and expression to be collected and important features and labels analyzed leading to far more progression into artificial intelligence in automatic emotion and face recognition Edsinger and O'Reilly (2000)

The research on social robots and humanoids makes the current era of research. Since technology and artificial intelligence have proved to be immensely successful with social beings, humans, humanoids or robots that look like humans and combine the appearance with human like behavior are the future of technology.

There are numerous successful research on social robots Hirose and Ogawa (2007), Ogura *et al.* (2006), Jokinen and Wilcock (2014), which primarily achieved results on biped movement analogous to the way human beings move. Face-to-face communication still forms the most important part of human communication or human-humanoid interaction. Human faces can display and convey messages through a wide array of not only dynamic facial expressions but static expressions as well. There has been considerable research on making the humanoid head more human like. Robot SAYA Hashimoto *et al.* (2006) that was developed for rich facial expressions is a very good example of facial design research. The research in Hashimoto *et al.* (2006) also deals with identifying and using action units that pertaining to human facial muscles that contribute to facial expressions.

Chapter 2

GOAL

The main aim of this thesis is to establish a quantitative framework for the objective evaluation of mechanical social robots. This thesis provides a study on the current generation of mechanical social robots that have been researched into and developed successfully.

While the algorithms chosen in this thesis provide a quantitative measure of humanness of the mechanical social robots, there are a large number of factors that can affect the results that are obtained in the future using the same algorithms and the approach. This can be attributed to not only advancement in the design process of the mechanical social robots but also advancements in the learning algorithms themselves. The effect can be positive or negative based on how advancements progress.

The approach however will remain the same where each of these robots can be measured for their humanness and the changes in their design aspects. This would not only help assess the mechanical social robots being considered at that moment but will continue in to help in modifying and improving the design process based on the results obtained from the method outlined in this thesis.

Chapter 3

LITERATURE SURVEY

3.1 Generation of Robots

In this review of the related literature, I propose that the advances in robotics can be generally divided into five successive generations:

(1) Accurate and Dexterous Robots,

(2) Mobile Robots,

(3) Autonomous Robots,

(4) Socially Capable Robots, and

(5) Socially Autonomous Robots.

While there are relatively clear distinctions between the five generations, the following observations hold:

1) There is some research that spans two generations together. For example, a single research effort can focus on the accuracy of manipulator motion (generation 1) as well as on some autonomous decision making to guide that motion (generation 2).

2) Each generation could use functionality of previous ones. For example, development of autonomous actions (generation 2) requires as a prerequisite the ability of accurate motion (generation 3). Thus, generations are apparently cumulative in nature.

3.1.1 First Generation: Accurate and Dexterous Robots

The field of robotics is often considered to have grown out of two earlier machines: the CNC manufacturing machine Newman *et al.* (2008) in 1949 and the Teleoperator Johnsen and Corliss (1971), or remote manipulator in 1945. Development of the Teleoperator provided the context for the derivation of manipulator kinematics for positioning and velocity and force control of an end effector via joint control. The CNC machine contributed to the development of automation in motion control for fast, accurate, and stable positioning of an end effector. The combination of the knowledge gained through the remote manipulator and the CNC machine provided the technical foundation for the first robots, capable of simple preprogrammed motions of a manipulator and end effector through complex paths. These accurate and dexterous robots were limited to specified motions, and not yet capable of what would be classified as movement autonomy.

3.1.2 Second and Third Generation: Mobile and Autonomous Robots

While first generation robots were progressing in their mostly arm movements and dexterity, they were not capable of motion other than either specific preprogrammed trajectories or directly human guided motions. Second generation robots focused on mobility while the third generation robots focused on autonomous mobility, with the development of path planning and collision avoidance algorithms. These algorithms allow the robot to identify and follow paths and execute motions other than human directed or preprogrammed motions according to algorithmic commands, rather than point to point motions. These mobile and autonomous robots began to have the capability to move around in increasing distances first on wheels, then on multiple spider like legs, and eventually on two legs to simulate human walking. Walking stability and incident recovery improved with further research. These stabile robots required an awareness of the surrounding environment in order to identify and avoid obstacles and problematic areas, a significant problem not present in first generation robots. Examples of these mobile and autonomous robots include the Mars rover, self driving cars, and walking robots.

3.1.3 Fourth Generation: Socially Capable Robots

This generation of robots was designed with consideration of primitive social behavior of the robots in interactions with human operators. Robot social behavior was addressed in two main categories:
1) Displaying humanlike output (talking, face impressions, gestures, gazing, paying attention, emotions)
2) Perceiving and interpreting human social behavior as an input (voice, movements, presence, attention, etc.)
Research in this generation mostly focused on implementing one or few social components (either through an actual robot implementation Embgen *et al.* (2012), or by a simulated lab experiment), and identifying the effect of a particular social behavior. Most recent and current research in human robot interaction can be classified within this generation. Some of these key studies are categorized and summarized here.

Comparison of Human Robot Communication With and Without Social Cues

Vossen *et al.* (2010) argues that delivering the same information to humans via a socially capable robot (using speech or physical appearance) works more effectively than just presenting the information as plain factual feedback. The work by Jayagopi *et al.* (2013) is an early attempt to identify and build a dataset of social cues including para-verbal and nonverbal cues and their meanings.

Investigation of Algorithms for Robot Perception of Human Behavior

Sheikhi and Odobez (2012) test the capability of a robot to identify the combination of a human's head pose direction and eye gaze direction to estimate what the human subject is looking at. It is known that humans look at objects by the combined movement of our heads as well as the movement of our eyes. Only the combined effect of both movements determines the object we are looking at.

Investigation of Algorithms for Robot Imitation of Human Behavior

Mathur and Reichling (2009) evaluate the level of human trust of robots based on robot appearance. The “”Uncanny Valley of Trust” is evaluated and confirmed, in which it was found that up to a certain degree, humans prefer highly realistic robot faces over mechanical faces. Castellano and McOwan (2009) provide a detailed analysis of how socially capable robots can affect human behavior. Finke *et al.* (2005) show how a robot can perceive human presence and interest, and how to attract their attention in a gentle, nonintrusive way, Nakagawa *et al.* (2013) measure how humans believe in a whispering robot as they do with a whispering human.

3.1.4 Fifth Generation: Socially Autonomous Robots

While 4th generation robots display important social action or response, they are mostly limited to simple and limited social actions. In contrast, 5th generation robots include three main improvements:

1) An integrated set of complex social behavior (for example a broad range of natural social action and reaction)

2) Robot stability and reliability over extended period of time.

3) A continuous, intelligent dialog between human and robot rather than one action and/or one retraction.

The 5th generation robots are capable of a continuous and prolonged interaction with humans without deteriorating performance or breakdown.

The robot built for use in healthcare Kuo *et al.* (2012) can be considered an early 5th generation robot with both advanced social capabilities as well as extended autonomy. The researchers integrate several complex social variables including environment awareness, user presence awareness, user's attentions, and user face impression detection. They focus on the system stability and uptime of the robot for up to 6 hours a day for 2 weeks period with minimal technical maintenance. The robots operated in two different environments: A private apartment and a public lobby.

Breazeal (2000) built an anthropomorphic robot with extended and continuous conversational capabilities with humans using turn taking dialog. Gaschler *et al.* (2012) integrate many nonverbal cues (108 interaction movements) taken from natural human body language. The robot successfully worked as a bartender for extended periods of time with high success rates (91.2% recognition rate.)

3.1.5 Current Research

Overall, most current HRI research is focused in the fourth generation. Only a few studies extend into the fifth generation. An excellent example of this is the research on establishing a platform for identifying and understanding nonverbal information of users of robotic technologies Burleson *et al.* (2004). I propose a study within the fifth generation of HRI research by investigating the ability of an artificial agent to generate trust in human artificial agent collaborative tasks in three dimensions:

1) The ability to display multiple social cues

2) The extended period of interaction

3) The adaptability of the robot social behavior depending on the progress of the interaction

The work proposed here relies upon the concept that human trust in a machine does not rely only upon the reliability of the machine for its task, but also upon human perception of the machine. A few previous studies have supported this concept. In a study by Merritt (2011), participants interacted with a fictitious automated system after watching video clips to induce positive or negative moods. This study found that participant perception of the automated machine varied depending upon the video clip that was watched prior to interaction with the machine. A similar study by Phillips and Madhavan (2013) showed that mood manipulations affected participant trust, confidence, and sensitivity to an automated decision making aid. Merritt and Ilgen (2008) found that individual perceptions of a machine account for 52% of variance in trust of the machine, when machine characteristics are held constant.

These previous works conclude that success of interaction is not simply a function of the effectiveness and efficiency of the communication, but also a function of human trust in an automated machine. The trust, in turn, is not only a function of the machine reliability, but also upon human perception of the reliability of the machine. In this study, we propose to extend previous work in two key aspects: to investigate whether the robot social behavior has an effect on performance of a human-machine team on a collaborative task.

3.2 Face Recognition and Classification Algorithms

For accurate classification of facial images, it is important to first extract a face from a given image. For this purpose, algorithms to detect faces in an image are useful. In this work, face detection is performed using an implementation Haar Feature-based Cascade Classifiers in OpenCV. Once a face is detected, classification is done by encoding using three algorithms:

1. Eigenfaces

2. Fischerfaces

3. Local binary pattern histograms.

3.2.1 Face Detection

The feature based classifier that is used in this work was introduced by Viola and Jones (2001) in 2001. This is an efficient machine learning based approach and full details are available in Viola and Jones (2001). Use of Haar wavelet basis as a feature set for image object recognition was introduced by Papageorgiou *et al.* (1998) and its adaptation as Haar Features was part of the work by Viola and Jones (2001). With a training data set of positive and negative images, Haar feature set is an efficient way to create a classifier with a limited number of images in training data set.

In the Viola-Jones based approach, the Haar features used for object detection are shown figure 3.1. Haar features are computed by the difference of the sum of pixel intensities in the bright part of the rectangle from the sum of pixel intensities in the dark part of the rectangles. For a base resolution of 24x24, the number of Haar features is 180000. The key contribution by Viola and Jones (2001) is the use of integral images for fast computation of Haar features recursively. This reduces the computation time of Haar features. An integral image at (x, y) is the sum of all the pixels above and to the left of (x, y), inclusive as defined by

$$I_{\sum}(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x', y') \tag{3.1}$$

where $I_{\sum}(x, y)$ is the integral image of $i(x, y)$. The computation of any rectangular sum is then a look up of four array references

$$\text{sum} = I(C) + I(A) - I(B) - I(D), \tag{3.2}$$

where A, B, C and D are four points in the integral image.The difference between two rectangular sums is eight references.

Using Adaptive Boosting or AdaBoost (Freund and Schapire (1997)), a machine learning algorithm, the feature set can be reduced to few thousands. In the original work by Viola and Jones, they reported 95% accuracy with 200 features. An example of first and second features selected by AdaBoost is shown in 3.2, which can be interpreted intuitively based on the intensity changes around the eye region in a face.

Further optimization was achieved by hierarchically applying the classifiers to eliminate the regions that are least likely to contain a face first. This is the concept of Cascade of Classifiers introduced by Viola and Jones (2001). In their original work, the cascade of filters was split to 38 stages 1, 10, 25, 25 and 50 features in first five stages.

The face recognition detector is then scanned across the image at multiple scales and positions. The scaling is done by scaling the detector and not the image. This can

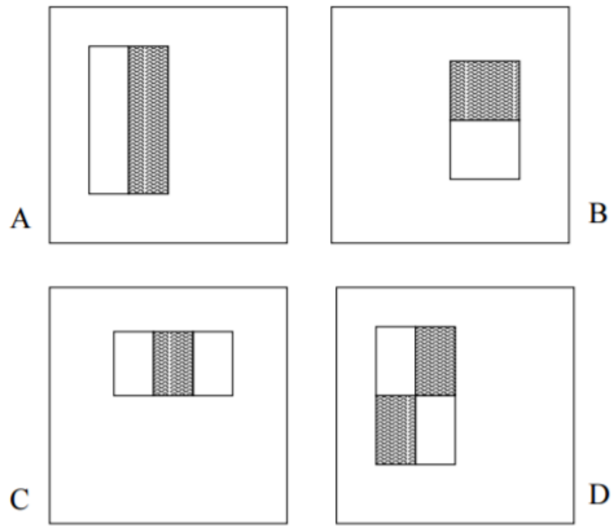


Figure 3.1: Examples of Haar Features Viola and Jones (2001)

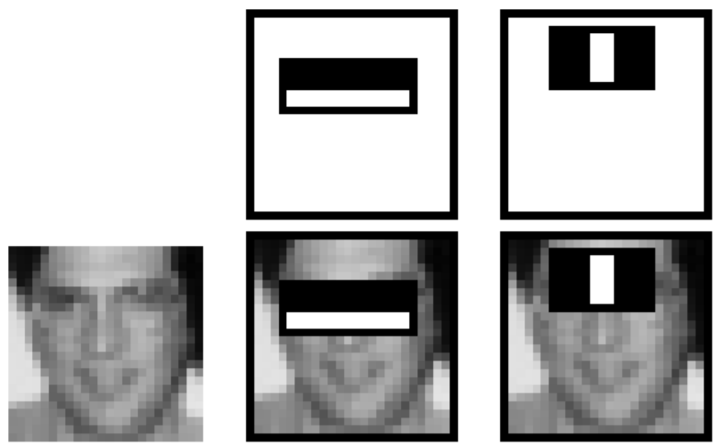


Figure 3.2: First and Second Feature Selected by AdaBoost. The Features Selected Correspond to Easily Interpreted Intensity Changes in a Face Viola and Jones (2001).

3.2.2 Face Classification

Eigen Faces

The basis of this algorithm is dimensionality reduction of images using Principal Component Analysis (PCA) Pearson (1901). A given image of 100x100 pixels has a dimensionality of 10^4. Through the application of PCA, an image can be decomposed to a linear combination of Eigen images such that the variance is maximized for these images. The basic idea here is a representation of an image as a linear composition of images which contain the most amount of information. If T is matrix of training images with each column containing the mean subtracted image, then the eigen faces are obtained by the solving the eigen value equation

$$\mathbf{S}\mathbf{v}_i = \mathbf{T}\mathbf{T}^T\mathbf{v}_i = \lambda_i\mathbf{v}_i \tag{3.3}$$

where $\mathbf{S}$ is the covariance matrix. For a typical image reconstruction, 300 eigen faces are sufficient for a reasonable reconstruction. For face recognition, the euclidean distance of the weights of the eigen faces for a given image is used as the basis for classification.

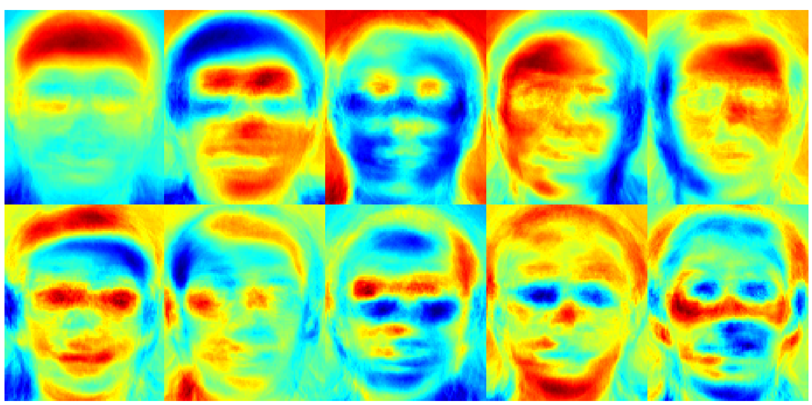


Figure 3.3: Set of Eigen faces Generated from the AT&T Data Set Using OpenCV Implementation of Eigen faces. Retrieved from Wagner (2017)

Fisher Faces

Although linear combination of Eigen faces is useful to encode data, PCA does not use any information about the classes in the training data set to compute the Eigen faces. This discriminatory information is lost when using PCA as splits the images into features that maximize the variance in different images. In many cases, the variance can be due to sources that are external and not related to the classification problem itself. An obvious example of this would be change in ambient lighting.

Fisher faces overcomes this limitation in encoding images as it uses a class specific decomposition. Fisher decomposition is based on Linear Discriminant Analysis, where the images are split into features that maximize the inter-class variation while minimizing the intra-class variation Belhumeur *et al.* (1997). Full details of algorithm is given in Belhumeur *et al.* (1997). Within class variation is estimated by

$$\mathbf{S}_w = \sum_{j=1}^{C} \sum_{i=1}^{n_j} (\mathbf{x}_{ij} - \mu_j)(\mathbf{x}_{ij} - \mu_j)^T, \tag{3.4}$$

where $\mathbf{x}_{ij}$ is the ith sample of class j, μ_j is the mean n_j the number of samples in class j. The inter-class difference is computed by

$$\mathbf{S}_b = \sum_{j=1}^{C} (\mu_j - \mu)(\mu_j - \mu)^T, \tag{3.5}$$

where μ is the mean of all classes. The problem of finding Fisher faces reduces to finding basis vectors $\mathbf{V}$ for which $\mathbf{S}_w$ is minimized and $\mathbf{S}_b$ is maximized. This is expressed as an Eigen value problem

$$\mathbf{S}_b\mathbf{V} = \mathbf{S}_w\mathbf{V}\boldsymbol{\Lambda}, \tag{3.6}$$

where $\boldsymbol{\Lambda}$ is a matrix of eigen values. An example of Fisher face decomposition is shown in Fig.3.4

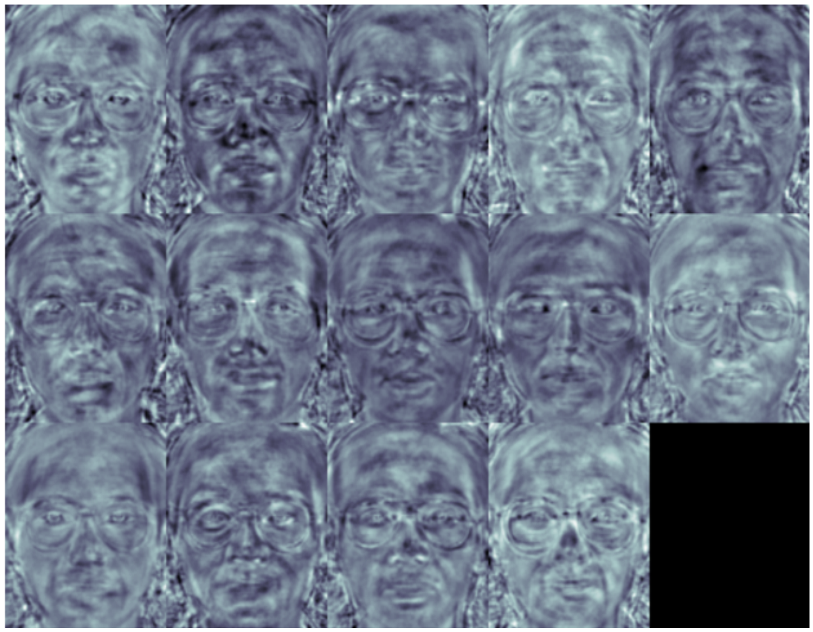


Figure 3.4: The First 16 Fisher faces Shown for an Image from Yale Data Base. Set of Eigen Faces Generated from the AT&T Data Set Using OpenCV Implementation of Eigen Faces. Retrieved from Wagner (2017)

Local Binary Patterns Histograms

Both Fisher faces and Eigen faces look at the image as a whole and any extraneous conditions that impact the image intensities can effect the classification. Even though Fisher faces is more resistant than Eigen faces, drastic changes in light conditions can impact classification. Especially in cases, where just one or two images are available under ideal conditions, the covariance estimates are not very accurate and lead to issues in classification. In such cases, an alternate approach of using local features that are invariant to gray scale transformations. Local Binary Patterns (LBP)is an example of such an encoding. LBP is formed by comparing a pixel intensity to its neighborhood Ahonen *et al.* (2004) as given by the LBP operator

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c), \tag{3.7}$$

where (x_c, y_c) is the pixel position with intensity i_c and i_n is the neighboring pixel intensity. s is the sign function defined as

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases} \tag{3.8}$$

For a 3x3 window, LBP calculation is shown in Fig. 3.5. An example taken from Wagner (2017) shows the invariance of LBP for images under different gray scale image transformations. Ahonen *et al.* (2004) propose splitting the image into m regions and concatenating histograms from each of these regions. Classification is done on the basis of comparing LBP histograms.

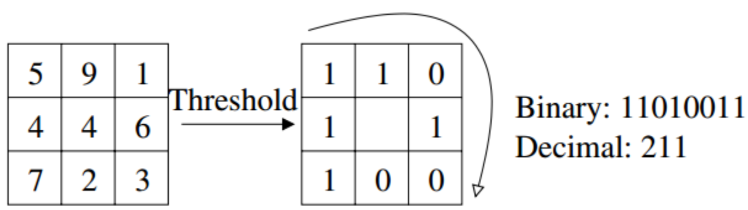


Figure 3.5: LBP Operator With a 3x3 Window Ahonen *et al.* (2004)

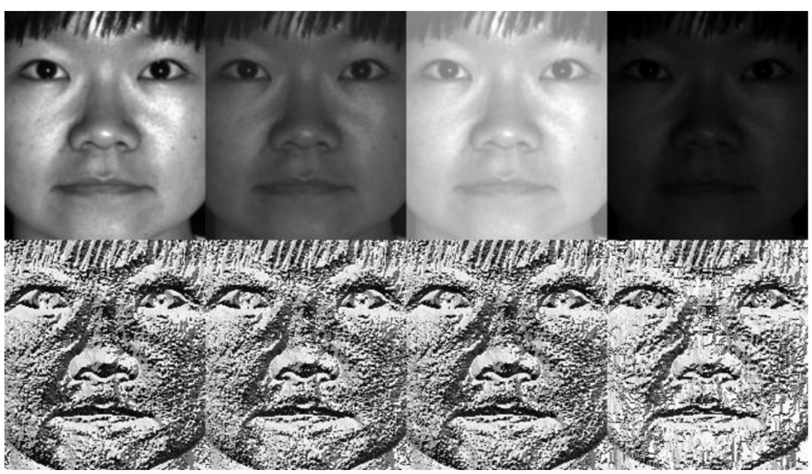


Figure 3.6: LBP for the Same Image Under Different Gray Scale Transformations. Retrieved from Wagner (2017)

Chapter 4

MOTIVATION

In the research that has been conducted in the field of digitization of the mechanical expressions of human faces in the form of animations and movies, most of them involve attaching actuators to human faces to understand key action units or muscles that contribute to the expressions. The same approach has been seen in the field of research where human expressions are converted into mechanical expressions on the humanoid heads, specifically. The technique of understanding and identifying certain action points over other can be also be seen in Artificial Intelligence algorithms, specifically face recognition algorithms, which can be used to replace some of the survey activity required to analyze and understand humanness of a robot's face and its expressions.

There have been certain changes in the way facial expressions are being recognized and categorized. As specified in Taheri *et al.* (2015), the paper tries to categorize the action units or the facial muscles that move when certain expressions are being displayed by the human face. But instead of 19 action units as mentioned earlier, 32 individual facial muscle action units are recognized. The paper also mentions automation of recognition of these action units from static expressions or profile of facial expressions. This is the trend that is trying to be adopted in this paper.

The reason why automation is chosen over manual surveys is motivated by the amount of human facial expression images that are available currently MIT (2003) Databases (2003). Most of these databases are still constructed from photographs taken in very controlled environments which add to the factor of how an image or an expressions or the subject itself is recognized. But this is not always the case in reality where controlled action is highly difficult to achieve. A result of controlled environment images is that the expressions between expressions, termed as dynamic expressions which do not result in movement of a majority of action units but would still result in some information being conveyed to the observer.

The idea here is to understand that the expressions in the human faces/ action units in the human faces represent some of the highest frequency points. These points are the ones that are common with most human faces that includes not only different expressions from a particular subject but over an array of subjects that are being considered to pick the muscles/ action units. This leads to selectively choosing points that are occurring with highest frequency and regarding the rest with negligible frequency as noise.

The research in this paper is motivated by three primary questions. The question about whether the features that are difficult to quantify exist in facial expressions, but are not selected due physical limitations, could actually add important information to the design process of artificial facial expressions. This follows by questioning why static expressions, the expressions that don't convey any emotion but still exist naturally between changing expressions in human faces, are not researched enough but which form a major part of humanness since we see in human expressions that human expressions are never robot like or artificially jumping from one expression to another with no in-between state. Even when there is no expression, there is a certain expression that is being exhibited. This can form a major part of making the humanoid's behavior more human like. Three, this again adds to the possibility of important information being lost when it is being ignored due to lower weightage or values given in the design consideration or "harmonics". This can be termed as "Design Harmonics". The effect of understanding "Design Harmonics" on facial expression design forms the last question.

So as to precisely state, the motivation for this thesis comes in the form of two major questions which combines the concerns mentioned above

1) Where do the human cognitive system and artificial cognitive system differ? Understanding what makes the complex human cognitive system where it is capable of realizing emotions as characterized by another human being by taking into account all the features that come in different frequencies, not matter how small the frequencies might be. It can be seen in a lot of research mentioned in chapter 2, about how face recognition or artificial cognition is put into effect, that a lot of the recognition takes into certain features while ignoring the other features spread throughout the spectrum due to existing limitations. The question whether this information, no matter how insignificant, might add to some important information in making the artificial cognition more human like is a major priority. This, as previously mentioned, is termed in this paper as “Design Harmonics”.

2) Since there is so much data available when it comes to human face images Martinez and Benavente (2007) Solina *et al.* (2003) Huang *et al.* (2007)that has been collected since more than a few decades now, understanding that automation, which uses these databases to their full extent eliminates the need to conduct manual surveys or aides in collecting any new face information, pushes to explore how it can help in the tedious process of making the humanoids behave and display emotions that are more human-like.

Chapter 5

RESEARCH CONTEXT

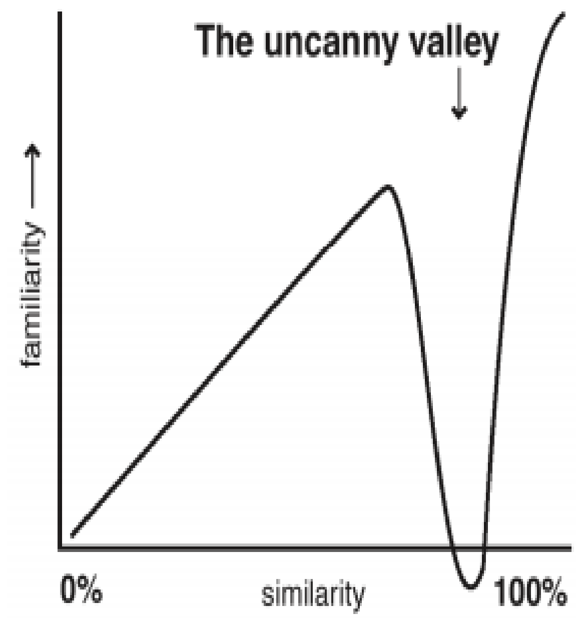
5.1 Need for Research

The research in this paper is intended to be used as part of a novel solution to the design of facial expressions of humanoid robots. A new model can be established based on this research that can form the basis of humanoid expressions and behavioral design that constitute a different approach to design. The work reflected in this thesis paper is important in today's world where the world is moving towards automation and humanoids, robots that look and function like humans, would serve as the best companions and assistive robots to the elderly and children, while also helping with situations that would otherwise prove very difficult, while the major application still remains education and healthcare. Health conditions such as schizophrenia and autism with activities between humanoid robots and the subjects under treatment as a remedial therapy is the latest era of research on the use of humanoid robots in the healthcare domain. Although not enough results have been produced so far Taheri *et al.* (2015), the area of humanoids in mental care is moving forward in the direction of being explored more and more.

5.1.1 Humanoid Robots and Pediatric Healthcare

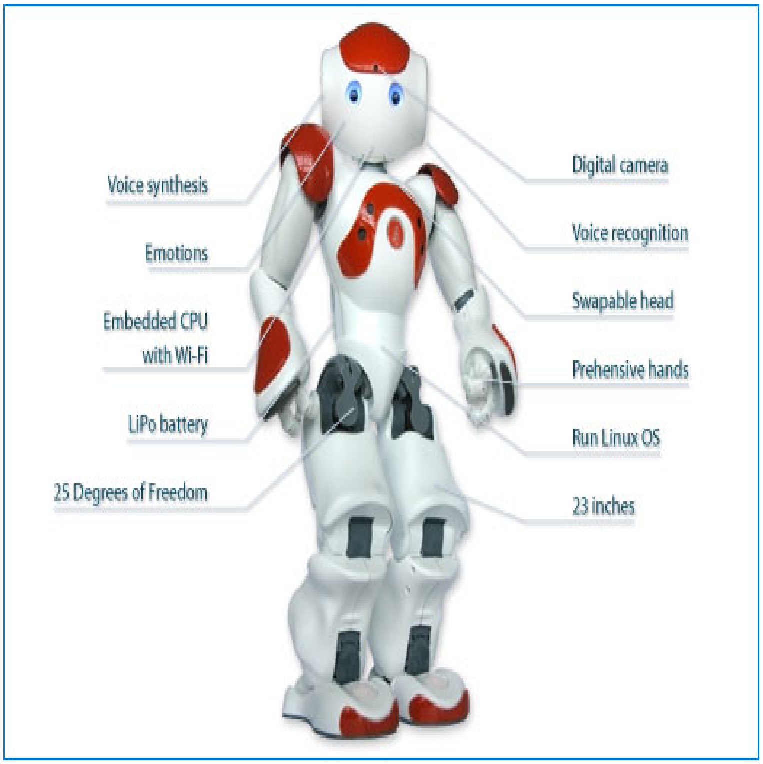
When it comes to pediatrics, humanoid robots can be seen as one of the most valuable companions. Humanoid robots are extensively being used in pediatrics healthcare as keenly seen in Jokinen and Wilcock (2014). A lot of the applications of the humanoid robots in this area comes with creating for the patient a calming effect or distraction from the procedures that are being performed at that moment. An important factor to be considered here is that the humanoids are expected to be human-like in such a way that it does not add to the discomfort of the subject in concern. This implies that the humanoid's body and face, including the humanoid's behavior and expressions, are completely human-like without adding any justification to the Uncanny valley theory Kozima *et al.* (2004). It can be seen from figure 5.1 how uncanny valley theory is proposed. The degree of acceptability depends on how close to looking human-like the humanoid under construction is and the more close it is, the more likable it gets, until a certain threshold, beyond which the resemblance enters an uncanny valley. This is the current state of most humanoids that have been developed. This acts as a major hurdle when it comes to acceptability of humanoids in every day world, thus making it the most important concern when it comes to humanoid facial design, since face is the first window of interaction between humans and humanoids.

Figure 5.1: Mori's "Uncanny Valley" Kozima *et al.* (2004)



Not only do these robots help in reducing pains but also put children at a higher chance of being vaccinated since the subjects already expect a distraction that can reduce or eliminate the pain caused during vaccination. As mentioned in Jokinen and Wilcock (2014), the humanoid robots help act as coaches for pain, motivation not just before the procedure is performed on the subject, but also during and after the procedure has been performed. While the humanoid can perform a number of tasks to ease the pain, the primary objective of acceptance by the subject still relies on how familiar/human-like it can look and behave.

Figure 5.2: NAO Robot's Features Jokinen and Wilcock (2014)



Humanoids are shown to be trained to demonstrate certain sets of exercises that are suitable to the particular subject that under therapy. For example, the robot NAO Jokinen and Wilcock (2014) has a solution called Zora Solution installed with it. This solution not only ensures the set of exercises are particularly suited for the patient, but its also important to note that, unlike human help, the robot that is trained to help the patient perform these exercises is capable of repeating the exercises accompanied with repeated verbal instructions as well which could or could not be supported by sound and music. This would not only serve as a solution with unlimited patience to provide the same set of services repeatedly with the same zeal every time, but also leads to major motivation to the children and seniors using these kinds of services.

Cerebral Palsy and Parkinson's are examples of a few applications of the humanoid robot with respect to therapy. Figure 5.2 shows how the NAO robot is built with its features highlighted.

5.1.3 Humanoids and Mental Health

Studies like Jokinen and Wilcock (2014) Taheri *et al.* (2015) show that the combination of humanoids and mental health can serve as a novel and successful combination in remedial services and psychological health. Autism is a condition where the subjects face difficulty in initiating and expressing speech. Humanoid robots can be trained to function in a certain way in a variety of scenarios that are medically suitable to help subjects with special conditions. As mentioned in Conti-Ramsden *et al.* (2006) , autism can be characterized by (1) poor communication skills, (2) abnormal play patterns, (3) difficult social interaction. Studies like these have shown that children with autism interact and react more positively with social robots than while interacting with their peers or parents. While some subjects concentrate extensively on a certain subject exhausting it completely, there are other subjects suffering with autism that fail to show any creativity or interest in any particular subject. There can also be subjects that can be obviously impaired with language abilities, failing to communicate at all, while there are other subjects that can constantly and continuously talk about their interests with letting their peers to be involved in the conversation. With humanoids, the structure and expressions that matter the most when it comes to holding an engaging conversation with subjects suffering with autism, can be tweaked and programmed in a way that suits the need of the situation. The expressions, again, is the most important factor to be considered when designing humanoids for autistic subjects.

When it comes to autism special education, research such as Jokinen and Wilcock

(2014) Taheri *et al.* (2015) have shown that humanoid robots can be designed so that they can interact and educate subjects with autism with applications designed specifically for such subjects. As mentioned in Jokinen and Wilcock (2014), the NAO robot was designed to be especially suitable for interaction with autistic subjects since the NAO robot is not only interactive, engaging, captivating and interesting but also adapt to the exact requirements of the classroom that the NAO robot is meant to deal with. The applications that can work with such robots can be designed to increase the level of comfort of such subjects while also making them confident thereby functioning as motivators based on the individual subject's personality.

5.1.4 Humanoids in Education

A major use of humanoid robots come from the field of education. There can be several different advantages of getting a humanoid involved in the daily classroom/ laboratory activities. As mentioned in Knoll (2002), the humanoid robots general usefulness in the field of education could come from two applications. One, the construction and programming of humanoids along with humanoid specific application development, can lead to high educational gains and understanding in how the humanoids work and how they can be improved. This kind of knowledge not only serves useful to the students themselves, but it also provides a possibility of useful advancements being contributed back by the students to the world of robotics which comes from thorough understanding of how the humanoids are built from scratch.

Figure 5.3: Example of Nonverbal Behavior of Humanoid in Educative Environments Brown and Howard (2013)

Gesture	Behavioral Meaning	Description of Motion
Eye Contact	Attention is directed towards an object	Head (eyes) is aligned with a specified target
Hand Wave	Goodbye	Arm is bent and raised next to head; forearm moves back and forth
Head Nod	Back-channel signal meaning continue; okay; yes	Head moves in an up and down motion
Head Shake	Negative connotation; sad; no	Head moves from side to side while facing the ground
Head Scratch	Confusion; lost	Arm/hand moves back and forth next to head (Fig. 3)
Fast Arm	Positive connotation; approval; excitement	Arm is bent and raised next to head; arm then quickly moves downward

Figure 5.4: Example of Verbal Responses of Humanoid in Educative Environments Brown and Howard (2013)

Answer	Speed	Phrase
Correct	Fast	"You're really good at this."
		"You're on fire!"
		"Awesome!"
	Slow	"You're doing great! I had trouble with that one too."
		"I appreciate the effort you're putting into this test."
		"This is hard, but we're doing great."
Incorrect	Fast	"Can you slow down a little so we can do it together."
		"You're leaving me behind."
		"Please wait for me."
	Slow	"This is really making us think."
		"This section is hard."
		"Don't sweat it; we'll get the next one."
None	Inactive	"Are you still there?"
		"Let's make an educated guess."
		"I was completely stumped on this one."

Figure 5.5: Example of Engagement Model Brown and Howard (2013)

Answer	Speed	User's Behavioral State
Correct	Fast	Engaged
		Not challenged enough
	Slow	Engaged
		Challenged
		Requires more time to think
Incorrect	Fast	Not engaged
		Unmotivated
		Not challenged (too hard/easy)
		Bored
	Slow	Engaged
		Challenged
		Struggling

As seen in figure 5.3, the humanoids that are deployed in environments for educationPantic and Rothkrantz (2004) can be used as entities to provide feedback in different forms. As it can be seen from the figure 5.3, the non-verbal responses form an important feedback mechanism to the students under training. This requires careful modeling and design of such features that make the robot more acceptable in an educational setting.

Along with nonverbal responses, the figure 5.4 shows verbal responses that can be used as feedback. The response can be programmed in various different ways as shown. It can be used for gentle motivation to aggressive celebratory tones. This kind of education helps the students in interacting with the humanoid more easy and more informative with the higher chances of students understanding in depth the topics discussed.

5.2 Research Questions

The research on humanoid robots and making them not only appear more human-like, that involves design of mechanical aspects of the humanoid and the material that is used to build the robot, but also making its behavior including expressions more human-like gives raise to a number of research questions.

Understanding that face expressions form one of the most important aspects of communication as the research in Schmidt and Cohn (2001) shows, it is of utmost interest to direct research towards facial expressions. Facial expressions are shown to form the most important part of human cognitive abilities Calder and Young (2005). Even so, a lot of the current research is concentrated and directed towards other aspects of humanoid design such as design of the mechanical body parts, like the research in Jokinen and Wilcock (2014) where the robot NAO is mentioned to be a robot capable of biped movement, but as seen from figure 5.2, the robot does not appear very human like, while still being able to communicate and show behavioral characteristics like humans. There's also a lot of research into just the biped movement of the humanoid like Miyakoshi *et al.* (1998) which shows that a lot of advancements have been going on when it comes to researching the biped movement of the humanoid robots. As shown in Yamaguchi *et al.* (1999) the two robots WABIAN and WABIAN-R Yamaguchi *et al.* (1999) are both developed to imitate human biped walking. While WABIAN is mentioned to be shorter with lesser degree of freedom, WABIAN-R is shown as a humanoid that is comparable to average human height while consisting of more degree of freedom. The question that is being raised in this paper is when such advancements have happened in the research of other aspects of humanoid robots, why is research on humanoid facial expressions so far behind? This boils down to to the question how the human cognition and artificially induced cognition/facial emotions function.

Another important question that is considered in this thesis is the extent to which the number of expressions exhibited by human beings are replicated in the artificial world. Darwin is known to have written about facial expressions, although the meaning of his research is not agreed upon by everyone. His greatest scientist status makes a frank assessment of his claims and its effects on what is now known of phylogenetic evolution is difficult. His 1872/1965 book has been read differently by different researchers owing to the vagueness in his conceptualization of emotion and expressions. His status has led to terming a lot of research on facial expressions as Darwinian, even when most of that work was not by Darwin himself. This time during which Darwin proposed his research was when facial expressions were thought of as being universal. Darwin not only recognized basic human emotions as "States of mind" that were mentioned in terms of those emotions but also as an effect of motivational, behavioral or personality traits, sensations and cognitive processes. A lot of research has been influenced from Darwin's ""States of mind" research Russell and Fernández-Dols (1997) and scientists have recognized 6 basic emotions Ekman (1992) and 21 total expressions. How many of these have been replicated and understood well is an important question identified here.

Even when we see a lot of research being put into the aspect of behavior of humanoids and advancing it towards making it more human like, as in the research of the social robot SAYA Hashimoto *et al.* (2006) it can be seen from figure 5.6 that the expressions are still very robot like.

Figure 5.6: Social Robot SAYA Making an Angry Face from Hashimoto *et al.* (2006)

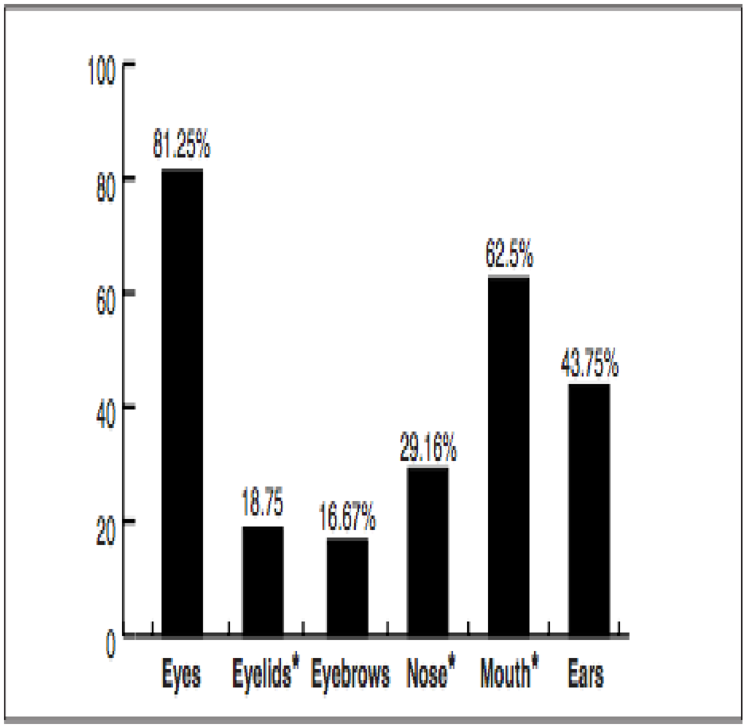
The disparity in the successful results of biped movements,movement of other body parts and the unsuccessful/ creepy results from facial expression design that qualify the robots as less human and more towards justification of uncanny valley 5.1 Kozima *et al.* (2004) is the next question raised in this paper. This thesis tries to determine the causes and any possible solutions to improve the facial expression design of the humanoid social robots.

As seen from chapter 2, the face recognition algorithms and classification algorithms especially the Eigenfaces recognition algorithm, use principal component analysis that identifies local and global features that are significant but may not be directly related to intuitive notion of face features such as the eyes, nose, lips and hair. This technique as opposed to the current design technique of identifying action units and muscles that contribute to a particular expression would serve as a good insight into how the design process of facial expressions. The identification of the implications of this change on this design process is the next question identified in this thesis.

5.3 Research Methods So Far

The design of humanoid robots' facial expressions is constituted with researching on identifying the muscles involved while certain expressions are displayed. Taheri *et al.* (2015) selects 32 action units or active muscles used in displayed different basic emotions. This method of using actuators to imitate the action units or active muscles on the humanoid robots' head along with appropriate material for skin is the current widespread research method involved. Survey has been the most important form of research used so far in understanding and analyzing whether the design of the humanoid constructed which is under evaluation qualifies the humanoid as human in the eyes of humans. DiSalvo *et al.* (2002) shows how surveys on 48 robots were used to understand how humans perceive them and the degree to which humans are comfortable with engaging in a conversation with a humanoid. The research and survey in DiSalvo *et al.* (2002) is used to understand crucial effects of facial features on humanoid robot's head as shown in figure 5.7.

Figure 5.7: Presence and Influence of Facial Features on Robot Heads from DiSalvo *et al.* (2002)



While this technique helps in understanding and extracting important behavioral information about humans including how they perceive different kinds of robots which also adds to the design process, it is obvious that a survey like this requires enough time and resources. Not only is the factor of time and resources a major influence on the results from a survey but the whole concept of survey requires that the subjects undertaking the survey will take it in a good spirit and fully-aware of the consequences of their responses. While some of the participants might take the surveys seriously, thereby contributing positively to the process of understanding the effects of the results from such surveys, there might be a larger population of the participants who do not take the survey without any exterior influences and factors. This will impact the results of the survey so that the conclusions obtained from such a survey may not be reliable. Fong *et al.* (2003) shows another research where survey is used to understand human preferences in the acceptance of humanoid robots. For example, Fong *et al.* (2003) mentions how the face of the humanoid head and degree of verbal communication incorporated into the humanoids make a difference in the acceptance of the robots by the participants in the survey. While research like this can add to important conclusions, Tourangeau and Yan (2007) mentions the issues in survey pertaining to sensitive questions in surveys, Cho and LaRose (1999) talks about privacy issues in surveys, these add to the other issues involved in correctness of the results from surveys.

Fong *et al.* (2003) also mentions how animation is used to generate 3D rendered images to model humanoid head with appropriate expressions. This is again used to conduct surveys and collect results about humanness of the humanoid 3D image.

So as to sum up, the research method currently used is designing and developing a social humanoid robot along with the features of its head and behavior that can qualify it as more human-like, following up with surveys that analyze how

5.4 Proposed Research Method

The research method proposed in this paper tries to take a new approach of incorporating deep learning into humanoid facial expression design. As mentioned in literature survey, under Facial recognition and classification algorithms, the EigenFaces, FisherFaces and Local Binary Pattern Histogram algorithms(references to which can be found in chapter 2 all work with the same idea in focus. The idea is of information retrieval. This idea of information retrieval is not only about retrieving important information but also making the information retrieval more efficient. The current approach that is used in facial expression design that involves actuators, is only somewhat efficient. This comes from the observation that it looks different to human eyes than what it is being represented as. Along with this experiments mentioned in later chapters also show that these images of social robots' facial expressions are not classified in par with the human facial expression images.

The observation picked up in this research is that while information leading to facial expression is identified, it is possible that some of the crucial information is being lost either due to implementation complexities or incorrect identification of important information as noise.

The research in this paper aims to bridge the gap in information retrieval by suggesting use of deep learning techniques for crucial and better information retrieval which is also efficient which can then lead to better design process.

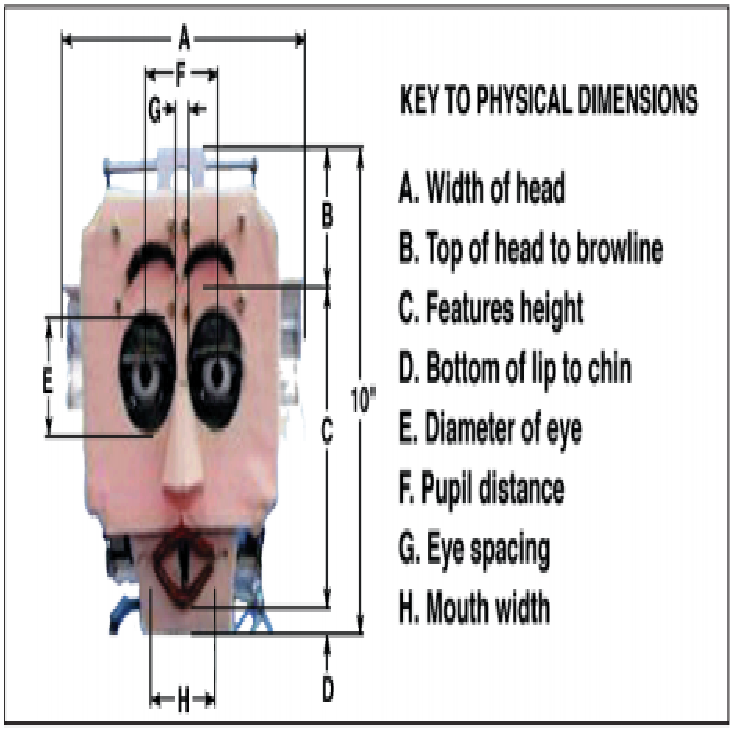
This deep learning technique is again used for the aspect of automating survey, where only images or information that pass of as being more human like through the software, which is more reliable and gives accurate results, are judged for acceptance by humans in the real world. This not only saves time but also is more efficient in the design process as well as understanding its humanness and acceptance.

Chapter 6

FACIAL DESIGN

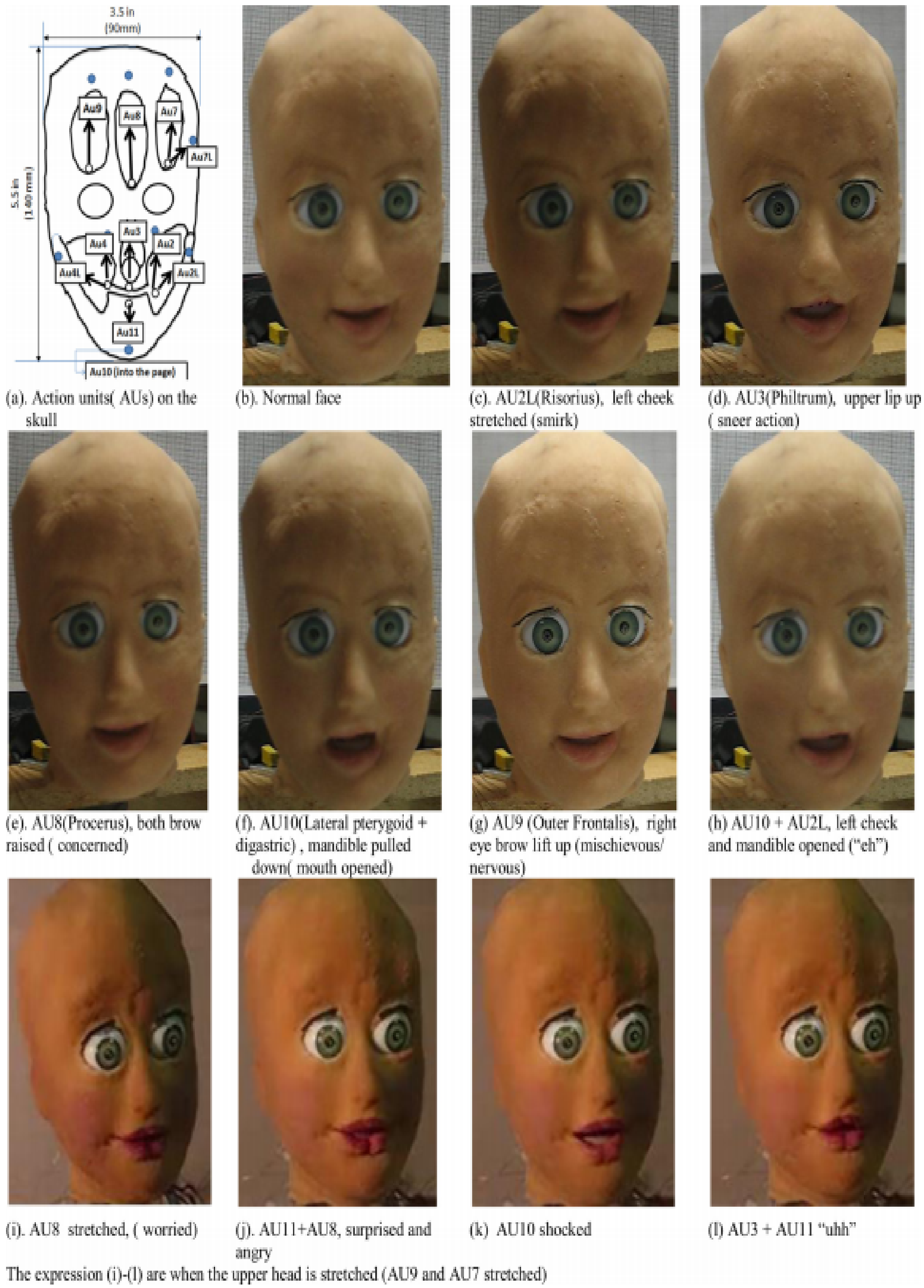
The human face has a varied array of functions. Not only does it display an individual's motivation, that makes a human's behavior more predictable, understandable to another observing human being, but also replaces/supplements it with signaling the speaker's attitude towards the information being spoken. As mentioned in earlier chapters, along with body language, facial expressions and face-to-face communication form a crucial part of human communication system. Human expressions can be classified into 6 basic expressions- happiness, sadness, fear, disgust, contempt, surprise. This has been identified in a number of works mentioned in earlier chapters and most of these works aim into incorporating these basic expressions into their humanoid head design. There have been a total of 21 expressions that have been identified as exhibited by human beings. While the aforementioned 6 emotions/expressions form the basis of face-to-face communication, with the action units/ keys points of movement identified in these 6 expressions, these action units are shown to be twitched to model different expressions. Facial gestures along with facial expressions form a major part of communication, such as a shrug to convey "I don't know" Breazeal (2004).

Figure 6.1: Physical Measures Taken for Humanoid Head Design DiSalvo *et al.* (2002)



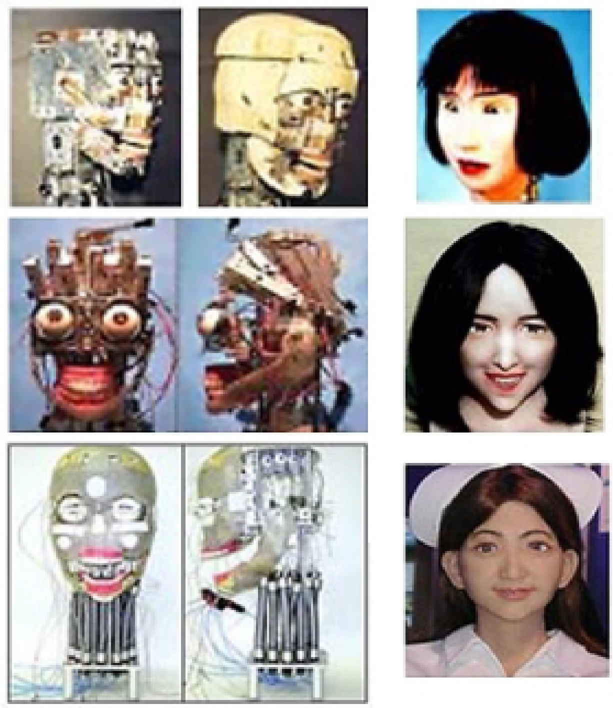
The figure 6.1 shows the measures of human head that is applied in the research of humanoid head design. These measures are again obtained from surveys involving a considerable number of human subjects. As mentioned in chapter 3, the research method of survey is used extensively for the humanoid head design. This forms as a basis to understand the key action points or active muscles involved in displaying the emotions/expressions. This can be seen from figure 6.2.

Figure 6.2: Various Facial Emotions from the Prototype Face Developed Using Shape Memory Alloy Actuators Tadesse and Priya (2012)



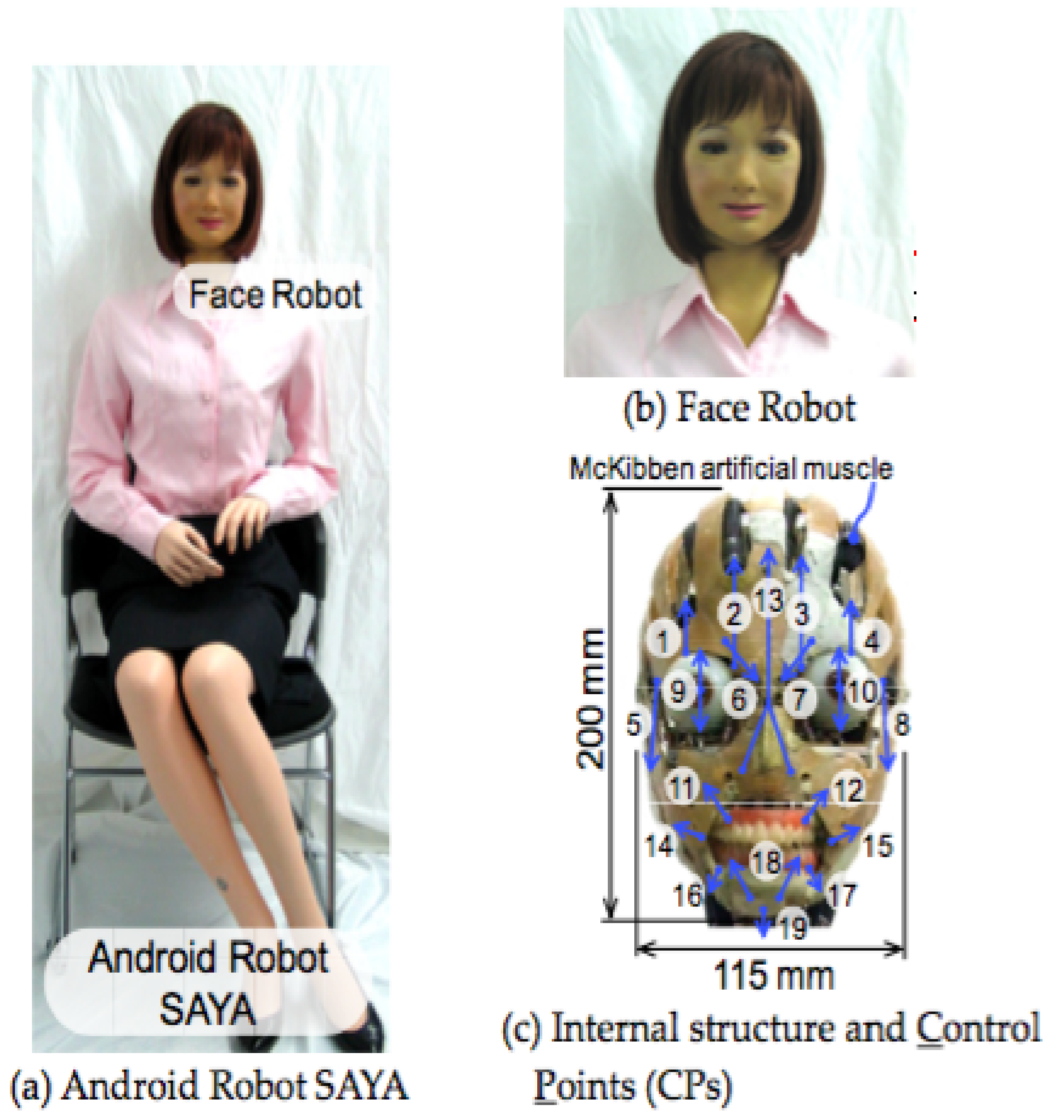
This highlights yet another important design feature as mentioned in [38], where action units are identified that contribute to display of facial expressions. Yet another example of this design technique is the figure 6.3 that shows the design of the social humanoid robot SAYA.

Figure 6.3: Saya Face Robots Fong *et al.* (2003)



The figure 6.3 incorporates several control points that are actuated under the "skin" to produce a varied array of facial movements and human expression. The control points can be seen in figure 6.4.

Figure 6.4: Saya Robot Hashimoto *et al.* (2006)



As mentioned in the chapter research context, the research method of using techniques from animation to model humanoid heads. The figure 6.5 shows a generated 3D image that models a humanoid head. The image is generated by graphical rendering and thus packs many degrees of freedom that are available for generating expressions. This face is rendered based on Delsarte's code for facial expressions Bruce *et al.* (2002).

Figure 6.5: Vikia-Computer Generated Face Fong *et al.* (2003)

Chapter 7

EXPERIMENTS

The research in this paper is motivated by the use of machine learning algorithms. As mentioned in chapter 2, under the section face recognition and classification algorithms, the recognition and classification algorithms work as information retrieval algorithms. The basic idea behind these various algorithms is efficient selection and information retrieval strategy that identifies and retains the most important information using which the quality of results obtained can be expected to have accurate measures. The recognition and classification algorithms in this thesis is first used to understand how it works on the data that is collected. The data consists of images acquired from already available databases. The recognition and classification algorithms work by shifting the bases on which information with highest variance is available to enable concentration of important information, as mentioned in chapter 2 as being the information with highest variance spread over the data set that is being considered, along one base and trying to reduce information on other bases. As detailed in earlier chapters these algorithms construct Eigen values and vectors which when added with the right co-efficient values give back almost all of the original image.

7.1 Limitations

The four experiments outlined below have some limitations that can be generalized between all of them. Firstly, only those algorithms that have been very successfully developed and proved to be highly useful through prior research have been used in these experiments. This includes EigenFaces, FisherFaces and Local Binary Pattern

Histograms. It has to be noted that these algorithms are used from the python library OpenCV.

The images used in these experiments have been taken in controlled environments. This indicates that the high performance of one of these algorithms on these images might be a factor of how these images have been taken owing to the dependence of these algorithms on the image. However, it is to be noted that when every time the algorithms are run on images clicked in the same manner as the ones in these experiments, the results can be expected to be obtained similarly. This eliminates the need for different kinds of images to be clicked and passed through these algorithms. However, it is possible to determine the effect of other kinds of images on these algorithms and repeat these experiments in the future.

The mechanical social robots developed might have their own limitations in development which might make it possible that the results remain consistently bad over many trials. This aspect, however, is out of scope in this paper and shall not be taken as a factor for the results obtained since the idea of these experiments is to understand the limitations of the mechanical social robots itself.

7.2 Experiment 1: Evaluation of EigenFaces, FisherFaces and LBPH Algorithms on Human Face Image Database

The experiment involves a training data set and a prediction data set both of which are fed to the three of the algorithms. The experiments work with first preparing the dataset to suit the format in which the three algorithms accept the input.

The pre-processing/preparation of data, in a nutshell, involves converting the image into grayscale and cropping the images by appropriate identification of face in the image, making sure that the entire training data set contains images of the same dimensions. The pre-processing step is then followed by creating the required classifier and training it with the training data set to identify the facial expressions. This is then followed by predicting the facial expression of the prediction data set.

The technical requirements of this experiment is:

Anaconda Version 4.2.13

PyCharm Community Edition 2016.3

Build PC-163.8233.8, built on November 22, 2016

JRE: 1.8.0_112-release-408-b2 x86_64

JVM: OpenJDK 64-Bit Server VM by JetBrains s.r.o

The experiment uses Anaconda OpenCV 2.4.8. This library contains functions extensively used in face recognition and classification.

The training set is a database of images of human faces exhibiting different facial expressions/empotions ranging across- happiness, sadness, fear, disgust, contempt, surprise, neutral. The images in the dataset are taken from Kanade *et al.* (2000). The images contain FACS coded files to indicate the label of the expression that the image represents. The FACS code is obtained from the FACS action units according to the table 7.1.

Table 7.1: FACS Action Units Kanade *et al.* (2000)

AU	Facial muscle	Description of muscle movement
1	Frontalis, pars medialis	Inner corner of eyebrow raised
2	Frontalis, pars lateralis	Outer corner of eyebrow raised
4	Corrugator supercilii, Depressor supercilii	Eyebrows drawn medially and down
5	Levator palpebrae superioris	Eyes widened
6	Orbicularis oculi, pars orbitalis	Cheeks raised; eyes narrowed
7	Orbicularis oculi, pars palpebralis	Lower eyelid raised and drawn medially
9	Levator labii superioris alaeque nasi	Upper lip raised and inverted; superior part of the nasolabial furrow deepened; nostril dilated by the medial slip of the muscle
10	Levator labii superioris	Upper lip raised; nasolabial furrow deepened producing square-like furrows around nostrils
11	Levator anguli oris (a.k.a. Caninus)	Lower to medial part of the nasolabial furrow deepened
12	Zygomaticus major	Lip corners pulled up and laterally
13	Zygomaticus minor	Angle of the mouth elevated; only muscle in the deep layer of muscles that opens the lips
14	Buccinator	Lip corners tightened. Cheeks compressed against teeth
15	Depressor anguli oris (a.k.a. Triangularis)	Corner of the mouth pulled downward and inward
16	Depressor labii inferioris	Lower lip pulled down and laterally
17	Mentalis	Skin of chin elevated
18	Incisivii labii superioris and Incisivii labii inferioris	Lips pursed
20	Risorius w/ platysma	Lip corners pulled laterally
22	Orbicularis oris	Lips everted (funneled)
23	Orbicularis oris	Lips tightened
24	Orbicularis oris	Lips pressed together
25	Depressor labii inferioris, or relaxation of mentalis, or orbicularis oris	Lips parted
26	Masseter; relaxed temporal and internal pterygoid	Jaw dropped
27	Pterygoids and digastric	Mouth stretched open
28	Orbicularis oris	Lips sucked

Table 7.2: FACS Action Units(continued) Kanade *et al.* (2000)

AU	Facial muscle	Description of muscle movement
41	Relaxation of levator palpebrae superioris	Upper eyelid droop
42	Orbicularis oculi	Eyelid slit
43	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Eyes closed
44	Orbicularis oculi, pars palpebralis	Eyes squinted
45	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Blink
46	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Wink

Some examples of the images and their associated FACS coded file are shown in the figures below. The training set contains of a total of 521 images each showing different sets of emotions by the same subject. The training dataset is passed through a face recognizer, which works along the explanation in literature survey of facial recognition algorithms. The recognizers used in this experiment are haarcascade_frontalface_default, haarcascade_frontalface_alt2, haarcascade_frontalface_alt, haarcascade_frontalface_alt_tree. Each of these are pre-trained recognizers avaiable as part of openCV library, haarcascades package.

Figure 7.1: Face Image 1 Kanade *et al.* (2000)

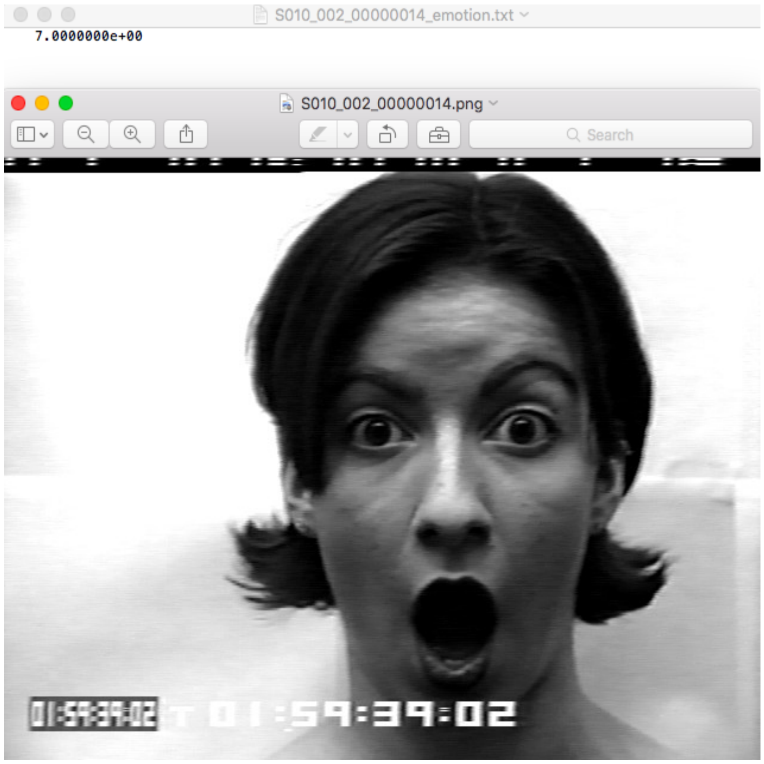


Figure 7.2: Face Image 2 Kanade *et al.* (2000)

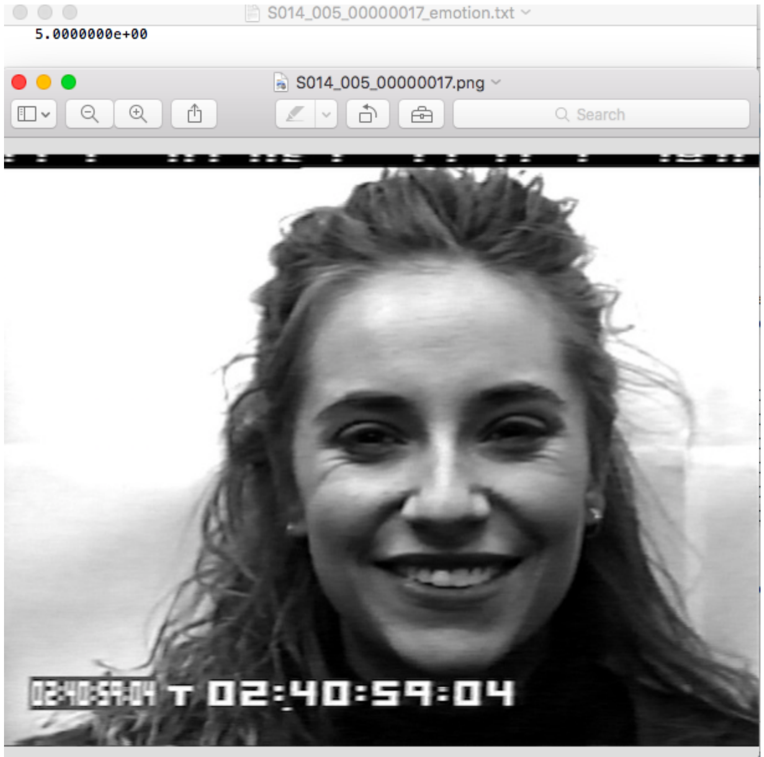


Figure 7.3: Face Image 3 Kanade *et al.* (2000)

Each of these images are identified in series for maximum feature recognition (details of which can be found under literature survey of face recognition algorithms) and cropped to the same dimensions. The cropped images are shown below

Figure 7.4: Cropped Grayscale Recognized Face Image 1 Kanade *et al.* (2000)

Figure 7.5: Cropped Grayscale Recognized Face Image 2 Kanade *et al.* (2000)

Figure 7.6: Cropped Grayscale Recognized Face Image 3 Kanade *et al.* (2000)

These images are then passed on to three different classifiers, FisherFaces, Eigen-Faces, Local Binary Pattern Histograms classifiers (references and details in chapter 2).

The prediction data set in this experiment is again human face databases with the aforementioned 7 expressions that is taken from Kanade *et al.* (2000) and a total of 119 images from the database are used for prediction at random selection. Some of the limitations of this experiment are:

1) Images clicked in a controlled environment

2) Only frontal view images

3) No profile views

4) Since the pictures are posed, expressions are not natural in a few sequences.

7.3 Experiment 2: Evaluation of EigenFaces, FisherFaces and LBPH Algorithms on Social Robot Face Images Database

The second experiment runs on the same principles as the first experiment outlined above. The first experiment is intended to understand how the classifiers and recognizers run on the database containing human face images only.

This experiment uses 521 human face images as part of training data set. A data set containing a total of 36 social robot head images is used as the prediction data set. The 36 robot images are taken from the internet.

The aim of this experiment is to identify the prediction success rates on robot face images as prediction set when the classifiers are trained on human face images showcasing similar expressions. This accounts to the humanness of the robot face images. The results can be used accurately since they do not depend on human altered surveys.

The second experiment's technical requirements do not change from the first experiment's requirements. The recognition and classification for this experiment is run on pycharm editor, using anaconda openCV library, the four recognizers used are haarcascade_frontalface_default, haarcascade_frontalface_alt2, haarcascade_frontalface_alt, haarcascade_frontalface_alt_tree recognizers and FisherFaces, EigenFaces and Local Binary Pattern Histogram classifiers.

Below shown are some of the robot face images and their images after face recognition, gray scaling and cropping.

Figure 7.7: Social Robot SOPHIA Neutral Image (http://www.hansonrobotics.com)

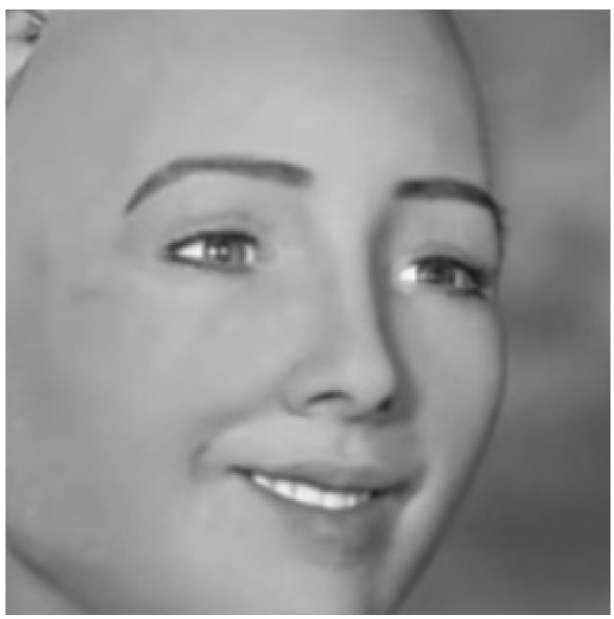
Figure 7.8: Cropped Grayscale Recognized Robot SOPHIA Image

Figure 7.9: Social Robot FACE Fear Image

Figure 7.10: Cropped Grayscale Recognized Robot FACE Image

Figure 7.11: Social Robot Sophia Happy Image(http://www.hansonrobotics.com)

Figure 7.12: Cropped Grayscale Recognized Robot SOPHIA Image



These are some of the robot images used in this experiment. The 36 images of robots used in this experiment consist of different fidelity designs of robots ranging from most human-like to least human like.

7.4 Experiment 3: Segregation of Robots With Different Design Fidelity Using the Prediction Confidence Levels of Classification Algorithms

The aim of this experiment is to evaluate and automate the identification of humanness of humanoid robot using the deep learning algorithms. The motivation behind this experiment is to evaluate how elimination of survey regarding humanness of a humanoid robot design can prove to be successful.

The three algorithms that are used in this experiment, FisherFaces, EigenFaces and Local Binary Pattern Histogram while predicting the images in the prediction set, also provide the confidence levels of every prediction. When a prediction is successful, the confidence levels can be used to establish thresholds beyond which a humanoid robot can be considered as high-fidelity or medium fidelity or low-fidelity humanoid robot. The three classifiers, FisherFaces, EigenFaces and Local Binary Pattern Histogram are trained using 521 human face images taken from Kanade *et al.* (2000). Examples of different kinds of robots used as prediction set in this experiment are shown below.

Figure 7.13: Image of Nexi Robot(robotic.media.mit.edu)

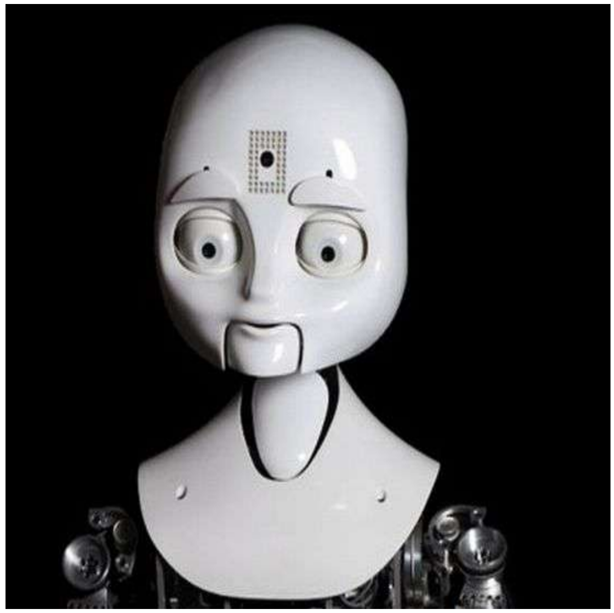


Figure 7.14: Image of Nadine Robot (http://imi.ntu.edu)

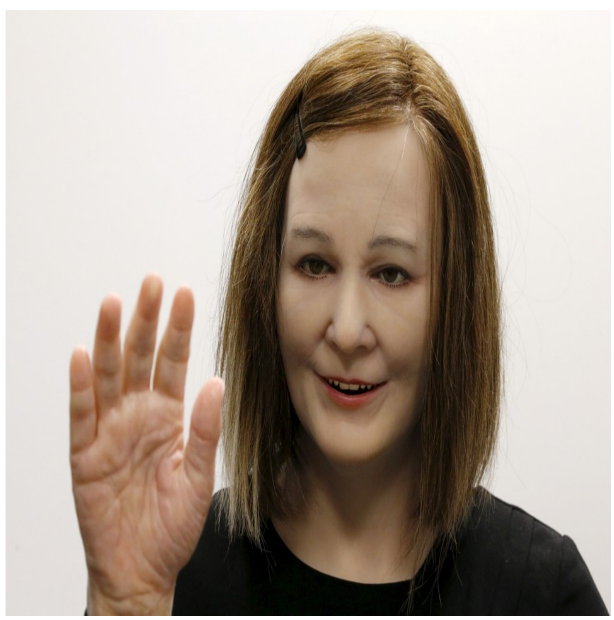
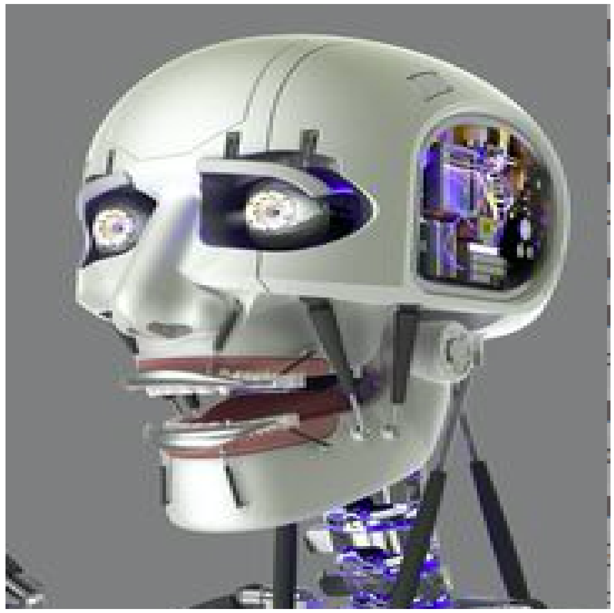


Figure 7.15: Image of Robot Displaying Fear



7.5 Experiment 4: Evaluating Performance of the Learning Algorithms on Ethnically Similar Training Dataset and Prediction Data Set

The aim of this experiment is to understand whether design of facial expressions of social robots is affected by the research involving subjects that are ethnically similar. The training data set in this experiment is taken from Lyons *et al.* (1998) and contains a total of 168 images of human faces with the following expressions: Happiness, sadness, fear, anger, disgust, surprise, neutral. The images are coded with the subject label followed by emotion label and the image number.

The examples of training data set is as shown in the figures below.

Figure 7.16: Human Face Image Displaying Anger Lyons *et al.* (1998)

Figure 7.17: Human Face Image Displaying Digust Lyons *et al.* (1998)

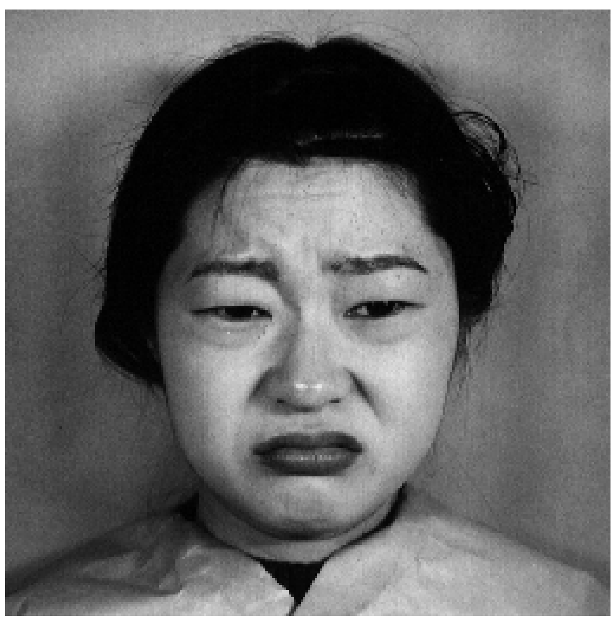
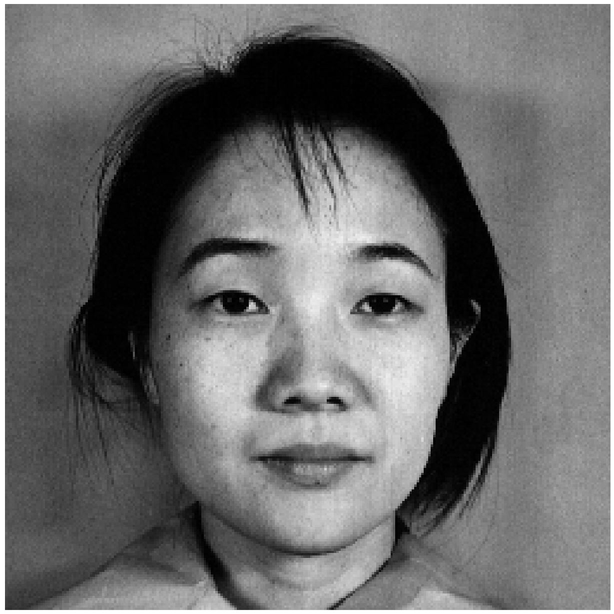


Figure 7.18: Human Face Image Displaying Happiness Kanade *et al.* (2000)



Since the appearance of humans of different ethnicities is different, it will serve as a good input into the research of facial expressions when the results of training set containing human face images from a particular ethnicity and prediction set images being of robots that look ethnically similar to the ones in the training set.
The prediction data set contains 29 robot images that are similar to training data set in ethnicity. The examples of some of the images chosen are shown in the figures below.

Figure 7.19: Robot Face Image 1 Displaying Anger

Figure 7.20: Robot Face Image 2 Displaying Anger

Figure 7.21: Robot Face Image 3 Displaying Anger

The ethnicity that was chosen in this experiment was Asian. This was motivated by the number of Asian-looking robots/ humanoids that have been developed so far. This gives the experiment enough data to work with while making the results more reliable.

The limitations of this experiments were:

1) Asian as an ethnicity is widely varying. There are facial differences within the ethnic group itself. This has not been accounted for in this experiment, citing the reason that the variance between Asian people does not vary by huge differences.

2) The training data is gathered in a controlled environment, where the expressions are similar to the others or not natural. This can be seen from figure 7.18.

7.6 Code snippets for Preparation, Recognition and Classification of Images

7.6.1 Preparing the Dataset Including Recognition of Face in the Dataset

The code snippet below is the code that is used to prepare the dataset for passing as input into the classification algorithms. The preparation step involves identification of faces in the face images of the current database that is being used as the dataset for training/prediction. The libraries used in this code are released under BSD license(open source). The code snippet is inspired from the blogpost by van Gent (2016).

```
import cv2
import glob
#using 4 face recognizers in order, any face not discovered in order is
   discarded
faceDet =
   cv2.CascadeClassifier("haarcascades/haarcascade_frontalface_default.xml")
faceDet2 =
   cv2.CascadeClassifier("haarcascades/haarcascade_frontalface_alt2.xml")
faceDet3 =
   cv2.CascadeClassifier("haarcascades/haarcascade_frontalface_alt.xml")
faceDet4 =
   cv2.CascadeClassifier("haarcascades/haarcascade_frontalface_alt_tree.xml")
#define emotions for classification
emotions = ["neutral", "anger", "contempt", "disgust", "fear", "happy",
   "sadness", "surprise"]
def detect_faces(emotion):
   files = glob.glob("jaffe-robots//%s//*" % emotion)
   # Get list of all images with emotion
```

```python
filenumber = 0
for f in files:
    frame = cv2.imread(f) # Open image
    gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY) # Convert image to
        grayscale

    # Detect face using 4 different classifiers
    face = faceDet.detectMultiScale(gray, scaleFactor=1.1,
        minNeighbors=10, minSize=(5, 5),
                                    flags=cv2.CASCADE_SCALE_IMAGE)
    face2 = faceDet2.detectMultiScale(gray, scaleFactor=1.1,
        minNeighbors=10, minSize=(5, 5),
                                      flags=cv2.CASCADE_SCALE_IMAGE)
    face3 = faceDet3.detectMultiScale(gray, scaleFactor=1.1,
        minNeighbors=10, minSize=(5, 5),
                                      flags=cv2.CASCADE_SCALE_IMAGE)
    face4 = faceDet4.detectMultiScale(gray, scaleFactor=1.1,
        minNeighbors=10, minSize=(5, 5),
                                      flags=cv2.CASCADE_SCALE_IMAGE)

    # Go over detected faces, stop at first detected face, return empty
        if no face.
    if len(face) == 1:
        facefeatures = face
    elif len(face2) == 1:
        facefeatures == face2
```

```
        elif len(face3) == 1:
            facefeatures = face3
        elif len(face4) == 1:
            facefeatures = face4
        else:
            facefeatures = ""

        # Cut and save face
        for (x, y, w, h) in facefeatures: # get coordinates and size of
            rectangle containing face
            print "face found in file: %s" % f
            gray = gray[y:y + h, x:x + w] # Cut the frame to size

            try:
                out = cv2.resize(gray, (350, 350)) # Resize face so all
                    images have same size
                cv2.imwrite("jaffe-robot-dataset//%s//%s.jpg" % (emotion,
                    filenumber), out) # Write image
            except:
                pass # If error, pass file
        filenumber += 1 # Increment image number


for emotion in emotions:
    detect_faces(emotion)
```

7.6.2 Training the Classifier With Training Dataset Prepared from Previous Step and Feeding the Prediction Dataset for Classification

This code snippet outlines the training of classifier, the code specifically shows training of FisherFace classifier while the same code is changed to achieve training of a different classifier. The code is then passed a prediction dataset and the prediction of all images is run for 10 times. The average of the results is evaluated as the final result of prediction success percentage. This code is also inspired from the blogpost by van Gent (2016).

```
import cv2
import glob
import random
import numpy as np
import pickle

emotions = ["neutral", "anger", "contempt", "disgust", "fear", "happy",
    "sadness", "surprise"] # Emotion list
fishface = cv2.createFisherFaceRecognizer()
# Initialize fisher face classifier
#the above line can be changed to train different classifiers
data = {}


def get_files(emotion): # Define function to get file list, randomly
    shuffle it and split 80/20
    files1 = glob.glob("jaffe-robot-dataset//%s//*" % emotion)
    files2 = glob.glob("robot-dataset//%s//*" % emotion)
```

```
    random.shuffle(files1)
    training = files1[:int(len(files1) * 0.8)] # get first 80% of file list
    prediction = files2
    return training, prediction


def make_sets():
    training_data = []
    training_labels = []
    prediction_data = []
    prediction_labels = []
    for emotion in emotions:
        training, prediction = get_files(emotion)
        # Append data to training and prediction list, and generate labels
            0-7
        for item in training:
            image = cv2.imread(item) # open image
            gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY) # convert to
                grayscale
            training_data.append(gray) # append image array to training
                data list
            training_labels.append(emotions.index(emotion))

        for item in prediction: # repeat above process for prediction set
            image = cv2.imread(item)
            gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
            prediction_data.append(gray)
```

```
            prediction_labels.append(emotions.index(emotion))

    return training_data, training_labels, prediction_data,
        prediction_labels


def run_recognizer():
    training_data, training_labels, prediction_data, prediction_labels =
        make_sets()

    print "training fisher face classifier"
    print "size of training set is:", len(training_labels), "images"
    fishface.train(training_data, np.asarray(training_labels))

    print "predicting classification set"
    cnt = 0
    correct = 0
    incorrect = 0
    for image in prediction_data:
        pred, conf = fishface.predict(image)
        if pred == prediction_labels[cnt]:
            correct += 1
            cnt += 1
        else:
            incorrect += 1
            cnt += 1
    return ((100 * correct) / (correct + incorrect))
```

```python
# Now run it
metascore = []
for i in range(0, 10):
    correct = run_recognizer()
    print "got", correct, "percent correct!"
    metascore.append(correct)

print "\n\nend score:", np.mean(metascore), "percent correct!"
```

Chapter 8

RESULTS AND DISCUSSION

8.1 Results

The results from experiment 1 is outlined in table 8.1. The results show a very high prediction rate with the FisherFaces algorithm as opposed to EigenFaces or Local Binary Pattern Histogram. This is justified by the literature survey in chapter 2. The FisherFaces algorithms classify the algorithms based on classes, thereby trying to automate the identification or reference image formation of the required emotion and/or person. The classification is segregated by classes. This helps better and efficient classification, thereby showing excellent results as opposed to the other two algorithms.
The result from this experiment, where it is observed that FisherFaces works best, is intended to be suggested for use in the facial expression design.

Table 8.1: Prediction Results on Human Images Database Over 10 Trials

Classifier	Average successful prediction (%)
FisherFaces	81.7
EigenFaces	46
Local Binary Pattern Histogram	48

The results of experiment 2 are shown in table 8.2. Here also, we can see that the FisherFace algorithm is the best algorithm when it comes to classification. But the major observation to be taken away from this experiment is that when the same

training set is used to train the classifiers and the prediction set is changed from human images to humanoid face images, the results of even the best working algorithm drops by more that half of when the human images were used as prediction set.
This experiment and the results from this experiment highlights the fact that even the most advanced facial designs fail at being recognized as the appropriate facial expression.

Table 8.2: Prediction Results on Robot Images Database Over 10 Trials

Classifier	Average successful prediction (%)
FisherFaces	25.9
EigenFaces	13.7
Local Binary Pattern Histogram	13.9

The results from experiment 3, which involved prediction set containing humanoid face images of different design fidelity, as intended failed to even pass the first recognition step for the low-fidelity humanoids. This is a major milestone in this experiment since the surveying of humanoid images to rank them by their humanness can be eliminated by just passing them through the recognizer. Any images that do pass through the recognizer are already enough human like and require additional research into the facial expressions. This can be said because the images that did pass through the recognizer failed to be classified as the correct emotion. This is seen as a result of two factors
1) The data set used for training was in controlled environment. The natural expressions of human beings are not staged. This has an effect on the way the prediction set is classified.
2) The facial expression design is not accurate enough.

Table 8.3: Prediction Results on Ethnically Similar Robot Face Images Database Over 10 Trials

Classifier	Average successful prediction (%)
FisherFaces	20.3
EigenFaces	21.9
Local Binary Pattern Histogram	13.1

The results from the last experiment can be seen in 8.3. The aim of this experiment was to analyze whether the performance of experiment 2 can be improved by using ethnically similar training and prediction data set. This experiment is used to evaluate whether design considerations of ethnicities and the functioning of classification algorithms work differently.
As seen from the table 8.3, the results from experiment 2 and this experiment are almost identical, with very negligible differences. This can be accounted to concluding that the facial recognition and classification of face images are not dictated by ethnicity, thereby suggestions made from analyzing the results of these experiments are valid in any condition.

8.2 Discussion

The motivation of this thesis was to bridge the gap between what humans see as human-like and what the humanoid design features portray as being human-like. It is evident from a couple of figures from chapter 6 that the facial expression that the humanoid is intended to display does not agree well with the human eye.
The first research question was how the human cognition and artificially constructed facial systems differ. This can be seen as an information retrieval problem. Human

cognitive system retrieves a lot more important information that help the proper identification of the expression that the other person is displaying. This also helps the humans with the display of expressions. Imitation is the key to learning of expressions.

The current research method extensively relies on actuating the facial muscles artificially on the robot faces to enable display of expressions. But as seen with the experiments, the facial recognition and classification algorithms fail to identify the expression. This shows that some information is being missed from the design of facial expression of social robots currently.

As mentioned in chapter 2, the facial classification and recognition algorithms work well because of reduction of bases and maximization of variance along the reduced base. The research method that is proposed after the careful consideration of the results from the experiments is that instead of actuation, the design should progress from first developing design outlines from data obtained through the learning algorithms, since learning algorithms preserve important information more accurately than just measuring facial features that help actuation. This would help in construction of a facial image since as mentioned in chapter two, the basis images that are obtained in the process of training the classifiers can be combined to obtain an image that is almost similar to the training image that was used.

This implies that a training data set containing of human images is to be used, which will then give rise to basis images, that which contain the most importance variance information, which then can be used to produce a robot facial design that will have the important features for it to be recognized as human.

The second research question was how automation or learning algorithms like these can help with survey aspect of facial expression design. As seen from the previous point, not only do these algorithms help in facial design, but can also be used to evaluate humanness, thereby eliminating the need to survey first.

Therefore, the image can be constructed using basis images' information, passed through the recognition algorithms for an initial humanness measure and then passed on to actual survey that would help in further modifications, This reduces the time and effort in surveying so many images, as well as improves the design process of facial expression of social robots.

Chapter 9

CONCLUSION

The present humanoid facial expression design technique that has been motivated by actuators fail to identify certain important features in the human face which is caught by the facial recognition and classification algorithms, as seen from the experiments. This implies that the with the actuation method information retrieval fails to account for important information. It is therefore suggested based on the experimental outcomes that when the design process progresses from establishing a design using basis images formed from the learning algorithms, most important information can be retrieved, which can then be used for successful construction of facial image that can be translated into the mechanical humanoid face design.

Chapter 10

FUTURE WORK

The major step after the identification of design technique of facial expression of social robots is how this design is translated to the mechanical design phase. This future work is understanding how the basis images can be combined in a useful manner and what each of the coefficients of the basis images should be chosen so that the reconstructed image can give rise to an accurate facial expression design. The main future work still remains understanding what these artificially learnt facial features that form the basis for humanizing the facial expressions correspond to in the mechanical face.

The information that is currently being missed in the actuation process is caught by human eye and the recognition and classification algorithms, but is unknown to the research world. Identification of the exact information that would add to the facial expression research can be made sense of using the learning algorithms. If identified correctly, it can be used to improve actuation process of facial design as well.

Understanding how clear statements like formal statements, example: programming languages, that are unambiguous, Natural Language Interface that depend on language and context, which are ambiguous and Body Language Interface that includes verbal, non-verbal and para-verbal(pronunciation) language which have the highest ambiguity affect the facial expression display and cognition/recognition.

The same set of experiments can be run over multiple trials on different databases with images clicked in different conditions, including angles, expressions and lighting. This might reveal further details about the the algorithms used in these experiments. Along with this, different algorithms can also be taken into consideration.

REFERENCES

Ahonen, T., A. Hadid and M. Pietikäinen, *Face Recognition with Local Binary Patterns*, pp. 469–481 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004), URL `http://dx.doi.org/10.1007/978-3-540-24670-1_36`.

Aly, A. and A. Tapus, "Prosody-driven robot arm gestures generation in human-robot interaction", in "Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction", pp. 257–258 (ACM, 2012).

Animation, P., "Pixar's animation process", URL `http://pixar-animation.weebly.com/pixars-animation-process.html` (1998).

Belhumeur, P. N., J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", IEEE Transactions on pattern analysis and machine intelligence **19**, 7, 711–720 (1997).

Berlin, M., C. Breazeal and C. Chao, "Spatial scaffolding cues for interactive robot learning", in "Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on", pp. 1229–1235 (IEEE, 2008).

Billard, A. and K. Dautenhahn, "Grounding communication in situated, social robots", in "Proceedings Towards Intelligent Mobile Robots Conference, Report No. UMCS-97-9-1, Department of Computer Science, Manchester University", (1997).

Billard, A. and K. Dautenhahn, "Grounding communication in autonomous robots: an experimental study", Robotics and Autonomous Systems **24**, 1-2, 71–79 (1998).

Breazeal, C., "Proto-conversations with an anthropomorphic robot", in "Robot and Human Interactive Communication, 2000. RO-MAN 2000. Proceedings. 9th IEEE International Workshop on", pp. 328–333 (IEEE, 2000).

Breazeal, C. L., *Designing sociable robots* (MIT press, 2004).

Brown, L. and A. M. Howard, "Engaging children in math education using a socially interactive humanoid robot", in "Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on", pp. 183–188 (IEEE, 2013).

Bruce, A., I. Nourbakhsh and R. Simmons, "The role of expressiveness and attention in human-robot interaction", in "Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on", vol. 4, pp. 4138–4142 (IEEE, 2002).

Burleson, W., R. Picard, K. Perlin and J. Lippincott, "A platform for affective agent research", in "Workshop on Empathetic Agents, International Conference on Autonomous Agents and Multiagent Systems, Columbia University, New York, NY", vol. 2 (Citeseer, 2004).

Calder, A. J. and A. W. Young, "Understanding the recognition of facial identity and facial expression", Nature Reviews Neuroscience **6**, 8, 641–651 (2005).

Castellano, G. and P. W. McOwan, “Analysis of affective cues in human-robot interaction: a multi-level approach”, in “Image Analysis for Multimedia Interactive Services, 2009. WIAMIS’09. 10th Workshop on”, pp. 258–261 (IEEE, 2009).

Chidambaram, V., Y.-H. Chiang and B. Mutlu, “Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues”, in “Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction”, pp. 293–300 (ACM, 2012).

Cho, H. and R. LaRose, “Privacy issues in internet surveys”, Social Science Computer Review **17**, 4, 421–434 (1999).

Conti-Ramsden, G., Z. Simkin and N. Botting, “The prevalence of autistic spectrum disorders in adolescents with a history of specific language impairment (sli)”, Journal of Child Psychology and Psychiatry **47**, 6, 621–628 (2006).

Craig, S. D., S. D’Mello, A. Witherspoon and A. Graesser, “Emote aloud during learning with autotutor: Applying the facial action coding system to cognitive–affective states during learning”, Cognition and Emotion **22**, 5, 777–788 (2008).

Das, D., M. M. Hoque, T. Onuki, Y. Kobayashi and Y. Kuno, “Vision-based attention control system for socially interactive robots”, in “RO-MAN, 2012 IEEE”, pp. 496–502 (IEEE, 2012).

Databases, F. R. H., “Face recognition homepage - databases”, URL `http://www.face-rec.org/databases/` (2003).

Dautenhahn, K., “Getting to know each other,artificial social intelligence for autonomous robots”, Robotics and autonomous systems **16**, 2-4, 333–356 (1995).

DiSalvo, C. F., F. Gemperle, J. Forlizzi and S. Kiesler, “All robots are not created equal: the design and perception of humanoid robot heads”, in “Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques”, pp. 321–326 (ACM, 2002).

D’Mello, S. K., S. D. Craig and A. C. Graesser, “Multimethod assessment of affective experience and expression during deep learning”, International Journal of Learning Technology **4**, 3-4, 165–187 (2009).

Edsinger, A. and U.-M. O’Reilly, “Designing a humanoid robot face to fulfill a social contract”, in “Proc. 9th IEEE Ro-Man”, (2000).

Ekman, P., “Are there basic emotions?”, (1992).

Embgen, S., M. Luber, C. Becker-Asano, M. Ragni, V. Evers and K. O. Arras, “Robot-specific social cues in emotional body language”, in “RO-MAN, 2012 IEEE”, pp. 1019–1025 (IEEE, 2012).

Finke, M., K. L. Koay, K. Dautenhahn, C. L. Nehaniv, M. L. Walters and J. Saunders, “Hey, i’m over here-how can a robot attract people’s attention?”, in “Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on”, pp. 7–12 (IEEE, 2005).

Fong, T., I. Nourbakhsh and K. Dautenhahn, “A survey of socially interactive robots”, Robotics and autonomous systems **42**, 3, 143–166 (2003).

Forbes-Riley, K. and D. Litman, “Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor”, Speech Communication **53**, 9, 1115–1136 (2011).

Freund, Y. and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting”, Journal of Computer and System Sciences **55**, 1, 119 – 139, URL `http://www.sciencedirect.com/science/article/pii/S002200009791504X` (1997).

Gaschler, A., S. Jentzsch, M. Giuliani, K. Huth, J. de Ruiter and A. Knoll, “Social behavior recognition using body posture and head pose for human-robot interaction”, in “Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on”, pp. 2128–2133 (IEEE, 2012).

Hashimoto, T., S. Hitramatsu, T. Tsuji and H. Kobayashi, “Development of the face robot saya for rich facial expressions”, in “SICE-ICASE, 2006. International Joint Conference”, pp. 5423–5428 (IEEE, 2006).

Hegel, F., S. Gieselmann, A. Peters, P. Holthaus and B. Wrede, “Towards a typology of meaningful signals and cues in social robotics”, in “RO-MAN, 2011 IEEE”, pp. 72–78 (IEEE, 2011).

Hirose, M. and K. Ogawa, “Honda humanoid robots development”, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **365**, 1850, 11–19 (2007).

Hizem, W., E. Krichen, Y. Ni, B. Dorizzi and S. Garcia-Salicetti, “Specific sensors for face recognition”, in “International Conference on Biometrics”, pp. 47–54 (Springer, 2006).

Huang, G. B., M. Ramesh, T. Berg and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments”, Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst (2007).

Jayagopi, D. B., S. Sheiki, D. Klotz, J. Wienke, J.-M. Odobez, S. Wrede, V. Khalidov, L. Nyugen, B. Wrede and D. Gatica-Perez, “The vernissage corpus: A conversational human-robot-interaction dataset”, in “Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction”, pp. 149–150 (IEEE Press, 2013).

Johnsen, E. G. and W. R. Corliss, *Human factors applications in teleoperator design and operation* (Wiley-Interscience New York, 1971).

Jokinen, K. and G. Wilcock, “Multimodal open-domain conversations with the nao robot”, in “Natural Interaction with Robots, Knowbots and Smartphones”, pp. 213–224 (Springer, 2014).

Kajita, S., M. Morisawa, K. Miura, S. Nakaoka, K. Harada, K. Kaneko, F. Kanehiro and K. Yokoi, “Biped walking stabilization based on linear inverted pendulum tracking”, in “Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on”, pp. 4489–4496 (IEEE, 2010).

Kanade, T., J. F. Cohn and Y. Tian, “Comprehensive database for facial expression analysis”, in “Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on”, pp. 46–53 (IEEE, 2000).

Knoll, A., “Editorial: Cui bono robo sapiens?”, Autonomous Robots **12**, 1, 5–12 (2002).

Kozima, H., C. Nakagawa, N. Kawai, D. Kosugi and Y. Yano, “A humanoid in company with children”, in “Humanoid Robots, 2004 4th IEEE/RAS International Conference on”, vol. 1, pp. 470–477 (IEEE, 2004).

Kube, C. R. and E. Bonabeau, “Cooperative transport by ants and robots”, Robotics and autonomous systems **30**, 1, 85–101 (2000).

Kuo, I.-H., C. Jayawardena, E. Broadbent, R. Q. Stafford and B. A. MacDonald, “Hri evaluation of a healthcare service robot”, in “International Conference on Social Robotics”, pp. 178–187 (Springer, 2012).

Le Maitre, J. and M. Chetouani, “Self-talk discrimination in human–robot interaction situations for supporting social awareness”, International Journal of Social Robotics **5**, 2, 277–289 (2013).

Lyons, M., S. Akamatsu, M. Kamachi and J. Gyoba, “Coding facial expressions with gabor wavelets”, in “Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on”, pp. 200–205 (IEEE, 1998).

Martinez, A. and R. Benavente, “The ar face database, 1998”, Computer Vision Center, Technical Report **3** (2007).

Mathur, M. B. and D. B. Reichling, “An uncanny game of trust: social trustworthiness of robots inferred from subtle anthropomorphic facial cues”, in “Proceedings of the 4th ACM/IEEE international conference on Human robot interaction”, pp. 313–314 (ACM, 2009).

Melhuish, C., O. Holland and S. Hoddell, “Collective sorting and segregation in robots with minimal sensing”, in “Proceedings of the fifth international conference on simulation of adaptive behavior on From animals to animats”, vol. 5, pp. 465–470 (1998).

Merritt, S. M., “Affective processes in human–automation interactions”, Human Factors: The Journal of the Human Factors and Ergonomics Society **53**, 4, 356–370 (2011).

Merritt, S. M. and D. R. Ilgen, “Not all trust is created equal: Dispositional and history-based trust in human-automation interactions”, Human Factors **50**, 2, 194–210 (2008).

MIT, “Face databases”, URL http://web.mit.edu/emeyers/www/face_databases.html (2003).

Miyakoshi, S., G. Taga, Y. Kuniyoshi and A. Nagakubo, “Three dimensional bipedal stepping motion using neural oscillators-towards humanoid motion in the real world”, in “Intelligent Robots and Systems, 1998. Proceedings., 1998 IEEE/RSJ International Conference on”, vol. 1, pp. 84–89 (IEEE, 1998).

Nakagawa, K., M. Shiomi, K. Shinozawa, R. Matsumura, H. Ishiguro and N. Hagita, “Effect of robot’s whispering behavior on people’s motivation”, International Journal of Social Robotics **5**, 1, 5–16 (2013).

Newman, S., A. Nassehi, X. Xu, R. Rosso, L. Wang, Y. Yusof, L. Ali, R. Liu, L. Zheng, S. Kumar *et al.*, “Strategic advantages of interoperability for global manufacturing using cnc technology”, Robotics and Computer-Integrated Manufacturing **24**, 6, 699–708 (2008).

Niculescu, A., B. van Dijk, A. Nijholt, H. Li and S. L. See, “Making social robots more attractive: the effects of voice pitch, humor and empathy”, International journal of social robotics **5**, 2, 171–191 (2013).

Ogura, Y., H. Aikawa, K. Shimomura, A. Morishima, H.-o. Lim and A. Takanishi, “Development of a new humanoid robot wabian-2”, in “Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on”, pp. 76–81 (IEEE, 2006).

Osawa, H., K. Ishii, S. Yamada and M. Imai, “Grounding cyber information in the physical world with attachable social cues”, in “Embedded and Real-Time Computing Systems and Applications (RTCSA), 2011 IEEE 17th International Conference on”, vol. 2, pp. 41–47 (IEEE, 2011).

Pantic, M. and L. J. Rothkrantz, “Facial action recognition for facial expression analysis from static face images”, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **34**, 3, 1449–1461 (2004).

Papageorgiou, C. P., M. Oren and T. Poggio, “A general framework for object detection”, in “Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)”, pp. 555–562 (1998).

Pearson, K., “On lines and planes of closest fit to systems of points in space”, Philosophical Magazine Series 6 **2**, 11, 559–572, URL http://dx.doi.org/10.1080/14786440109462720 (1901).

Phillips, R. and P. Madhavan, “The role of affective valence and task uncertainty in human-automation interaction”, in “Proceedings of the Human Factors and Ergonomics Society Annual Meeting”, vol. 57, pp. 354–358 (SAGE Publications Sage CA: Los Angeles, CA, 2013).

Pon-Barry, H., K. Schultz, E. O. Bratt, B. Clark and S. Peters, “Responding to student uncertainty in spoken tutorial dialogue systems”, International Journal of Artificial Intelligence in Education **16**, 2, 171–194 (2006).

Russell, J. A. and J. M. Fernández-Dols, *The psychology of facial expression* (Cambridge university press, 1997).

Salem, M., S. Kopp, I. Wachsmuth, K. Rohlfing and F. Joublin, "Generation and evaluation of communicative robot gesture", International Journal of Social Robotics **4**, 2, 201–217 (2012).

Schmidt, K. L. and J. F. Cohn, "Human facial expressions as adaptations: Evolutionary questions in facial expression research", American journal of physical anthropology **116**, S33, 3–24 (2001).

Sheikhi, S. and J.-M. Odobez, "Recognizing the visual focus of attention for human robot interaction", in "International Workshop on Human Behavior Understanding", pp. 99–112 (Springer, 2012).

Sinha, P., B. Balas, Y. Ostrovsky and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about", Proceedings of the IEEE **94**, 11, 1948–1962 (2006).

Solina, F., P. Peer, B. Batagelj, S. Juvan and J. Kovač, "Color-based face detection in the" 15 seconds of fame" art installation", (2003).

Tadesse, Y. and S. Priya, "Graphical facial expression analysis and design method: An approach to determine humanoid skin deformation", Journal of Mechanisms and Robotics **4**, 2, 021010 (2012).

Taheri, A., M. Alemi, A. Meghdari, H. Pouretemad and S. Holderread, "Clinical application of humanoid robots in playing imitation games for autistic children in iran", Procedia-Social and Behavioral Sciences **176**, 898–906 (2015).

Tourangeau, R. and T. Yan, "Sensitive questions in surveys.", Psychological bulletin **133**, 5, 859 (2007).

van Gent, P., "A tech blog about fun things with python and embedded electronics", URL `http://www.paulvangent.com/2016/04/01/emotion-recognition-with-python-opencv-and-a-face-dataset/` (2016).

Viola, P. and M. Jones, "Rapid object detection using a boosted cascade of simple features", in "Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001", vol. 1, pp. I–511–I–518 vol.1 (2001).

Vossen, S., J. Ham and C. Midden, "What makes social feedback from a robot work? disentangling the effect of speech, physical appearance and evaluation", in "International Conference on Persuasive Technology", pp. 52–57 (Springer, 2010).

Wagner, P., "Opencv, `http://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec_tutorial.html#duda01`", URL `http://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec_tutorial.html#duda01`, accessed: 2017-03-20 (2017).

Yamaguchi, J., E. Soga, S. Inoue and A. Takanishi, “Development of a bipedal humanoid robot-control method of whole body cooperative dynamic biped walking”, in “Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on”, vol. 1, pp. 368–374 (IEEE, 1999).

Yamasaki, F. and Y. Nakagawa, “Education using small humanoid robot”, in “Proceedings of the 3rd International Symposium on Autonomous Minirobots for Research and Edutainment (AMiRE 2005)”, pp. 248–253 (Springer, 2006).