RISE OF PERSONAL ROBOTS IN OUR LIVES
-The Robots Are About People

A **seminar report** submitted in partial fulfillment of
the requirements for the degree of

Bachelor of Engineering

by

Mayur N. Talole
(GR No. 111227)

Under the guidance of
Prof. S. M. Tayde

DEPARTMENT OF ELECTRONICS ENGINEERING
VISHWAKARMA INSTITUTE OF TECHNOLOGY PUNE
2013 - 2014

Bansilal Ramnath Agarwal Charitable Trust's

VISHWAKARMA INSTITUTE OF TECHNOLOGY, PUNE - 37

(An Autonomous Institute Affiliated to University of Pune)

CERTIFICATE

This is to certify that the **seminar report** entitled **RISE OF PERSONAL ROBOTS IN OUR LIVES -The Robots Are About People** has been submitted in the academic year **2013-14** by

Mayur N. Talole
(GR No. 111227)

under the supervision of **Prof. S. M. Tayde** in partial fulfillment of the requirements for the degree of Bachelor of Engineering in **Electronics Engineering** as prescribed by University of Pune.

Guide/Supervisor
Name:
Signature:

Head of the Department
Name: Prof. A. M. Chopde
Signature:

Internal Examiner
Name:
Signature:

Contents

List of Tables

List of Figures

Abstract

The adoption of personal service robots will depend on how well they interact with users. This report was motivated by a desire to facilitate the design of usable personal service robots. This report reviews the ongoing research about the designing of personal robots. Robots are just now becoming part of our everyday life and being used by ordinary people. The next generation social today activities. Future robots will work in hospitals, elder care centers, schools, and homes.

Personal robotics is a new and attractive use of robotic technologies. Imagine a day when you are talking to your grandma who lives at far distance and you will get to know whether she is in happy mood or not. If you get a sad face on phone you will try to cheer her out to make her happy. Same is possible with personal robotics. Also she can play with your children using a personal robot, while you are busy in work. Same can be implemented in hospitals for a doctor to operate on patient even when you are in airplane through a robot. Robots can even clean your room to make it tidy.

This report will discuss wide applications of personal robotics and basically how vision system for a robot is made. An important goal of this work is to use socially interactive robots as a scientific tool to understand human behaviour, to design machines that can engage us on social and emotional levels as well as learn from people, and to use these insights to create robotic technologies that can enhance human performance and quality of life with specific applications in healthcare, education, entertainment, and telecommunication.

Chapter 1

Introduction

A Personal robot is one whose human interface between machine design make it useful for all human beings. This is by contrast to industrial robots which are generally confined to some task and usually used by robotic specialists. A personal robot enables individual to automate the repetitive part of work of home or office making it more productive.

Robotic vacuum cleaners and floor-washing robots that clean floors with sweeping and wet mopping functions. Outdoor robots are domestic robots that perform different chores that exist outside of the house. Robotic lawn mowers are one type of outdoor robot that cut grass on their own without the need for a driver. There are also small humanoid remote controlled robots. Electronic pets, such as robotic dogs, can be companions for children. They have also have been used by many universities in competitions such as the RoboCup .There are also phone-powered robots for fun and games, such as Romo which is a small robot that employs smart phones as its brain. By using another mobile device and a cross-platform app, the user can drive it, make it produce animated facial expressions, direct it to dance, or turn it into a spybot. Also there are educational robots that will help teachers to teach students.

At present age of robotics, there is evolution of special type of robots-Humanoids. These are the robots which have shape in resemble with human body. Researchers need to understand the human body structure and behaviour (biomechanics) to build and study humanoid robots. So this report is going to expose the actions of hardware and coding algorithms on which these robots work.

Chapter 2

What is PERSONAL ROBOTICS?

Romotive Nerds get this question a lot, so I would like to take this opportunity to define Personal Robotics. We all know Rosie from the Jetsons or Roomba from iRobot; both serve as classic examples of personal robots. They exemplify the application of robotics to everyday life for us as individuals, and thats how we define personal robotics.Now, improving everyday life can cover a range of tasks from coffee retrieval to companionship, which is why robotics can involve a wide range of subjects. Roboticists cant just think about gears and wires. They also have to consider the world and people with whom these robots will interact.

While robots currently fulfil many industrial and research needs, theyre still a rare sight in our day to day living. Society once saw computers in the same way as machines used strictly in industrial and academic settings. Personal robotics is the dream that robots will make the same leap into everyday life and provide value for everyone. The field of personal robotics provides unique insights into human psychology. We are interested in how robots are influenced by behaviour selection, regulate social interactions, and promote learning in a social context.

It is the study of the mechanisms that enable an individual to acquire information or skills from another individual has been a seminal topic in many areas of cognitive science. Traditionally, autonomous robots are designed to operate as independently and remotely as possible from humans, often performing tasks in hazardous and hostile environments . However, a new range of application domains (domestic, entertainment, health care, etc.) are driving the development of robots that can interact and cooperate with people as a partner, rather than as a tool.

Chapter 3

Block Diagram of Personal Robots

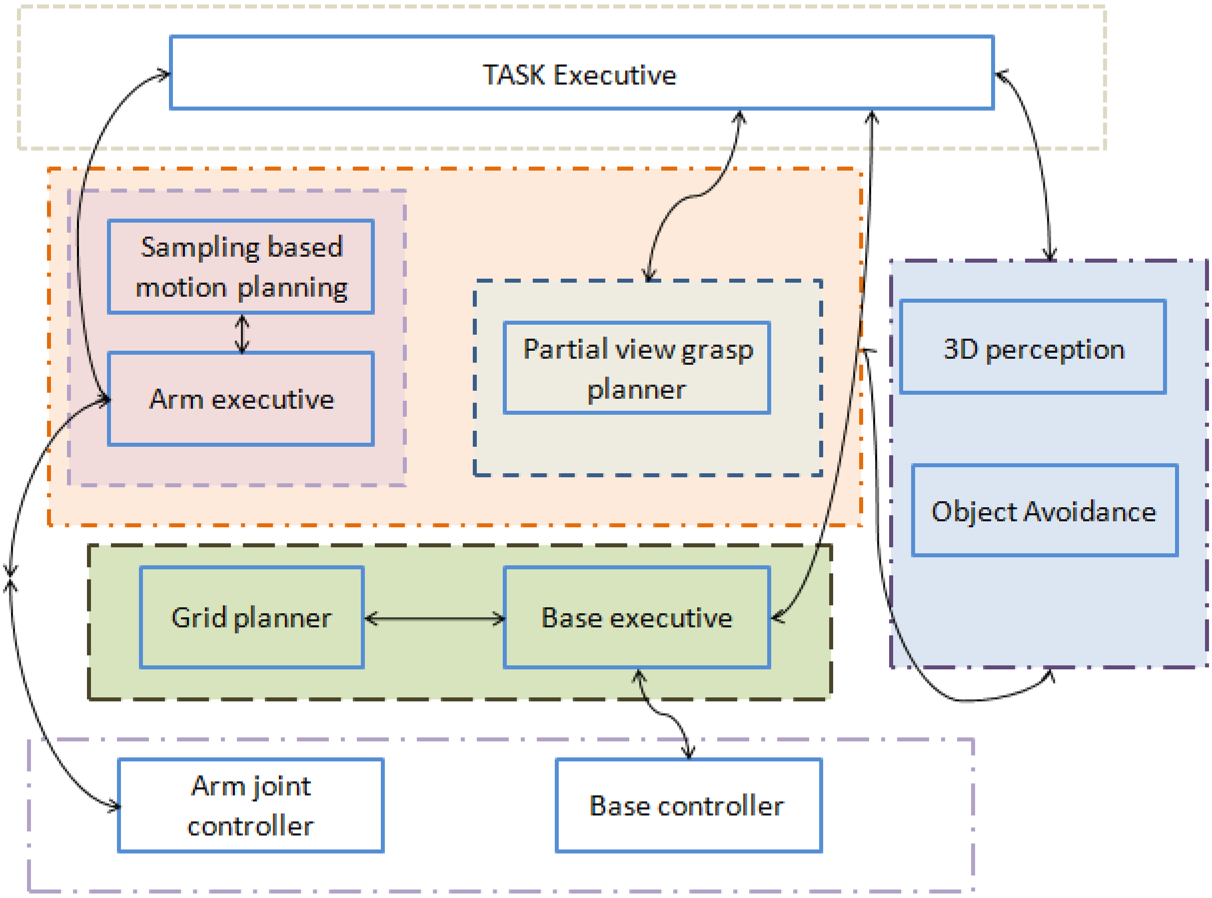

Figure 3.1: General block diagram of personal robot

It is necessary to provide robots with safe and goal-directed real-time manipulation behaviour. The safe and goal-directed real-time manipulation behaviour. The personal robots should be capable of performing well while using realistic sensors and making few assumptions about the environment setup. The personal robotic system should produce a system that is safe and robust, can run autonomously; can handle failures and situations it has not been presented before [16]. Reliability and exibility is not only achieved through the capabilities of the individual components (perception, planning, etc.) but also through the task management that synchronizes and parameterizes the activities in dynamic, context-specic ways and detects and locally recovers from failures. That is why the personal robotic system should as the general block diagram in figure.

3.1 3D perception

This system will take data from sensors such as kinect, ultrasonic distance sensor and 3D laser scanner sensor, based on the algorithm which is coded in DSP processor the system will generate a dynamic obstacle map. This map is used by system to guide the robot to destination.

3.2 Arm executive

The arms of robot will be useful to grasp any object. Generally this system will be consists of motors, encoders circuitries. The motor will be useful for actuation of mechanisms but the movement of the mechanism will be dependent on encoder output. The encoders will control motor movement upto desired angle like in servo motors.

3.3 sampling based motion planner

This block takes the generated map provided by 3D perception system for collision detection. To perform this task the leading-edge motion planning algorithms and tools for optimizing and smoothing motion trajectories.

3.4 Base executive

This block contains all actuation circuitries which will drive motors, also with the feedback for sensors and inputs from processor. This block will be responsible for not allowing excessive current to motors on sensors circuitries which will protect them from getting damaged.

3.5 Task executive

This is a task layer which is manipulation system which solves many complex components such as 3D perception, motion planning and more. All software components necessary to perform the tasks discussed in this work are modular so they can be reused individually. To be capable, personal robots need a few core technologies to function and engage within a human environment, including: sensing, moving, actuation and intelligence. Visual perception requires high bandwidth and is computationally demanding. In the early stages of human vision, the entire visual eld is processed in parallel. This will model the ambient data received from sensors to implement them as human vision system. A system that integrates action, perception, attention, and other cognitive capabilities can be more exible and reliable than a system that focuses on only one of these aspects.

Chapter 4

Present day personal robots

The first home robot was Roomba is an automated vacuum cleaning robot first released in 2002. Roomba is powered by a rechargeable battery, and many models are available with a docking station to which the Roomba should return to recharge at the end of its cleaning cycle. Scooba is iRobot's floor washing robot. The product became commercially available in limited quantities in late 2005 before a full product release in 2006, although it still is not available in many overseas markets.

Irobot has an extensive line of robots designed for the use in military such as PackBOT, SUGV, Warrior, R-GATOR,, Negotiator, Transphibian, Chembot,Alram, Ember. Negotiator is a man-portable civil-response surveillance and reconnaissance robot. Transphibian is a man-portable UUV and bottom crawler that autonomously inserts itself into the water and operates in a shallow area. It is designed for mine detection, harbor defense and surveillance. Ember is a prototype miniature, tracked robot, weighing around 1 lb and costing so little to make that it is intended to be virtually disposable [5]. Ember moves at walking pace, can right itself when it is turned over and is controlled by a simple touchscreen application on an Apple iPhone. Also personal robots in medical field are developed RP-VITA, or Remote Presence Virtual + Independent Telemedicine Assistant, is a medical robot jointly produced with In Touch Health. The robot will be cloud-connected and have access to a patient's medical record, and will also be able to plug in diagnostic devices such as stethoscopes, otoscopes, and ultrasound.

The most famous personal robots are of Massachusetts Institute of Technology personal robotics group which are stated as follows.

4.1 MeBot

The MeBot is semiautomatic robotics avatar that gives a person the richer way of interacting with audience than is allowed with phone and video conferencing. The robot is able to convey the non-herbal channels of social communication. It is able to communicate with anybody posture, a wide range of head movement and very expressive hand gesture. It takes the advantage of the current advanced technology in wireless communication and ever expanding capabilities of mobile devices [20]. MeBot is push toward the future where remote presence can achieved easily that will save travelling time but still achieves the same experience of being there.

Figure 4.1: MeBot at MIT Media Lab

4.2 Kismet

The robot with the most sophisticated visual-motor behaviour is Kismet. This robot is an active vision head augmented with expressive facial features. Kismet is designed to receive and send human-like expression behaviours. Kismet has three degrees of freedom to control gaze direction, three degrees of freedom to control its neck, and fteen degrees of freedom in other expressive components of the face (such as ears and eyelids). The cameras in Kismets eyes have high acuity but a narrow eld of view[3]. Between the eyes, there are two central cameras for wide view of environment. A wide eld of view is needed for search tasks, for tracking multiple objects. Visual behaviour can be conceptualized on four different levels. These levels correspond to the social level, the behaviour level, the skills level, and the primitives level. This decomposition is motivated by distinct temporal, perceptual, and interaction constraints that exist at each level.

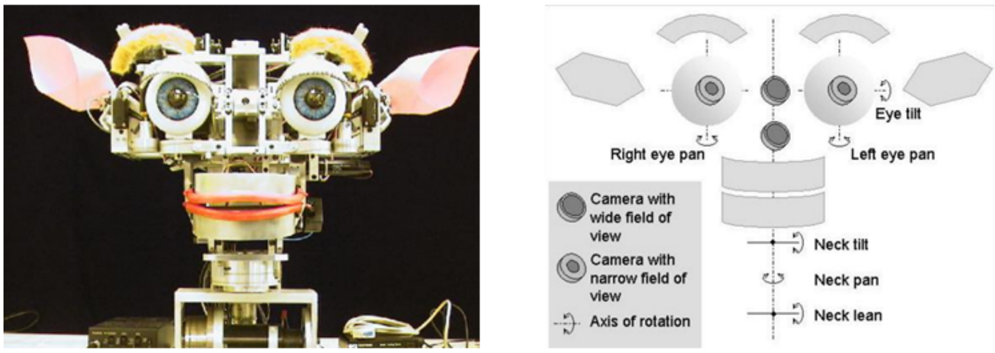

Figure 4.2: Kismet, the sociable robot.

- The Social Level: The social level explicitly deals with issues pertaining to having a human in the interaction loop. This requires careful consideration of how the human interprets and responds to the robots behaviour in a social context[4]. Using visual behaviour to help regulate the transition of speaker turns during vocal turn-taking is an example.

- The behaviour level: The behaviour level deals with issues related to producing relevant, appropriately persistent, and opportunistic behaviour. Actively seeking out a desired stimulus and then visually engaging it is an example.

- The motor skill level: The motor skill level is responsible for guring out how to move the motors to accomplish that task[1]. The skills level must also deal with coordinating multi-modal motor skills (e.g., those motor skills that combine speech, facial expression, and body posture).

- The motor primitives level: The motor primitives level implements the building blocks of motor action. Kismet actually has four distinct motor systems at the primitives level: the affective vocal system, the facial expression system, the oculomotor system, and the body posturing system.

There are some expressions done by Kismet such as sad, calm, anger, sleep, fear as shown below[4].

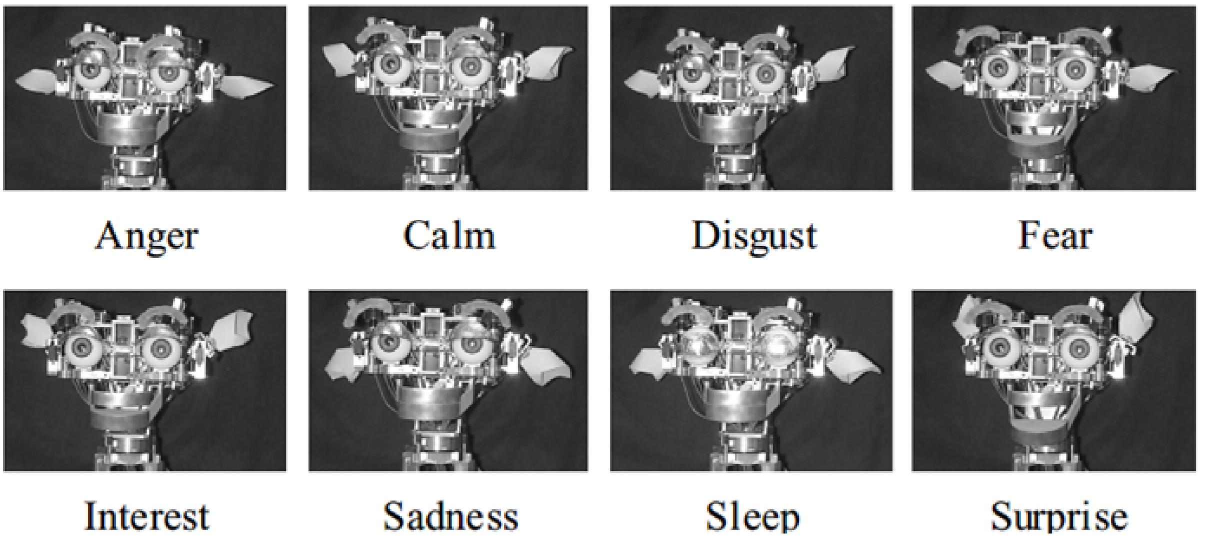

Figure 4.3: Static extremes of Kismet s facial expressions. During operation, the 11 degrees-of freedom for the ears, eyebrows, mouth, and eyelids vary continuously with the current emotional state of the robot..

4.3 AIDA

AIDA is part of the Sociable Car - Senseable Cities project which is collaboration between the Personal Robots Group and the Senseable Cities Group at MIT[20]. The aim of the Personal Robot's portion of the project is to expand the relationship between the car and the driver with the goal of making the driving experience more effective, safer, and more enjoyable. This channel would be modelled on fundamental aspects of human social interaction including the ability to express and perceive affective/emotional state and key social behaviours. This interface is a research platform, which can be used as a tool for evaluating various topics in the area of social human-automobile interaction.

Currently the AIDA research platform consists of a fully functional robotic prototype embedded in a stand-alone automobile dash[20]. The robot has a video camera for face and emotion recognition, touch sensing, and an embedded laser projector inside of the head. Currently a driving simulator is being developed around the AIDA research platform in order to explore this new field of social human-automobile interaction. Its intention is that a future version of the robot based on the current research will be installed into a functioning test vehicle.

Figure 4.4: AIDA system attatched in vehicle

4.4 Leonardo

MIT have christened this new character collaboration with a name that embodies art, science and invention. Hence, the name ”Leonardo” – namesake of Leonardo DaVinci, the Renaissance scientist, inventor and artist. Leonardo has 69 degrees of freedom — 32 of those are in the face alone. As a result, Leonardo is capable of near-human facial expression.

4.4.1 Construction

Unlike the vast majority of autonomous robots today, Leonardo has an organic appearance. It is a fanciful creature, clearly not trying to mimic any living creature today. These controllers support simultaneous absolute position and velocity feedback, allowing good dynamic performance without the need for lenghthy calibration phase at power-up. The motor drivers are standard FET H-bridges; recent advances in FET process technology permit surprisingly low RDS on losses, and switching at relatively low (1-10kHz) frequencies reduces switching losses. Hence, the power silicon (and thus the package as a whole) can be reduced in size. The sixteen channels each support current feedback, encoder feedback, and analog feedback, and the system is controlled by a custom SoC motion controller with an embedded soft processor core implemented in a Xilinx Virtex FPGA.

Figure 4.5: the outer body of Leonardo

4.4.2 Learning faces

MIT Media Lab have developed a real-time face recognition system for Leonardo that can be trained on the fly via a simple social interaction with the robot[20]. The interaction allows people to introduce themselves and others to Leonardo, who tries to memorize their faces for use in subsequent interactions. In order to learn new faces, Leo keeps a buffer of up to 200 temporally-contiguous or near-contiguous views of the currently tracked face. This buffer is used to create a new face model whenever the person introduces themselves via speech. When a new model is created, principal component analysis (PCA) is performed on the entire face image data set, and a spline manifold is fitted to the images of the new face [14]. Since the face recognition module runs as a separate process from Leo's other cognitive modules, the addition of a new face model can be done without stalling the robot or the interaction [12]. The face recognition module receives speech input provided by the Sphinx-4 speech recognition system. The speech recognition system allows people to introduce themselves via simple phrases: ”My name is Marc” or ”Leo, this is Dan.”

4.5 Social learning

Personal robots must be able to learn new skills and tasks while on the job from ordinary people. Human-teachable robots must be able to move flexibly and opportunistically along the GUIDANCE/EXPLORATINON spectrum. Along the GUIDANCE dimension, many prior works require a constant (often high) involvement of a human teacher for the robot to learn anything.

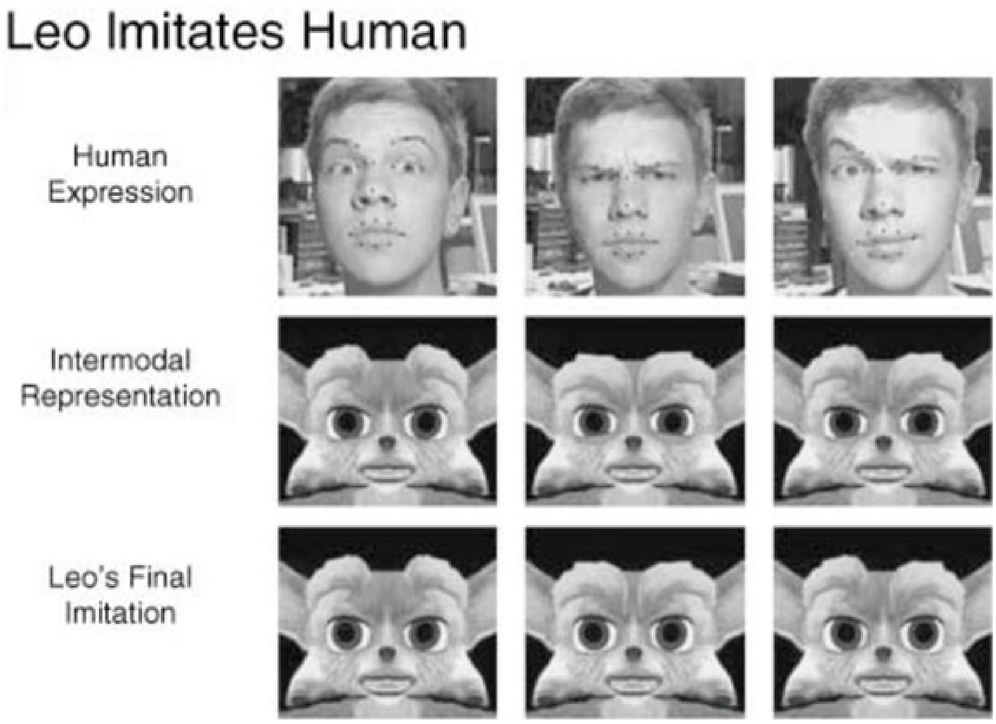

Figure 4.6: Leonardo can imitate facial configurations that involve combining intermodal representations for different regions of the face. By searching each of its motor systems (left eye region, right eye region, and mouth) for the closest match in the overall pose.

Perceiving similarities between self and other is an important part of the ability to take the role or perspective of another, allowing people to relate to and to empathize with their social partners[10]. This sort of perspective shift may help us to predict and explain other's emotions, behaviours and other mental states, and to formulate appropriate responses based on this social understanding. To imitate people, the robot needs to convert the 3D joint angle data it perceives about the human.

Chapter 5

Basic functionalities in software required to build personal robot

5.1 Real time Face detection and tracking

Face detection and tracking is the process of determining whether or not a face is present in an image. Unlike face recognitionwhich distinguishes different human faces, face detection only indicates whether or not a face is present in an image. In addition, face tracking determines the exact location of the face.

5.1.1 Algorithm

Color segmentation has been proved to be an effective method to detect face regions due to its low computational requirements and ease of implementation. Compared to the featured based method, the color-based algorithm required very little training.First, the original image was converted to a different color space, namely modified YUV. Then the skin pixels were segmented based on the appropriate U range. Morphological filtering was applied to reduce false positives[11].

5.1.2 Modified YUV Color Space

Converting the skin pixel information to the modified YUV color space would be more advantageous since human skin tones tend to fall within a certain range of chrominance values (i.e. U-V component), regardless of the skin type.

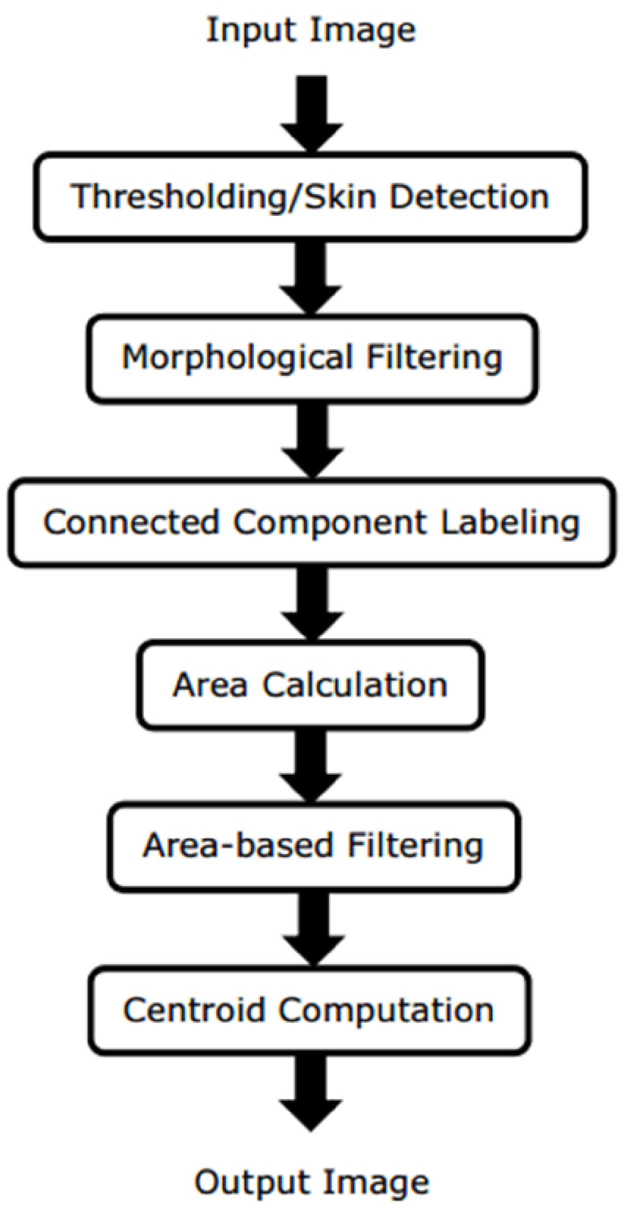

Figure 5.1: Real Time Face Detection Algorithm

5.1.3 Thresholding /Skin Detection

After skin pixels were converted to the modified YUV space, the skin pixels can be segmented based on the following experimented threshold.

$$10 < U < 74 \tag{5.1}$$

$$-40 < V < 11 \tag{5.2}$$

Realistically, there are so many other objects that have color similar to the skin color. Applying morphological filteringincluding erosion and hole filling would, firstly, reduce the background noise and, secondly, fill in missing pixels of the detected face regions, MATLAB provided built-in functionsimerode and imfill for these two operations.

5.1.4 Connected Component Labeling and Area Calculation

After each group of detected pixels became one connected region, connected component labeling algorithm was applied. This process labeled each connected region with a number, allowing us to distinguish between different detected regions. The built-in function bwlabel for this operation was available in MATLAB. In general, there are two main methods to label connected regions in a binary imageknown as recursive and sequential algorithms.

Figure 5.2: MATLAB implementation :Result after thresholding and Connected Component Labeling

5.1.5 Area-based Filtering

Note that morphological filtering only removed some background noise, but not all. Filtering detected regions based on their areas would successfully remove all background noise and any skin region that was not likely to be a face[11]. This was done based on the assumption that human faces are of similar size and have largest area compared to other skin regions, especially the hands. Therefore, to be considered a face region, a connected group of skin pixels need to have an area of at least 26 percentage of the largest area. This number was obtained from experiments on training images. Therefore, many regions of false positives could be removed in this stage.

Figure 5.3: MATLAB implementation :Result after area-based filtering

5.1.6 Centroid Computation

The final stage was to determine face location. The centroid of each connected labeled face region can be calculated by averaging the sum of X coordinates and Y coordinates separately. Centroid is shown as blue stars.

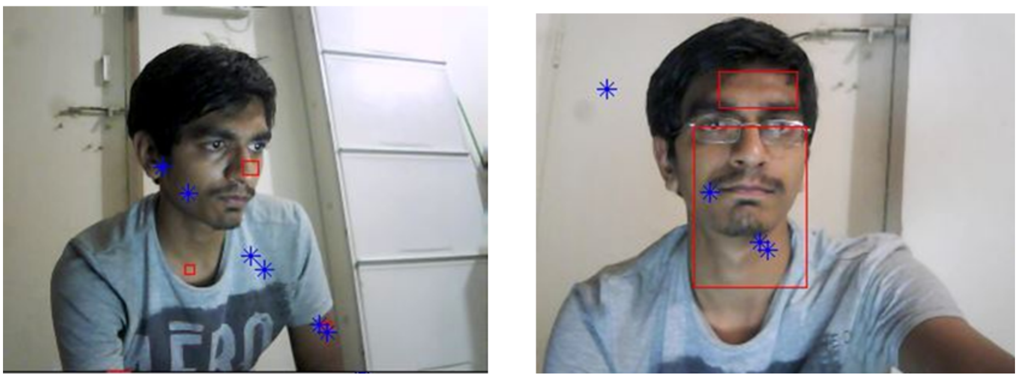

Figure 5.4: MATLAB implementation of Real time detection and tracking of face

5.2 speech recognition

It is because of large number of accents spoken around the world that this conundrum still remains an active area of research. Speech Recognition finds numerous applications including health care, artificial intelligence, human computer interaction, Interactive Voice Response Systems, military, avionics etc. Real Time Speech Recognition takes as an input the time domain signal from a microphone and performs the frequency domain feature extraction on the sample to identify the word being spoken.

5.2.1 Principle

Speech recognition systems can be classified into several models by describing the types of utterances to be recognized. These classes shall take into consideration the ability to determine the instance when the speaker starts and finishes the utterance. In our project we aimed to implement Isolated Word Recognition System which usually used a rectangular window over the word being spoken. The Speech Recognition Engines are generally classified into 2 types, namely Pattern Recognition and Acoustic Phonetic systems. While the former use the known/trained patterns to determine a match, the latter uses attributes of the human body to compare speech features[18].

5.2.2 Data acquisition and detection

The speech signal is essentially analog in nature. Hence, the signals must be converted to digital data in order to be read and processed. The very next issue to be addressed was to detect whether the word was being spoken. So for this the FFT Module is used which output the source signal which is a signed block exponent which accounts for the internal signal values during FFT computation.

5.2.3 Feature Extraction and Transformations

Mel Frequency Cepstrum

Mel Frequency Analysis of speech is based on perception in which the filter bank is used to concentrate only on certain perceptible frequency ranges. Mel-Frequency Cepstrum (MFC) is a representation of the short-term power spectrum of a speech, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency[18].

The triangular mel-filters in the filter bank are placed in the frequency domain so that each filter's center frequency follows the mel scale so that each filter bank represents different perceptual effect for different frequency bands. MFCCs combine consideration of aspects of human hearing (logarithmic frequency perception, the mel scale).

Performing Discrete Cosine Transform on Mel Cepstrum

A DCT computes a sequence of data points in terms of summation of cosine functions oscillating at various frequencies [18]. The idea of performing DCT on Mel Scale is rather difficult to imagine on what the transformations might look like physically. After obtaining a DCT for each chunk of 32 ms of the speech signal, an algorithm to find the shortest distance match for the sequence of these 32 blocks is to be done.

5.2.4 Comparison of Words Yes and No

The following images were derived using MATLAB for the words Yes and No to efficiently make out both time domain and frequency domain differences between the two spoken words:

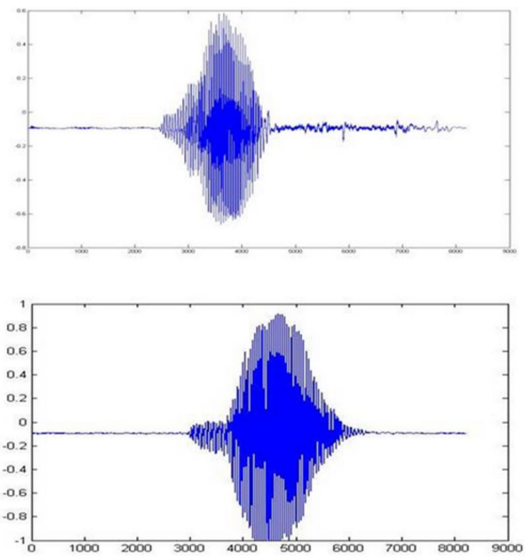

Figure 5.5: MATLAB implementation Time Domain Representation:For words Yes and 'No'

5.2.5 Facial expression recognition

An efficient, local image-based approach for extraction of intransient facial features and recognition of four facial expressions from 2D image sequences can be used. The algorithm uses edge projection analysis for feature extraction and creates a dynamic spatio-temporal representation of the face. Firstly it is necessary to detect mouth to get lip enhancement transform [8]. In case of colored images, lip pixels significantly differ from those of skin in YCbCr color space. Edges for lips occur both in horizontal and vertical direction. In the bounding box computed by the generic algorithm, closed contours are obtained by applying Laplacian of Gaussian operator at zero thresholds.

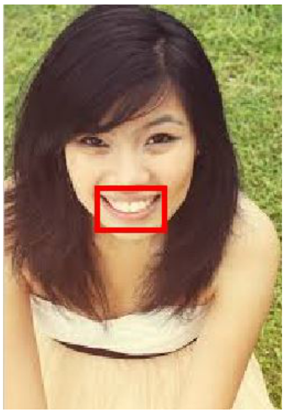

Figure 5.6: MATLAB implementation for detection of mouth and then selected part is processed for detection of facial expression

5.2.6 LIP ENHANCEMENT TRANSFORM

Sadeghi proposed a lip segmentation method based on Gaussian mixture modeling of mouth area images, followed by Bayesian decision-making system, which labels the pixels as lip or non-lip. This approach gives a binary image, on which usual feature extraction algorithms cannot be applied. Approach by Lievin using Bayesian segmentation also results in a binarized image. To follow the usual feature extraction algorithms, the method would be to first convert the vector (color) image to a scalar (grayscale) image by

$$S_1 = a_1R + a_2G + a_3B \tag{5.3}$$

Where (R,G,B) are the components in the RGB color space. This is followed by other operations like calculating edge maps to find the lip region in a color image. This approach would use intensity information only. This compromises the accuracy of lip segmentation, because intensity difference need not be significant between the lip and skin pixels, and lightening variations may also be present.

5.2.7 FEATURE VECTOR AND CLASSIFICATION

A spatio-temporal representation of the face is created, based on geometrical relationships between features using Euclidean distance. Such a representation allows robust handling of partial occlusion. Seven parameters form the feature vector F

$$F = \{H_e, w_e, H_m, W_m, R_{ul}, R_{ll}, N_L\} \quad (5.4)$$

All components of the vector are normalized against the first frame to achieve scale independence. Radii of curvature of the upper and lower lips Rul and Rll are computed by approximating the binary mask of the lips with two parabolas [8]. NL is the number of distinct peaks detected for upper and lower lips during edge projection analysis, indicating whether mouth was open or closed. Such dynamic characteristic of the feature vector provides shape independence.

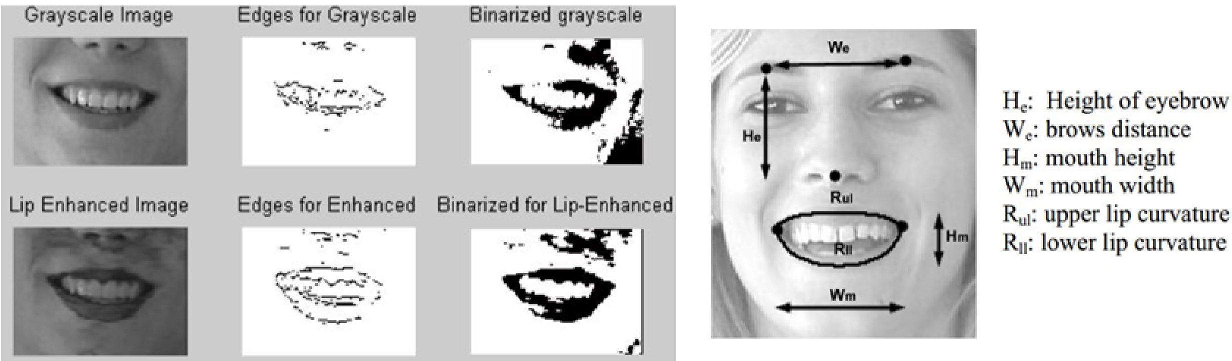

Figure 5.7: . First Row: (A) Color image converted to grayscale,(B) Edges for A, (C) Binary image obtained by thresholding A.Second Row: (A) Transformed color image, (B) Edges for A,(C) Binary image obtained by thresholding A. Notice the improvement in the second row, in terms of edges for lips, and binarized image.

5.2.8 Feature Extraction

Each grayscale image sequence in the database depicted one of the expression classes (smile, surprise, sad and disgust against neutral). The first image in the sequence was a neutral image. Confidence level of each expression was calculated for each of the subsequent images against the neutral image. Expression having the highest total confidence level was declared as the expression of the sequence.

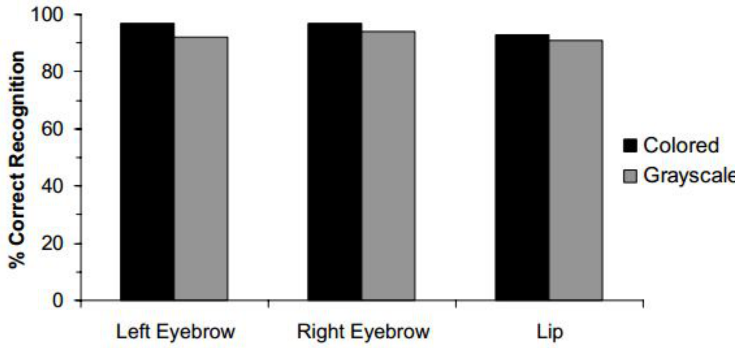

Figure 5.8: Accuracy for facial feature extraction

5.3 Gesture detection using RGB color space

Edge detection process has a widespread usage in computer vision applications. But it has a different output when its input image changes from color to grayscale. Edges are sudden variation in the gray level or color of image pixels, edge detection is an important operation for reducing processed data with preserving useful and informative object boundaries [13]. xThe step of algorithm that was proposed by Dutta and Chaudhuri calculates a directional color difference. This step applied a transformation on the RGB pixel values to make it one value instead of three to reduce the computational overhead of color vector.

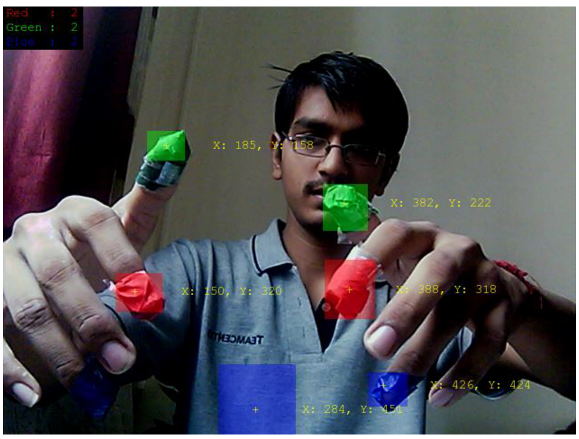

Figure 5.9: MATLAB implementation of real time RGB coordinate detection

5.4 Pose estimation

Human pose estimation from a single viewpoint is a challenging and important problem in computer vision. As cameras become standard computer peripherals there are many possible applications for looking at people. There has been substantial progress on estimating human pose from a single viewpoint, however the problem remains quite difficult. Many recent approaches are based on using a tree-structured model that captures the kinematic relations between parts such as the torso and limbs[19]. In time linear in the number of body parts, while capturing what is arguably the most important source of constraint on human body pose, the joints connecting the limbs.

The key idea is to introduce a small number of latent variables to represent residual correlations between parts that are not captured by a tree model. This kind of approach has recently been investigated in more general graphical models. In a bit more detail, we start with a kinematic tree model and identify parts where correlation in their locations violates the conditional independence assumption of the tree model. We then use factor analysis to nd the best common factor that accounts for these correlations [15]. This common factor is added to the tree model as a latent variable. The resulting common factor model preserves the underlying tree structure, which allows a variant of the Viterbi algorithm to be used for efficient pose estimation [10].

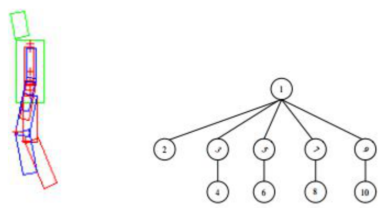

Figure 5.10: pose estimation of one side of body

However this model of a person also illustrates some limitations of tree-structured models. The locations of sibling parts are independent when conditioned on their parent. For instance, given a location for the torso, the locations of the upper arms and legs are independent.

Chapter 6

Robots of future

A generation of robots, like the da Vinci surgical robot, has given surgeons the dexterity and mobility to operate in the confined and claustrophobic interior of the body. But combined with advances in imaging, the technology allows surgeons to see through the body like never before. One of the ways to combine humans and robots is an experimental technique called " perceptual docking", where the eye movements of the surgeon are tracked in order to teach the robot the cognitive processes and decision-making paths involved in surgery.

Already under development, for example, is a military humanoid robot that will fight fires on ships. The innocuously named Shipboard Autonomous Firefighting Robot, or Saffir, actually looks suspiciously much like a skinless version of the Terminator, though its mission is much more peaceful.

Saffir, which is being developed by university researchers in cooperation with US navy scientists, is designed to fight fires on naval ships. But even these firefighting robots are being developed specifically to work with human counterparts, so a major focus of the work is on getting the robots to respond to human gestures and speech.

Chapter 7

Conclusions

It can be concluded that robot assisted daily work will be very precise and easy.In order to increase the relationship between humans and robots we should look forward for making of human like robots.With robots guided surgery we can achieve maximum accuracy and also ,success.

This report will help one to design a basic personal robot. Some of the algorithms specified in this report are implemented using MATLAB.We can program the DSP processor using the same algorithm.The processor can be fitted in robotic system. Using sufficient current requirement we can drive motor upto some angle using proper sensors.

”You have a human, which is pretty good in terms of decision-making, and learning. You have a robot, which is good at doing precise movements. Why not use a combination of both?”

Bibliography

[1] C. Breazeal and B. Scassellati (2002). ”Robots that imitate humans,” Trends in Cognitive Science, 6, pp. 481-487.

[2] C. Breazeal (2003), Towards sociable robots, T. Fong (ed), Robotics and Autonomous Systems,vol. 42(3-4), pp. 167-175.

[3] Deep Learning for Detecting Robotic Grasps, Ian Lenz, Honglak Lee, Ashutosh Saxena. Robotics: Science and Systems (RSS), 2013.

[4] Dutta, S., Chaudhuri, B.: A Color Edge Detection Algorithm in RGB Color Space. In: Proc. International Conference on Advances in Recent Technologies in Communication and Computing. pp. 337 - 340, (2009).

[5] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer,Optimizing binary mrfs via extended roof duality, in CVPR,2007.

[6] C. Breazeal and B. Scassellati (2000), Infant-like social interactions between a robot and a human caregiver, in K. Dautenhahn (ed), Adaptive Behavior, 8:1, 4772.

[7] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in CVPR, 2005.

[8] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray,Visual categorization with bags of keypoints, in Workshop on statistical learning in computer vision, ECCV, 2004.

[9] A. Torralba, Contextual priming for object detection, IJCV,vol. 53, no. 2, pp. 169191, 2003.

[10]] C. Li, A. Kowdle, A. Saxena, and T. Chen, Towards holistic scene understanding: Feedback enabled cascaded classication models, in NIPS, 2010.

[11] Real time face detection using FPGA arrays.

[12] Hallucinated Humans as the Hidden Context for Labeling 3D Scenes, Yun Jiang, Hema S. Koppula, Ashutosh Saxena. In Computer Vision and Pattern Recognition (CVPR), 2013

[13] Y. Jiang, S. Moseson, and A. Saxena. Efcient grasping from RGBD images: Learning using a new rectangle representation. In ICRA, 2011.

[14] Learning to Place New Objects in a Scene, Yun Jiang, Marcus Lim, Changxi Zheng, Ashutosh Saxena. In International Journal of Robotics Research (IJRR), 2012.

[15] S. Coradeschi, H. Ishiguro, M. Asada, S. Shapiro, M. Theilscher, C. Breazeal, M. Mataric, H. Ishida (2006) Human-Inspired Robots. IEEE Intelligent Systems21(4), 74-85.

[16] C. Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati (2001). ”Active vision systems for sociable robots,” in K. Dautenhahn (ed.), IEEE Transactions on Systems, Man, and Cybernetics, Part A, 31:5, pp. 443-453.

[17] Efficient Grasping from RGBD images: Learning using a new Rectangle Representation, Yun Jiang, Stephen Moseson, Ashutosh Saxena. In International Conference on Robotics and Automation (ICRA), 2011.

[18] Speech Algorithms On Wikipedia.

[19] Pose estimation from a single depth image for arbitrary kinematic skeletons, Daniel Ly, Ashutosh Saxena, Hod Lipson. In R:SS workshop on RGB-D cameras, 2011.

[20] MIT Media Lab Personal robotics group.

[21] A. Saxena, L. L. S. Wong, and A. Y. Ng. Learning grasp strategies with partial shape information. In AAAI, 2008.