University POLITEHNICA of Bucharest

Faculty of Automatic Control and Computers

Computer Science and Engineering Department

DOCTORAL THESIS

Contributions to the human body analysis from images

by

ALABBASI Hesham Adnan Abdelameer

Supervised by

Prof. dr. ing. Florica Moldoveanu

Bucharest, January 2016

UNIVERSITATEA POLITEHNICA DIN BUCURȘTI

Facultatea de Automatică și Calculatoare

Departamentul de Calculatoare

TEZĂ DE DOCTORAT

Contribuții la analiza corpului uman din imagini

Autor: **ALABBASI Hesham Adnan Abdelameer**

Conducător de doctorat: **Prof. dr. ing. Florica Moldoveanu**

Preşedinte	Prof. dr. ing. Adina Magda Florea	de la	Universitatea POLITEHNICA din Bucureşti
Conduc ător de doctorat	Prof. dr. ing. Florica Moldoveanu	de la	Universitatea POLITEHNICA din Bucureşti
Referent	Prof. dr. ing. Vasile Manta	de la	UniversitateaTehnică ”Gheorghe Asachi” din Iaşi
Referent	Prof. dr. ing. Ştefan-Gheorghe Pentiuc	de la	Universitatea ”Ştefan cel Mare” din Suceava
Referent	Conferenţiar dr. ing. Irina Mocanu	de la	Universitatea POLITEHNICA din Bucureşti

Bucureşti, Ianuarie 2016

ACKNOWLEDGEMENTS

Firstly, great thanksgiving and praise to God for giving me strength, patience and reconcile to complete my Ph.D. study. All the time I feel that God supporting me.

I would like to express my sincere gratitude to my supervisor Prof. Florica Moldoveanu for the continuous support of my Ph.D. study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor.

My sincere thanks also go to Prof. Alin Moldoveanu for the assistance given by him, I would also like to thank Dr. Victor Asavei, Dr. Anca Morar and Alexandru Grădinaru, also all the other members in the Faculty of Automatic Control and Computers, Computer Science Department during my time at the University POLITEHNICA of Bucharest.

I would like to express how grateful I am for the infinite support and love I have received throughout these years in the Ph.D. program, without which, I would not have been able to reach this stage successfully.

Many thanks go to my colleagues, whose support, friendship, and encouragement made the Ph.D. journey an enjoyable and unforgettable experience for me.

I would also like to thank all my close friends and brothers in my second country Romania (Bucuresti) who supported me, and invited me to strive towards my goal.

Last, but not least, I would like to thank my amazing family. My mom, wife, sons, and daughter, for their unconditional love and support. Words cannot express how grateful I am to my mother; your prayer for me has what sustained me thus far. My wife for all of the sacrifices, I am grateful for having them in my life and I thank God for this blessing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1. MOTIVATION

The human body is the most complex mechanism in the world. Like all other living organisms, the human body is composed of millions of micro cells that, combined, form tissues and functional complex body parts like arms, legs, torso, neck or the head that works as a very powerful processor.

The complexity of the human body allows us to perform many physical actions, communicate and express, build great societies, innovate and survive alongside other living beings. Still, we still know too little about it.

Our body, usually allows us to communicate verbally with others, but very often words are not enough so we have to gesture and express our feelings, thoughts and intentions in various ways using our body. This is called body language and we can be aware of it and communicate like this in a conscious way, but very often our body has many unconscious reactions that can be interpreted. Many of them are very important when it comes to psychology or medicine, but also in social interactions (twitches but or only).

It is very important to know about the movements of the human body in everyday and the technology nowadays have greatly improved in this area of interest where a wide range of applications are being extensively developed: biomechanics, medicine, sign language, game, movies and so on. Still, many more practical approaches and experiments are required to further understand this highly complex mechanism known as the human body.

A very complex part of the human body is the head where all the main sensory organs and also, a less studied part - the face, are located. Then face has probably the most important role in the body language when it comes to emotions. Facial expressions are the most natural way for humans to express their feeling in certain situations (hence the widespread emoticons or smileys that are represented as faces or facial expressions).

All of this body language can nowadays be captured and interpreted using current technology. It can, of course, detect if a person is happy, but also, more important, when a person is sad or sick. The automatic recognition of basic facial expressions in different contexts can lead to a much better psychological and medical diagnosis and can contribute to the general human welfare.

The medical field could also benefit from body tracking by easily recording and interpret human movements for impaired persons, for a better understanding of each patient's condition and even for the rehabilitations.

1.2. GOALS OF THE RESEARCH

The main goal of this thesis is to investigate the effectiveness of using the last product of the Microsoft Kinect V2 Sensor to build some applications for monitoring and analyzing the human face and body.

We have implemented and described in this thesis four systems.

The first system is capable of detecting the face in real-time with different depths in the image in both indoor and outdoor environments. The main contribution related to this system is the use of skin color as a basic feature for detection, by a combination between the result of segmentation in RGB and HSV color spaces, the result of segmentation using the Elliptical model in YCbCr color space and edge information.

The second system can recognize facial emotion expressions in real time, we focus on the emotion recognition from facial expressions by using Microsoft Kinect V2 sensor with its face tracking SDK to recognize eight expressions. The implementation of our emotion recognition application was made with Visual Studio 2013 (C++) and Matlab 2014.

The third system, has extended the second system to connect the emotional expressions with brain activity, so this system is capable of detecting the facial expression and which part of the brain is associated with each emotion.

Finally, the fourth is a system for motion tracking and evaluation, a virtual sports training system by using the same Kinect. This system can help trainee in different kinds of exercises, or other persons for rehabilitation.

1.3. SCIENTIFIC PUBLICATION IN CONNECTION WITH THIS THESIS

1. Hesham. ALABBASI, F. Moldoveanu," Human face detection from images based on skin color", Proceedings of the 18th International Conference on System Theory, Control and Computing, Sinaia, Romania, October 17-19, 2014, pp 538-543.
2. Hesham. ALABBASI, Florica. Moldoveanu, and Alin Moldoveanu, "Real Time Facial Emotion Recognition using Kinect V2 Sensor", IOSR Journal of Computer Engineering (IOSR-JCE), Vol.17, Issue 3, Ver. II, May- 2015, pp 61-68.

3. Hesham. ALABBASI, Florica Moldoveanu, Alin Moldoveanu and Zaid Shhedi," Facial Emotion Expression Recognition With brain activities Using Kinect Sensor V2", International Research Journal of Engineering and Technology (IRJET), vol. 02, Issue 02, May-2015, pp 421- 428.
4. Zaid Shhedi, Alin Moldoveanu, Florica Moldoveanu, and Hesham. ALABBASI," Evaluating Smart Communication and Monitoring Technologies for Hospital Hygiene Workflows", Journal of **I**nformation **S**ystems & **O**perations Management, accepted paper.
5. Hesham ALABBASI, Alex Gradinaru, Florica Moldoveanu, and Alin Moldoveanu "Human Motion Tracking & Evaluation using Kinect V2 Sensor", The 5th IEEE International Conference on E-Health and Bioengineering - EHB November 2015, Iasi, Romania (in press).
6. Hesham. ALABBASI, Alex Gradinaru, Florica Moldoveanu, and Alin Moldoveanu "Virtual sports training system using Kinect V2 sensor", University POLITEHNICA of Bucharest Scientific Bulletin, C series: Electrical Engineering and Computer Science, 2016, accepted paper.

1.4. STRUCTURE OF THE THESIS

Chapter 2 outlines first the basic concepts of a face detection system with its challenges, some faces detection approaches, modern face detection techniques, skin color methods, color models adequate for skin detection, and then our proposed method for face detection and the experimental results. We limit our discussion to those aspects that are relevant to the content of this thesis and we encourage the reader to consult the references for further details.

In Chapter 3, we introduced the "emotions" definitions, theories of emotions, explained human facial emotions, facial emotion recognition approaches, a brief introduction to the markup languages and Microsoft Kinect V2 sensor. After an intuitive description of the affective computing, we presented the approaches based on Kinect sensor to recognize facial emotional expressions, and then our proposed method to recognize facial emotion expressions using Kinect V2 sensor with the experimental results.

In Chapter 4, we presented the relation between human brain and emotions, we explained the brain parts associated with emotions through the studies of many researches related to this field. An emotional brain-computer interface was presented, and then our proposed approach to recognizing brain activities using facial emotion expressions.

In Chapter 5, we introduced motion tracking and analysis systems, optical marker, mark less, and 3D markers scanners. A human motion tracking system using Kinect sensor, the Kinect sensor and its joint tracking features, techniques, and related researches on motion and gesture recognitions were presented. Finally, we presented our proposed approach for Virtual sports training system, based on human motion tracking and evaluation using Kinect v2 sensor.

Chapter 6 contains the original contributions of the thesis and future work.

CHAPTER 2. HUMAN FACE DETECTION

Face detection is necessary and used by various applications like face recognition, gender identification, face tracking in video sequences, gender classification, facial emotion expression recognition, biometric identification, human computer interaction systems, and others [HA14].

The human face is a dynamic object and incorporates a high degree of variability in its looks, that makes face detection a troublesome drawback in computer vision. Many techniques and approaches have been proposed and presented, starting from easy edge-based algorithms to composite high-level approaches using advanced pattern recognition approaches [IK05].

2.1. FACE DETECTION AND RECOGNITION SYSTEM

Face detection is the first stage in a face recognition system. The input to the face detection system can be a digital image or video sequence. The output is an identification or verification of the subject or subjects that appear in the image or video. Some approaches outlined a face recognition system as a 3 stage process. Face detection and feature extraction stages may run at the same time.

Figure 2.1. Typical face recognition system

The aim of face detection is to identify the face region in an image for further processing, usually for face recognition [HA14]. So, the system positively identifies a certain image region as a face. This procedure has several applications like face tracking, pose estimation or compression. The next stage, feature extraction involves obtaining relevant facial features from the data. These features might be certain face regions, variations, angles or measures, which might be human-relevant (e.g. Eyes spacing) or not. This stage has alternative applications like face tracking or facial emotion expression recognition. Finally, the system recognize the face in the images. In an identification task, the system would report identity from a database. This stage involves a comparison approach, a classification algorithm, and an accurate measurement.

These stages can be merged, or new ones can be additional. Therefore, we can notice many alternative engineering approaches to face recognition. Face detection and recognition could be performed in tandem, or proceed to an expression analysis [RV10].

2.2. FACE DETECTION CHALLENGES

Face detection is a complicated and challenging task due to many problems like [DD13]:

- Out-of-plane rotation: frontal, 45 degrees, profile, upside down.
- The presence of a beard, mustache, glasses etc.
- Facial expressions.
- Occlusions by long hair, hand.
- Image conditions.
- Lighting condition.
- Distortion.

2.3. FACE DETECTION APPROACHES

There are several approaches to face detection, that can be mainly classified into four categories: *Knowledge-based*, *Template matching*, *Appearance-based*, *Feature invariant based approaches* (see Figure 2.2) [MH02].

Generally, face detection methods combine some or all of the four approaches to realize high face detection accuracy and a low false detection rate. Detection rate and the number of false positives are necessary factors in evaluating face detection systems. Detection rate is the ratio between the number of correct faces detected by the system and the real number of faces within the image [EA02] [RV10].

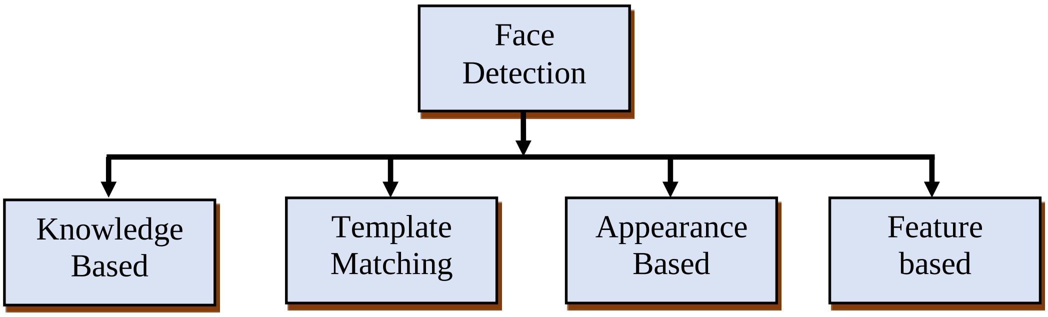


Figure 2.2. Types of face detection techniques

1. ***Knowledge-based approach:*** this approach depends on the human knowledge, and uses this knowledge as rules to detect the faces. An example, a face usually contains 2 eyes, one nose and a mouth that have certain distances and known positions relative to each other [MH02] [MC14].
2. ***Template matching approach***: this approach uses many templates to search out the face category and extract the face features. Template-matching compares the closely values from the face candidate image with the face template, measures the extent of similarity and concludes whether or not it's a human face. Gray color space was chosen because the most effective result is experimentally obtained [MM14] [MC14].
3. ***Appearance-based approach:*** this approach depends upon a collection of representative training facial images to find out the face patterns. Usually, appearance-based approach has shown superior performance compared to other approaches [BC10] [MC14].
4. ***Feature invariant based approach***: in this approach, the input image is processed to extract and identify structural features. Standard statistical pattern recognition methods are then used to locate the face by differentiating between facial and non-facial regions [MM14] [MC14].

2.4. Modern Face Detection Techniques

Nodaway, face detection has an important role in authentication & identification; therefore it has become a pioneer issue in today's computer vision. The following are some of the modern techniques used today for face detection.

1. ***Motion Base***: this technique is applied when a video sequence is available. Many objects have silhouettes in the video; this technique uses image subtraction to extract the moving foreground from the static background. The face is then located by examining the silhouette or the color of the difference image. This approach will not work well when there are a lot of moving objects in the video sequence [JC] [MM14].
2. ***Gray Scale Base***: one of the important features within the face is gray scale information. Eyebrows, nose, and lips appeared darker than other surrounding facial regions. Distinctive recent feature extraction algorithms were used to search for a neighborhood with low sphere inside segmented facial regions. These algorithms enhanced the input facial images by contrast-stretching and grayscale morphological procedures to improve the quality of neighborhood dark patches and in this manner make detection easier. The extraction of dark patches is accomplished by low-level grayscale threshold [MM14].
3. ***Edge detection approach***: it's an essential area in computer vision. Edges characterize the boundaries between regions in a digital image, which assists with segmentation and object recognition. They can indicate where shadows fall in an image or some other particular change in the intensity of an image. The images are enhanced by applying a median filter to remove noise and histogram equalization for contrast adjustment. For the second step, the edge images are constructed from the enhanced image by applying the Sobel operator, and then a novel edge tracking approach is applied to extract the sub-windows from the improved image taking into account edges. The nature of edge detection is very subject to lighting conditions, the vicinity of objects of comparative intensities, the density of edges in the scene, and noise. While each of these issues can be taken care of by adjusting certain qualities in the edge detector and changing the limit for what is considered an edge, no great method has been determined for setting these qualities in an automatic way, so they should be manually modified by an administrator whenever the detector is run with an alternate arrangement set of data [AR13] [MM14].
4. ***The neural network approach:*** numerous neural network algorithms have been used with face detection. The benefit of utilizing neural networks for face detection is the feasibility of training a system to capture the complicated category conditional density of face patterns. However, one negative mark is that the network structure

must be extensively tuned (number of layers, the number of nodes, training rates, etc.) to induce exceptional performance [MM14]. Detects faces by sub-sampling various regions of the digital image to a standard-sized sub image and after that passing it through a neural network filter. Generally, the algorithm performance is so good for frontal-parallel faces, but performance deteriorates when extended to different views of the face [DD13].

5. ***Depth-based approach:*** essential facial features are restricted on the basis of facial depth information. In the initial step, sets of images containing frontal perspectives are sampled from the input video sequence, and then point correspondences over a substantial divergence range are determined using the multi-resolution hierarchical matching algorithm. Finally, the facial features are located taking into account the depth information [DD13].
6. ***Skin color model-based approaches:*** color is one of the most widely used visual feature in face detection. Numerous methodologies used a skin color as a feature; convert the color image into an appropriate color space like RGB, HSI, YCbCr, or YIQ, to find skin color. These color spaces are more vigorous to the lighting conditions than the RGB color space and, subsequently, are perfect for face detection with distinctive illuminations. The mean or covariance matrices of the skin color are then computed from the skin colors and, finally, the results of these computations are used to find the probability that each pixel within the input image is, for sure, a skin color [HA14].
7. ***Viola Jones Method*:** this method depends on the use of Haar-like features that are evaluated rapidly through the use of a new image representation. Based on the concept of an “Integral Image”, it generates a large set of features and uses the boosting algorithm “AdaBoost” to reduce the over complete set and the introduction of a degenerative tree of the boosted classifiers provides for robust and fast interferences [PV04]. The detector is applied in scanning way and used for grayscale images, the scanned window that is applied can also be scaled, and additionally the features evaluated [MM14]. The fundamental benefit of this method is that it has uncompetitive detection speed while moderately high detection precision, comparable to much slower algorithms [VG14].
8. ***Gabor Feature Method:*** A Gabor filter is a linear filter whose impulse response is characterized by a harmonic function multiplied by a Gaussian function [AK09].

Within the field of image processing, a Gabor filter is used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been observed to be especially suitable for texture representation and differentiation. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave [IW]. *M. Sharif et al*. proposed an Elastic Bunch Graph Map (EBGM) algorithm that successfully implements face detection using forty different Gabor filters for face detection [MS11].

9. ***Support Vector Machine (SVM):*** SVM introduced first by *Osuna et al*. for face detection [EO97]. SVMs act as another paradigm to train polynomial functions, neural networks, or radial basis function (RBF) classifiers. SVMs works on induction principle known as structural risk minimization, which focuses on minimizing an upper bound on the expected generalization error. An SVM classifier is a linear classifier where the isolating hyperplane is selected to minimize the expected classification error of the unseen test patterns. Ion Marqu´es, proposed that the multi-class pattern recognition system can be obtained by combining two-class SVMs [ION10]. SVMs have also been used to detect faces and pedestrians in the wavelet domain
10. ***Local Binary Pattern (LBP):*** one of the efficient techniques used to describe the image texture features is the local binary pattern. It has benefits like, rapid-speed computation and fixed roundness, which make the process of image retrieval, texture examination, face recognition, image segmentation, etc. much easier. In LBP, every pixel is assigned a texture value, which can be actually combined with a target for tracking thermographic and monochromatic video. The major uniform LBP patterns are used to recognize the key points in the objective region and then form a mask for joint color-texture feature selection [VG14]. Within the detection of moving objects, LBP was applied using background subtraction [TA04].
11. ***AdaBoosting approach:*** is an approach to machine learning based on combining many relatively weak and incorrect rules, to create a highly accurate prediction. The AdaBoost approach was the first practical boosting algorithm, and one among the foremost wide used and studied with applications in numerous fields. The boosting algorithm used to train a classifier which is capable of processing images quickly while having high detection rates. AdaBoost is a learning algorithm which produces

a great classifier by selecting visual features in a group of simple classifiers and combining them linearly. Despite the fact that AdaBoost is more impervious to overfitting than numerous machines learning algorithms [RM03], it is repeatedly sensitive to noisy data and outlier's. AdaBoost is called adaptive because it uses multiple iterations to generate a single composite strong learner. AdaBoost creates the strong learner (a classifier that is well-correlated to the true classifier) by iteratively adding weak learners (a classifier that is only slightly correlated to the true classifier) [VG14].

2.5. SKIN COLOR BASED METHODS

The human skin color is an important feature, many types of research focus on skin color in human face detection. There are several advantages from using skin color as a feature for face detection. Human skin has a consistent color, which is distinct from many objects from images and is highly robust to geometric variations of the face pattern, invariant to face orientation and scale, stable against occlusions. Skin color has proven to be a useful and robust cue for face detection, localization and tracking [HA14].

Many types of analysis in face detection found that skin color is the best factor for identification. Skin detection could be a highly popular and helpful technique for detecting and tracking physical body regions [DD13]. It receives a lot of attention chiefly due to its wide range of applications corresponding to face detection and tracking, naked people detection, hand detection and tracking, people retrieval in databases and web, etc. The main goal of skin color detection or classification is to create a decision rule that may discriminate between the skin and non-skin pixels. Distinguishing skin color pixels involves finding the range of values that most skin pixels would fall in, during a given color space.

In general, a decent skin color model should have a high detection rate and reduce the false negative rate. That is, it should find most skin pixels, whereas minimizing the quantity of non-skin pixels classified as skin.

D. DUAN et al., showed that even though of different races, different ages, and a different gender, the difference in color chrominance is far less than the difference in the brightness. Skin distribution shows clustering distribution in the skin-color space without luminance influence [DD09].

There are two main approaches in face detection supported skin color [FA11]:

1. The first approach is pixel-based model, which relies on processing the pixels for all regions of the human skin color. In this approach, every pixel is processed independently, then based on facial structure or other choices, it will decide which set of points that are marked as skin belongs to the face or not.
2. The second approach, based on the status of a region of the image. In this approach at first, the necessary attempt is done to segregate the region that may produce a face within the given image. And then using the previous information and knowledge, it will be decided that specifies if the region belongs to face or not.

2.5.1 COLOR SPECIFICATION

Since the beginning of 1990's, the cost of quality color images has become competitive with black-and-white. Obviously, to a human operator, color images are greatly useful for identification, since cues, for example, hair color and appearance can be used. Assuming that the image resolution and noise performance are adequate, the use of color images to identify faces should provide numerous advantages tantamount to those for manual identification [AM07].

Colors can be represented by three independent variables, either the LMS responses or the tristimulus values under the primaries of a given color system, such as RGB. In many situations (e.g., computer image processing), it is more convenient to represent a color from a different set of three independent variables HSL (or HSI, HSV). These are defined as

- Hue: the dominant wavelength, the redness of red, the greenness of green, etc.
- Saturation: the purity of the color, or how much white is contained in the color. For example, red and royal blue are more saturated than pink and sky blue, respectively.
- Luminance (intensity, value): the intensity of the light [RO99].

Each colored image contains values of the three above components which make human skin hues cover a wide range of the color spectrum, this makes it exhausting to discover the human skin color based on the RGB color values themselves. Hence, the requirements for various color spaces that reduce the variance in human skin colors arises [AM07]. Reflected light that shine to an object will be received by the human vision system cells then processed by the nervous system, and eventually leads to the perception of color by a human. In this manner, the color is a substitution for recognitions that people can get by reflected lights from an object surface. In addition to object color, in color image processing fields, also shape, structure, and low-level color characteristics of an object are important and necessary information.

Considering that forty-fifth percentage of the ordinary human face comprises of skin, skin color is a suitable choice, as a benchmark, to discover the faces among an image. To detect faces in color images, the color models should be established and substantial color ranges using predefined information should be used. In spite of the fact that, at the first look, it appears that there are large diversity and distinction in people's skin colors, the result from many researchers showed that the skin color ranges are considerably small and even in several cases, the variations are due to differences in image light intensity [FA11]. By tolerating these outcomes, we can set up face detection systems, based on the human color characteristics, with less calculation and more accuracy.

2.5.2 SKIN COLOR MODELS

The color is a useful information for skin detection. The skin color is the first and a common approach for face detection, which helps to avoid the comprehensive search for faces in an image. To conduct a research on the face processing and detection, and also the implementation of such systems based on the color characteristics of the human face, it is necessary to study and analyze the color, color model, and color images. Researches that have been done so far include the existing definitions of color, available color models (YUV, RGB, HSI, HSV, Normal RGB, and YCrCb) and comparisons of such models [AS03].

Although, in general, color face detection systems are the same, but in detail they have some differences. Now we discuss three types of color spaces commonly used. In fact, all of the available color spaces can be converted into each other using one or more linear or non-linear transforms. The RGB and red, green, blue are known as main color-space and colors respectively. Other colors and color spaces can be achieved directly from the main space. The goal of selecting a particular color model is facilitating color description in a standard. Color modeling, actually, is determining a multidimensional coordinate system and the subspace inside it, in which each color, is expressed only by a point [FA11].

A. RGB COLOR SPACE

The RGB color space is a cube in a 3D coordinate system, with the three additive primary colors, Red, Green and Blue, varying from 0 to 1 on the three axes. This color space contains all the colors that can be obtained by combinations of the three primary colors. It is specific to most of the computer graphics devices, but it is not adequate for many image processing algorithms because the red, green and blue color components are highly correlated. Also, the RGB color model is light sensitive [DC01] [HA14].

B. HSV COLOR SPACE

The HSV color space is also based on three color components: H - the hue component, which defines the color, S - the saturation component, which specifies how pure the color is, and V - the value component, which specifies the brightness (intensity). By considering only the H and S components, we can make abstraction of lighting conditions. The HSV color space is a hexacone in a 3D coordinate system. H values vary from 0 to 1 on a circular scale, H=0, and H=1 representing the same color. S values vary from 0 to 1, 1 representing a color with 100% purity. V values vary from 0 to 1. Colors with S=0 represent different gray levels (the H component is not important). Colors with low S values cannot be well differentiated [DC01] [HA14].

The transformation from RGB to HSV is defined by the following formulas:

$$H = \begin{cases} \left(6 + \frac{G - B}{MAX - MIN}\right) \times 60°, if\ R = MAX \\ \left(2 + \frac{B - R}{MAX - MIN}\right) \times 60°, if\ G = MAX \\ \left(4 + \frac{R - G}{MAX - MIN}\right) \times 60°, if\ B = MAX \end{cases} \qquad (1)$$

$$S = MAX - MIN \qquad (2)$$

$$V = MAX \qquad (3)$$

where MAX = max (R, G, B), MIN = min (R, G, B).

C. YCBCR COLOR SPACE

YCbCr is one of primary color spaces used to represent digital video information. In the YCbCr color space, a color is represented by the brightness and two color difference signals. Y is the brightness (Luma) component, which represents the luminance and is computed as a weighted sum of RGB values. Cb and Cr are the chrominance components: Cb is computed as the difference between the blue component and a reference value, Cr is the difference between the red component and a reference value [DC01] [HA14].

As in the case of HSV space, by separating the luminance component from chrominance, makes the YCbCr color space luminance independent and more adequate than RGB for face detection by skin color recognition.

The transformation from RGB to YCbCr can be made by using the following formulas:

$$Y = 0.299R + 0.587G + 0.114B \quad (4)$$

$$Cb = -0.169R - 0.332G + 0.500B \quad (5)$$

$$Cr = 0.500R - 0.419G - 0.081B \quad (6)$$

2.5.3 HUMAN SKIN COLOR

Human skin color is distinctive from the color of numerous objects and, along these lines the statistical measurement of these attributes are of awesome significance for face detection. The color is a prominent feature of the human face. Utilizing skin color as a primitive feature for detecting face region has many benefits. Especially, the processing color is much faster than processing other facial features. Besides, color information is invariant to face orientation.

On the other hand, even under a fixed encompassing lighting, people have different skin color appearance. So as to successfully exploit skin color for face detection, one needs to find feature, in which human skin colors, cluster tightly together and resides remotely in background colors. Several approaches within the literature used different detection approaches, either based on the RGB, chromatic (CbCr) or Hue and Saturation (HSV) space [DO01] [AL]. In YCbCr space, the brightness (luminosity) is stored as a simple element (Y) and the value of chrominance as two different elements (Cb and Cr). The values Cb and Cr represent the difference between light blue and the current calculated value, as well as the difference between red and the current calculated value, respectively. Figure 2.3 represents the chromatic distribution of the human skin area with respect to Cb and Cr values [SO03].

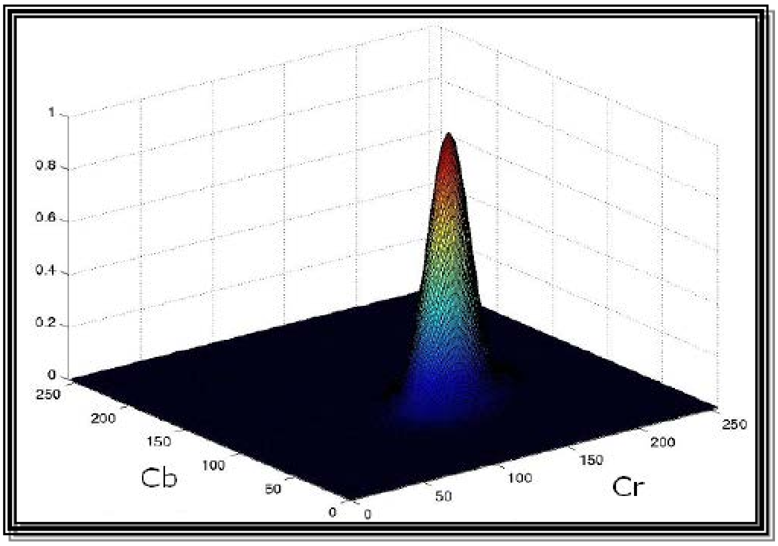
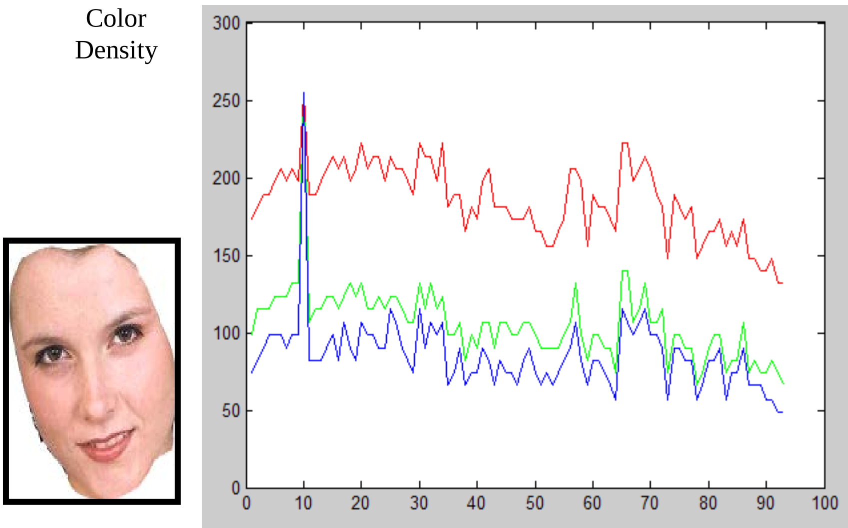


Figure 2.3. Human skin chromatic distribution

2.5.4 THE SKIN COLOR AS A FEATURE

Faces often have a characteristic color which is possible to separate it from the rest of the image (see figure 2.4) [BA01]. Numerous methods exist to model the skin color, essentially using Gaussian mixtures or simply using lookup tables. In some studies, skin color pixels are filtered, from the sub-image corresponding to the extracted face, using a look-up table of skin color pixels. The skin color table was obtained by collecting, over a large number of color images, RGB (Red-Green-Blue) pixel values in sub-windows previously selected as containing only skin. The weak point of these methods is the color similarity of hair pixels and skin pixels. For better results, the face bounding box should be used, thus avoiding as much hair as possible.

Color
Density



No. of pixel value

Figure 2.4. Image and RGB distributions of filtered skin color pixels

As is often done in skin color analysis studies, the histogram H (g) of R, G, and B, pixel components for different face images, which is computed as the number of the pixel at gray level g. Such histograms are characteristic of a specific person, but also discriminate among different persons (see Figure 2.5).

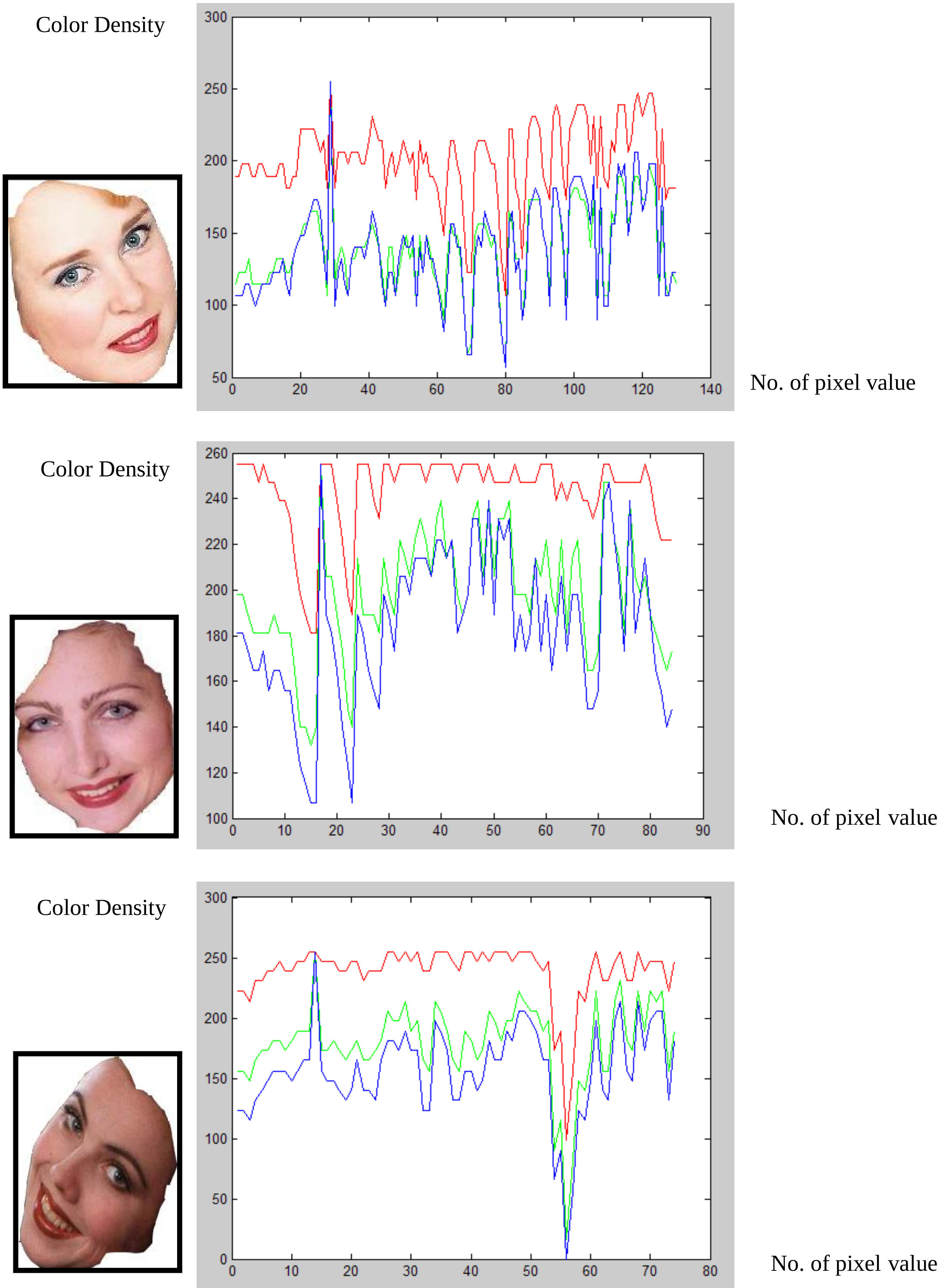


Figure 2.5. Images and RGB distributions of skin color pixels of different persons

2.6. PROPOSED FACE DETECTION APPROACH

The proposed system presents a suitable face detection algorithm that can detect the face with different depths in the image in both indoor and outdoor environments. A light correction step was used to adjust the illumination of the input image, and a feature invariant approach based on the skin and edge information was used for face detection. Figure 2.6 illustrates the steps of our algorithm for face detection.

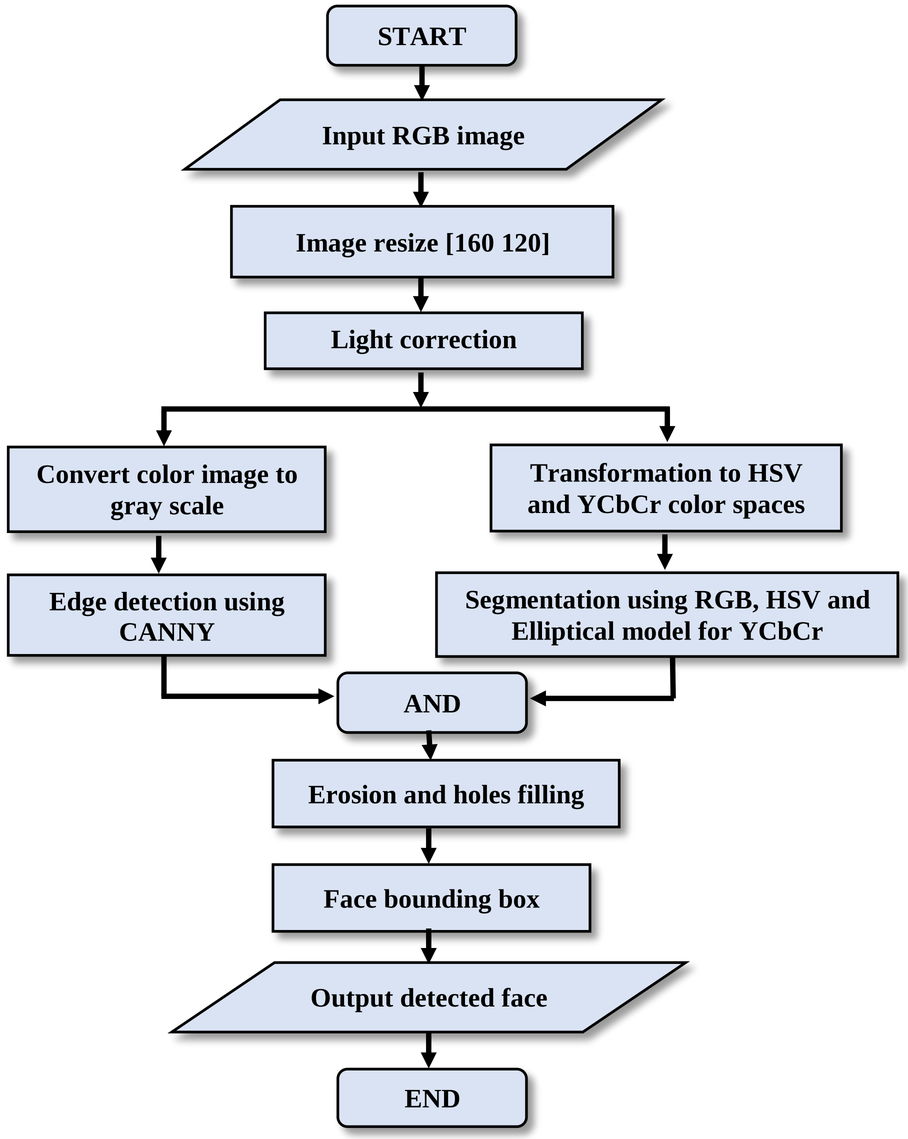


Figure 2.6. Block diagram explaining proposed face detection system

IMAGE RESIZING AND LIGHT CORRECTION

The input image is a true RGB color image (24-bits). First, the image is resized to (160 height × 120 widths) by using "resize" MATLAB function. This size is preferred because the shape of the human face can be usually framed in a rectangle of this form. The resized image is decomposed into its original three color bands (Red 160×120, Green 160×120, Blue 160×120).

The skin color is often affected by lighting conditions, which can create deviations from the real color of skin. Thus, the second step is light correction. The aim of this step is to adjust the luminance of the image so that all images can be considered as obtained under the same lighting conditions. This step was implemented using a lighting compensation algorithm which is named Gray World Theory (GWT) [KH11]. The (R', G', B') light corrected values of an (R, G, and B) color are computed as follows:

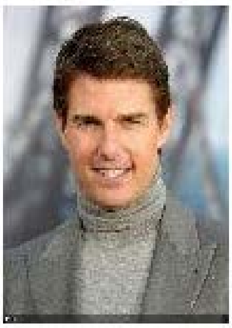
$$K = \left[\frac{R_{average} + G_{average} + B_{average}}{3}\right] \qquad (7)$$

$$R' = R * \left[\frac{K}{R_{average}}\right] \qquad (8)$$

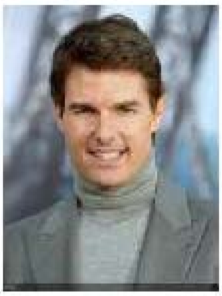
$$G' = G * \left[\frac{K}{G_{average}}\right] \qquad (9)$$

$$B' = B * \left[\frac{K}{B_{average}}\right] \qquad (10)$$

Where $R_{average}$, $G_{average}$ and $B_{average}$ are the averages of the R, G, B values of the image pixels and K is the average for all colors in the image. Figure 2.7 shows the original image, the resized and light corrected image.



(a)



(b)

Figure 2.7. (a) Original image (b) Resized and light corrected image

SKIN SEGMENTATION WITH RGB AND HSV COLOR SPACES

The image obtained from the previous step is then converted to two types of images:

1- Gray scale image.
2- HSV color image.

In the RGB color space, the detection was based on the following set of conditions [AM07]:

(a) For uniform daylight illumination:

$$\begin{gathered} \mathrm{R} > 95 \; AND \;\; \mathrm{G} > 40 \;\; AND \;\; \mathrm{B} > 20 \; AND \\ \mathrm{R} > \mathrm{G} \;\; AND \; \mathrm{R} > \mathrm{B} \; AND \; |\mathrm{R} - \mathrm{G}| \; 15 \;\; AND \\ (\mathrm{Max}\{\mathrm{R}, \mathrm{G}, \mathrm{B}\} - \mathrm{Min}\{\mathrm{R}, \mathrm{G}, \mathrm{B}\}) > 15 \end{gathered} \tag{11}$$

(b) Under flashlight or daylight called lateral illumination:

$$\begin{gathered} \mathrm{R} > 20 \; AND \; G > 210 \; AND \; B > 170 \; AND \\ |\mathrm{R} - \mathrm{G}| < \; 15 \; AND \; R > G \; AND \;\; G > B \end{gathered} \tag{12}$$

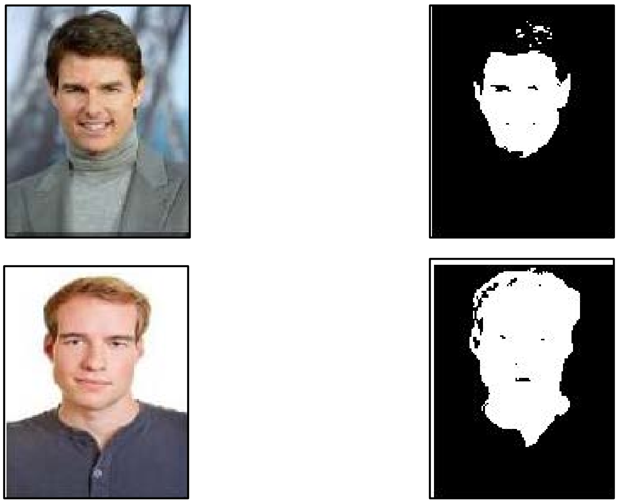
In the HSV color space, we can work independently with the intensity (V) and the two chrominance components, H (hue) and S (saturation). *Teskeridou* and *Pitas* [FG12] worked within the HSV color space and found the following thresholds for the pixels having skin-like colors:

$$\mathrm{V} \geq 40 \tag{13}$$

$$0.2 < \mathrm{S} < 0.6 \tag{14}$$

$$0° < \mathrm{H} < 25° \;\; or \;\; 335° < \mathrm{H} < 360° \tag{15}$$

The segmented image by using these thresholds may include non-skin regions like background, hair, dress, etc. Any pixel from the input resized and light corrected image whose value satisfies conditions from (11) to (15) will have the value 1 (skin color) in the segmented image, otherwise the value is 0 (non-skin). Figure 2.8 shows the result of skin segmentation using RGB and HSV color spaces, original image after resized and light correction.



(a) (b)

Figure 2.8. (a) Original image after resize and light correction (b) Skin segmentation using RGB and HSV color spaces

Skin Segmentation with YCbCr Color Space

The resulted segmented image is converted to its original color space by multiplying it with the light corrected image color components (Red, Green, and Blue). We saved the result as a new image, named "imrgbhsv.jpg".

We used a built-in MATLAB function to convert the imrgbhsv image to YCbCr color space. After that, we used the Elliptical model for the segmentation of the skin tones.

The expression of face skin color elliptical model is [JQ12]:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} C_b - C_x \\ Cr - C_y \end{bmatrix} \tag{16}$$

$$\frac{(x - ec_x)^2}{a^2} + \frac{\left(y - ec_y\right)^2}{b^2} = 1 \tag{17}$$

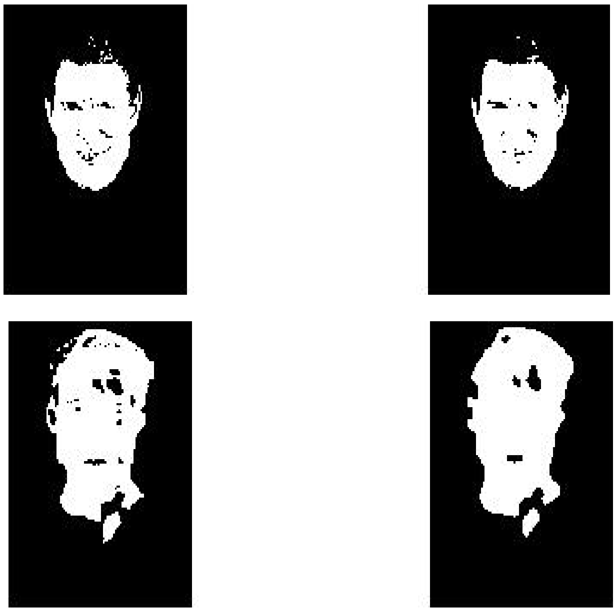
The elliptical parameters are as follows:

C_x = 109.38, C_y = 152.02, ec_x = 2.41, ec_y = 2.53, a = 25.39, b = 14.03 and θ = 2.53 (in radians).

x and y are calculated for each pixel of the YCbCr image by using equation (16). Those pixels which satisfy (17) are considered as skin color.

The result is a binary image with 1 for skin regions and 0 for non-skin regions. We used the median filter to remove the noise in the result binary image, implemented by "medfilt2" MATLAB function. This filter removes noise, with less attenuation of edges. Figure 2.9

shows the result of skin color segmentation using the three color spaces, before and after noise removal.



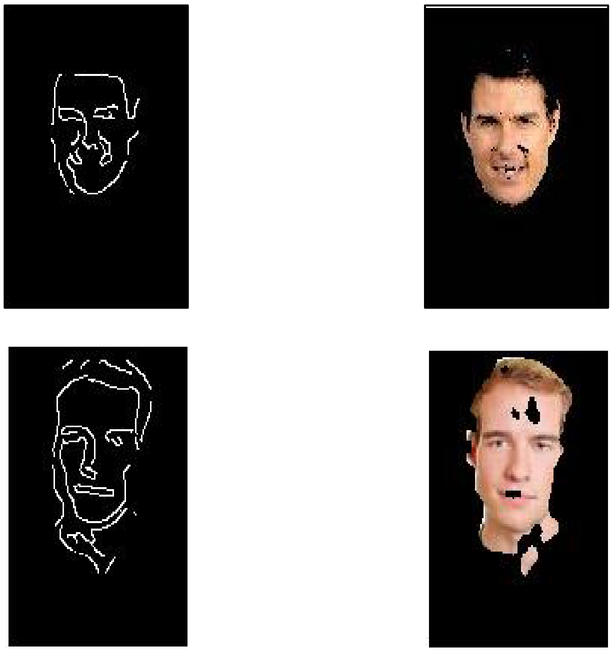
(a) (b)

Figure 2.9. (a) Binary image obtained by segmentation with RGB, HSV and YCbCr color spaces (b) Binary image after skin color segmentation and noise removal

EDGE DETECTION

An image can contain background regions with a color similar to skin color. After skin color segmentation, such regions can be connected to the face region. In order to easily localize the face in the image, it is necessary to separate such background regions from face region. This can be done using the edges of the initial converted gray scale image. Some common methods are used for edge detection like Sobel, Prewitt, Roberts and Canny edge detectors, also the LOG (Laplacian of Gaussian) edge detector. CANNY edge detector can detect weak edges and retain strong boundary. Therefore, we used the CANNY edge detector to determine the edges in the image. The output of the Canny edge detector is a binary image where the boundary pixels have value 1 and the others values are 0.

Then, we combined the segmented image from the previous step with the complemented output image of the Canny edge detector using the AND logical operation. Figure 2.10 shows the output of the Canny edge detector and the skin face region.



(a) (b)

Figure 2.10. (a) The output of the Canny edge detector (b) Skin face region

MORPHOLOGICAL OPERATIONS

Morphological techniques are well developed for binary images, but many methods can be successfully extended to grayscale. For a binary image, white pixels are normally taken to represent foreground regions, while black pixels denote background. Virtually all mathematical morphology operators can be defined in terms of combinations of erosion and dilation along with set operators such as intersection and union. The operators are particularly useful for the analysis of binary images and common usages include edge detection, noise removal, image enhancement and image segmentation [RG98].

Morphological operations like erosion and hole filling widely used in image processing. By erosion, pixels situated on an object boundary are iteratively removed, thus leading to object thinning. Hole filling operation fills the gaps with white spaces so as to make it a solid white blob. We used the MATLAB function "imerode", with a neighborhood specified by a square structuring element of width 4 (se = strel ('square', 4), then applied MATLAB function "imfill" on the eroded result image. Figure 2.11 shows the resulted images for this stage.

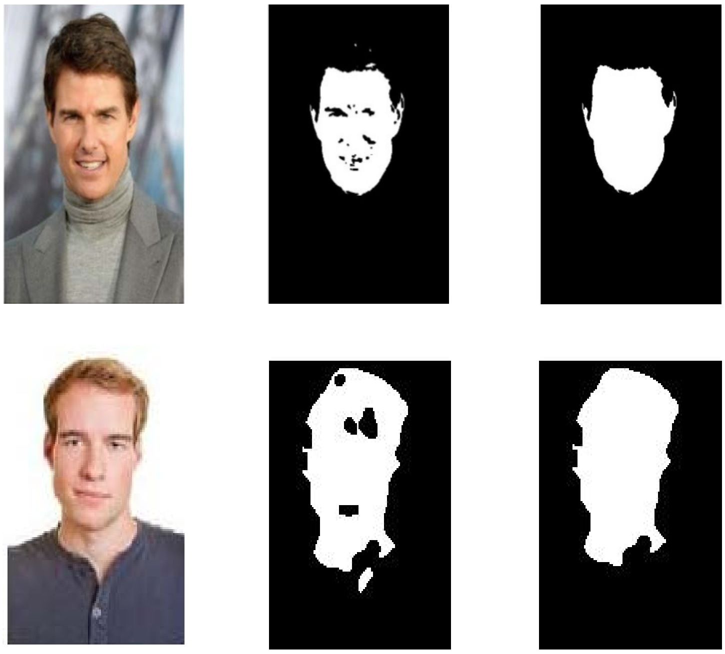


Figure 2.11. (a) Original images (b) Eroded Images (c) Image after holes filling

EXTRACTING FACE AREA

In order to extract the minimum bounding box that contains the face and with less hair and clothes, we applied our algorithm to the result of the morphological processing. We used the original image, let us call it (I), and P_1, the image from the previous stage. The minimum bounding box P_2, that contains the face, is constructed by determining the threshold values for image height and width. The rows and columns in the image that belong to the face region are those that exceed these thresholds. The following algorithm is used to find the minimum bounding box and extract a new image which contains this boundary.

Algorithm: Find minimum bounding box

Input: Binary image P_1, and original image I

Output: New image with bounding box P_2

Process:

1. Find the sum of rows and columns in the input image I and set it to S, also find the vectors row and col, which represent the summation of rows and the summation of columns in the image P_1.

$$S = \sum_{x=0}^{X} \sum_{y=0}^{Y} I(x,y)$$

$$row(y) = \sum_{x=0}^{X} P_1(x,y)$$

$$col(x) = \sum_{y=0}^{Y} P_1(x,y)$$

2. Calculate row ratio *row_r* and column ratio *col_rt.*

   row_rt= S/no. of rows; row ratio of the whole image.

   col_rt= S/no. of columns; column ratio of the whole image.

3. Calculate height threshold ht and width threshold vt

   ht = 0.66 * row_rt; height threshold.

   vt = 0.43 * col_rt; width threshold.

4. Apply the height and width thresholds horizontally and vertically to extract a row_new and col_new.
5. Calculate the bounding box dimensions (x_start, x_end, y_start, y_end).

   x_1=image width, x_2=0, y_1=image height, y_2=0.

   For each row i and column j of the original image

6. Extract the face image from the original image by using the dimensions of the bounding box.

   For i = x_start to x_end

   For j = y_start to y_end

   P2(i,j,1)=R(i,j)

   P2(i,j,2)=G(i,j)

   P2(i,j,3)=B(i,j)

The values (0.66) and (0.43) in the above-proposed algorithm, which are sensitivities to the threshold for height and threshold for the width of the box respectively, are calculated by trial. They are computed according to the ratio estimate of facial height and width. Figure 2.12 shows the original images and minimum bounding box (final image).

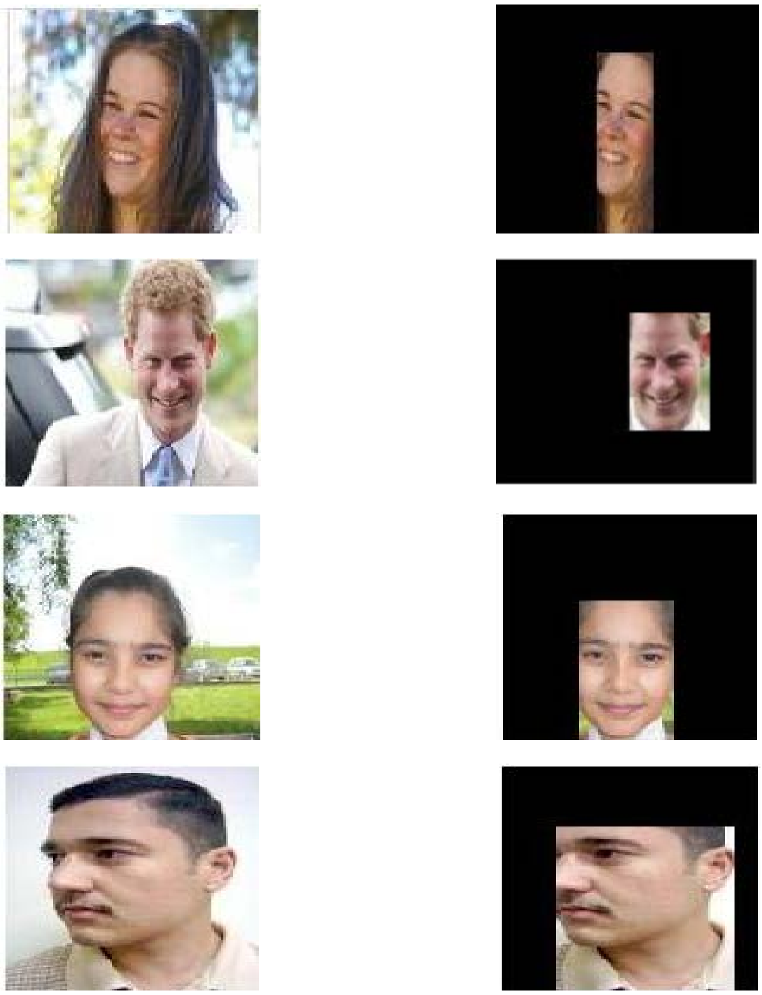
(a) (b)

Figure 2.12. (a) Original images (b) Minimum bounding box (final image)

EXPERIMENTAL RESULTS

We implemented our human face detection approach using the image processing toolbox from MatlabR2012a and tested it on 114 color images. Some of the images were downloaded from the internet, others have been collected from friends, all images containing one face in different poses. The image set contains different indoor and outdoor scenes with various lighting conditions.

The results of our experiment demonstrate that our approach can detect the facial region in a very good manner, with less hair and clothes. Figure 2.13 illustrates the results of skin region segmentation of the proposed approach.



(a) (b)

Figure 2.13. (a) Original images (b) Final images

CONCLUSIONS

Face detection techniques are increasingly used in real-world applications and products. For instance, most digital cameras today have built-in face detectors, which can help the camera to do better auto-focusing and auto-exposure. Our proposed approach is a very good algorithm for face detection from single face images, which is based on skin color and edge information and gives us high correctly detected face over the actually used images. Using skin segmentation in three color spaces combined with information of input image edges is a useful cue. Also, applying morphological operations, like erosion and holes filling, and de-noising the skin regions which were extracted proved to be valuable enhancements. Some false alarms happened, but the overall performance of the proposed algorithm is still quite satisfactory. The algorithm can be used in real-time applications with indoor and outdoor images and for advanced used like face recognition.

CHAPTER 3. REAL TIME FACIAL EMOTIONS RECOGNITION

Facial expression is composed of specific positions of the muscles beneath the skin of the face. The movement of facial muscles in such positions conveys the emotional state of an individual to observers. Facial expressions are a form of nonverbal communication. The emotional facial expression is an imperative component in human communication, alongside different kinds of non-verbal communication, for example, gestures and posters. The human facial expression, with its cunning and moment movements, conveys an amazing amount of information that can reflect passionate emotions. Observing people's facial expressions can offer some assistance with understanding their feelings.

3.1 THE THEORIES OF EMOTIONS

Emotions have always been subject to great debate. They were first just researched by philosophy and assumed to be disorganizing and irrational. *Darwin* outlined the primary trendy theory of emotions, recognizing their essential role in the survival and their accommodating function. Many researchers presented other developmental theories, proposing diverse sets of basic emotions with different including criteria. This assortment in the names and number of the basic emotions considered led to criticism and doubt on the terrible existence of basic emotions [AO90].

Definition

Emotion is a crucial aspect within the interaction and communication between individuals. Despite the fact that emotions are intuitively known to every person, it is difficult to define emotion. The Greek philosopher *Aristotle* considered of emotion as a stimulus that evaluates experiences based on the potential for gain or pleasure. In the seventeenth century, *Descartes* considered emotion to mediate between stimulus and response [CL04]. These days, there is still little agreement about the meaning of emotion. *P. Kleinginna* and *A. Kleinginna* gathered and analyzed ninety-two meanings of emotion from the literature that day [PK81]. They inferred that there is little consistency between diverse definitions and proposed the following comprehensive definition:

"Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems", which can:

1- Offer ascent to full of feeling encounters, for example, pleasure/displeasure;

2- Generate cognitive procedures, for example, emotionally relevant perceptual effects, appraisals, labeling processes;

3- Activate widespread physiological adjustments to the arousing conditions; and

4- Lead to behavior that is often, but not always, expressive, goal-directed, and adaptive.

This definition shows the diverse sides of emotion. From one viewpoint, emotion creates particular feelings and impacts someone‘s behavior. This portion of emotion is surely understood and is much of the time noticeable to a man himself or to the outside world. Emotion additionally adjusts the human brain state and directly or indirectly influences several processes [CL04].

In spite of the difficulty of defining it, emotion is omnipresent and an important factor in human life. Individual's moods intensely impact their way of communicating, additionally their acting and productivity [FL].

Theories of emotions stretch back at least as far the stoics of ancient *Greece*, as *Plato* and *Aristotle*. Also, there are sophisticated theories in the work of the philosopher as *Baruch Spinoza*, *Rene Descartes*, and *David Hume*. Later theories of emotions have a tendency to be educated by advances in observational research. Many theories have been discussed of emotion, yet five of them are the most generally discussed. These theories based on somatic, which claim the bodily responses rather than judgments essential to emotions. The following is a brief explanation of the five theories.

1. ***James-Lange Theory***: it's the oldest theory of the five theories, *William James* (1884: 1889) reports that emotions happen when the impression of an exciting fact causes a collection of substantial changes, and "our feeling of the same changes as they occur is the emotion". The same thought occurred to *Carl Lange* (1984) around the same time [JP03]. This theory says that emotion is not directly caused by the perception of an event, but rather by the bodily response caused by the event. This means that, in order to experience emotion, we must first experience the bodily response (e.g., fast breathing, racing heart, sweaty hands) that corresponds to the emotion [WJ84]. Moreover, James-Lange theory concerns itself with a sort of instantaneous interpretation to decide which emotions we feel. For example, if something frightening happened, we would notice our heart rates instinctively increase, and our skin perspiring. We would then take these physical

symptoms and based on the situation, may exert the emotion of "fear" and react further accordingly. So, it is only because of our body's preparation for something scary or happy, do we experience the emotion attached to the physical side-effect. The problem here lies within the proof of causality, and this is indeed hard to empirically observe and prove. The lack of substance in this theory is a bit baffling, and while it may have some merit, in particular, situations that require "fight-or-flight" thinking (those that are almost instinctive) [PR11]. M. Marin, expressed this theory in these words: "Environmental stimuli trigger physiological responses and bodily movements, and emotion occur when the individual interprets his or her visceral and muscular responses. I must be afraid because my heart is pounding and I am running like crazy" [MM06].

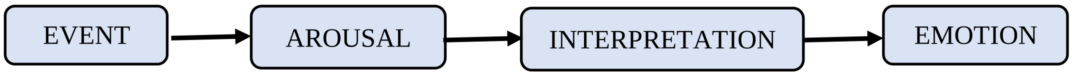


Figure 3.1. James-Lange theory of emotions

2. ***Cannon-Bard Theory*****:** *Philip Bard* concurred with *Cannon*, and kept inspecting emotion in the brain. *Cannon-Bard* theory (also known as the thalami theory) is an interesting reversal. This theory states that, the brain is stimulated as a result of a situation, and instead of separately going through the thought process (the cerebral cortex) or just to the autonomic system, it goes to both at the same time, giving the simultaneous display of physiological symptoms and the perceived "emotion" [CW27]. It is possible to explain classical conditioning this way, and since such conditioning is subconsciously present within everyone on some level to certain things (scary animals, criminals, etc.), it could very well be a rational explanation that the lack of a mechanic is the most accurate mechanic. For instance, when your heart is racing, it may mean you are angry, yet it might likewise mean you are energized positive. This implies our brain cannot simply rely on our bodily reactions to know which emotion we are encountering (i.e., there must be something else that lets us know whether we are angry or excited) [WC27]. From their research, *Cannon* and *Bard* concluded the experience of an emotion does not rely on upon input from the body and how it is reacting. Both the experience of the emotion and the bodily response to occur at the same time independently of each other, and once again *could* be very logical. The theory simply puts physiological symptoms on the same part of progression as the emotion itself. The underlying reason for this is, in this theory, there is a biological component. M. Marin, expressed this theory in these words, " Emotion is a

cognitive event that is enhanced by bodily reactions. Bodily reactions do not cause emotion, but rather occur simultaneously with the experience of emotion. “I am afraid because I know bears are dangerous” [MM06].

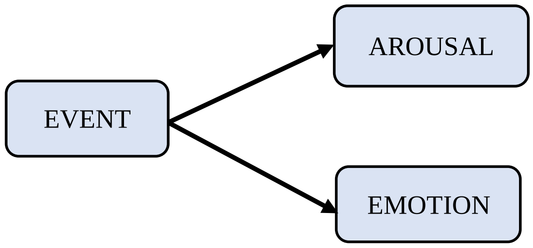


Figure 3.2. Cannon-Bard theory of emotions

3. ***Schachter-Singer Theory*****:** The two-factor theory of emotion was expanded by *Schachter* and *Singer*. This theory separates cognitive functions and physiological factors as a means to an end emotion. The subject experiences physiological arousal first, but then must analyze the reason for this arousal. This may sound similar to the James-Lange theory, but has a critical difference; one would not directly analyze the physiological symptoms, but the reason behind them [SS62]. The bodily reaction is the same; a person might experience very different emotions depending on the type of situation he is in. This is perhaps the most complicated mechanical, as it very much separates the physiological and cognitive processes, but maybe through this provides the most versatile explanation for emotion. The most interesting part of the theory, however, is the situation when there are no clues as to why physiological events are happening. Historically, the theory was tested with injections of epinephrine, which increases heart rate and makes a person start to perspire. The theory very controversially places the human as a “blank slate” of emotion, and unless there is a definite explanation for the reason behind the physiological symptoms, a person will not know what emotion to display. M. Marin expressed this theory in these words, “Emotions depend upon a kind of double cognitive interpretation: We appraise the emotion-causing event while also evaluating what is happening with our bodies. “I am afraid because I know bears are dangerous and because my heart is pounding” [MM06].

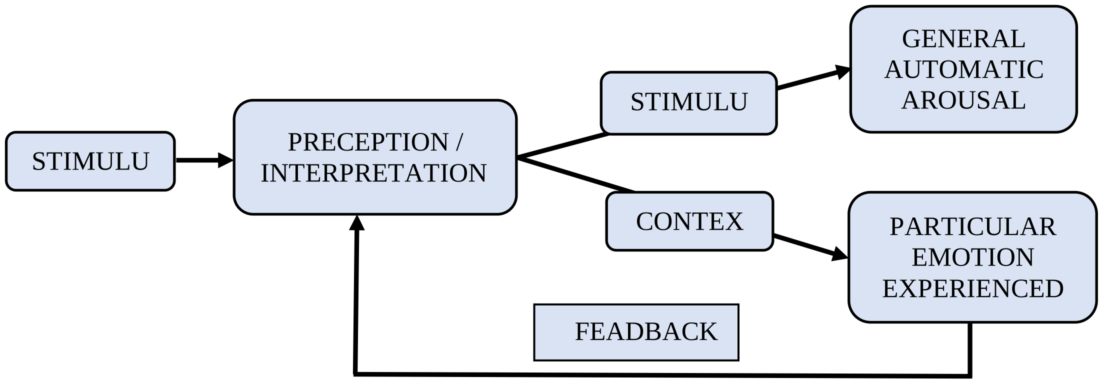


Figure 3.3. Schachter-Singer theory of emotions

4. ***Cognitive Appraisal Theory*:** this theory was developed by *Magda Arnold* [MA60] and *Richard Lazarus* [RL91]. According to appraisal theories of emotion, thinking must happen first before the experience of emotion. *Richard Lazarus* was a pioneer in the field of emotion, and this theory is frequently referred to as the *Lazarus* theory of emotion. Appraisal theory opens the door to a much more complex assessment and description of emotions.

   Appraisal literally is as it sounds, a person's appraisal in their mind of a situation affects their emotion thereof [RL91]. In light of this theory, the sequence of events first includes a stimulus, followed by the idea, which then prompts the simultaneous experience of a physiological reaction and the emotion. For instance, in the event that we experience a bear in the forested areas, we may immediately start to believe that we are in extraordinary danger. This then prompts the emotional experience of fear and the physical reactions associated with the fight-or-flight reaction. Appraisal works on a situational basis and has a person deciding upon an emotion based upon their mindset and current motivations. For instance, going to an interview for a job may induce the emotions of happiness, as well as fear. The ramifications of pre-planned actions and ones currently taking place in relation to the wants, needs, and motivations of a person directly affect the pattern of emotionality in the absence of physiological effects. The theory also opens up possibilities to account for multiple stage behaviors. For instance, if one knows that feeling excited about something, and being positive can lead to success which will lead to even more happiness, one will work to achieve this end emotion. This theory can be used to link directly to many other cognitive patterns, namely coping methods, which are used to combat certain emotions [MM06]. This theory addresses the main problem with previous emotional theories and removes the necessity of physiological symptoms for emotions. Cognitive processes and the appraisal and internal evaluation thereof can be self-governing bodies of emotion, without any pre-existing physiological symptoms.

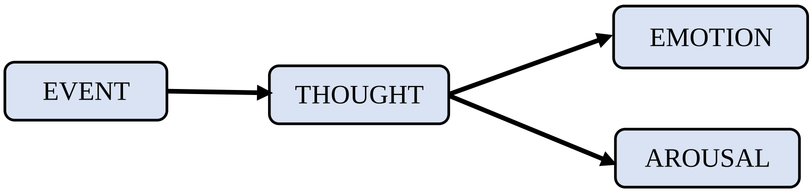


Figure 3.4. Cognitive Appraisal theory of emotions

5. ***Facial-Feedback Theory:*** its proposed early in the 19th century by *Charles Darwin* and *William James*, this theory states that without a physical change in specifically the facial area, there are only thought about an emotion, and it is not actually experienced [JA72] [JA84]. The facial-feedback theory of emotions suggests that facial expressions are connected to the experience of emotions. *Charles Darwin and William James* [JA90] both noted early on that sometimes the physiological responses often had a direct impact on emotion, rather than simply being a consequence of the emotion. Supporters of this theory recommend that emotions are specifically attached to changes in facial muscles. For instance, people who are forced to smile pleasantly at a social function will have a superior time at the event than they would if they had frowned or carried a more neutral facial expression. The theory itself sounds fairly simple, but perhaps the most interesting are the experiments performed to test it, and the empirical evidence that followed. *Strack, Martin, and Stepper* [ST88] tested the hypothesis involved holding a pen in one's mouth in both a way that made one frown or smile and a control group that had no pain in their mouth. The "smiling" pen group resulted in rating material more "funny" than the frowning group. It could very well be that simple facial muscle contractions do have specific effects on our emotional output.

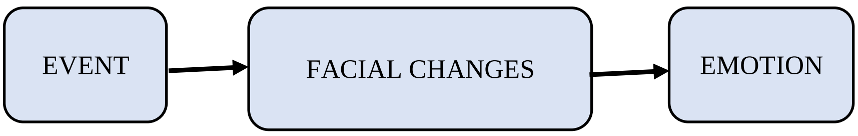


Figure 3.5. Facial-Feedback theory of emotions

3.2 AFFECTIVE COMPUTING

Affective computing was first popularized by *Rosalind Picard's* book. ”Affective Computing” was introduced for research into automatic sensing, detection, and the interpretation of affect and identified its possible uses in human-computer interaction (HCI) contexts. An automatic affect sensing has attracted a great interest from different fields and research groups, including brain science, cognitive sciences, phonetics, computer vision, speech analysis, and machine learning. The advancement in automatic affect recognition depends on the progress in all of these seemingly disparate fields. Affective computing has developed and enhanced over the previous decades [WR97].

There are numerous fields where the automatic affect sensing would be helpful:

1. ***Automatic tracking***: the instances of such systems are, air traffic control, nuclear power plant surveillance, and operating a motor vehicle. These systems have a tracking tool which makes these systems more secure and proficient, as a result of earlier detection of negative affective states, alert the administrator or others around him, thus helping to avoid accidents [WR09].
2. ***Automatic monitoring***: (Safety critical & medical environments). Affect sensing systems could also be utilized to monitor patients in hospitals, or when medical staffs are not promptly accessible or overburdened. It could also be used in helping living situations to monitor the patients and inform the medical staff during emergencies. *A. Ashraf et al*. [AA09] proposed in medical applications of affective computing, an automatic detection of pain. *F. Cohn. et al*. [FC09], presented another example of a developed system to the automatic detection of depression from facial and auditory signals.
3. ***Entertainment industry***: Automatic detection affects computer games, giving the players a more customized experience if the emotional state of the player was known to the game.
4. ***Affective information****:* this information could be utilized to increase restricted channels of communication, for example, text messaging. *K. Hook* [KH09], developed a system and is named eMoto. It is a phone with an expanded SMS service where users, besides sending a text message, are permitted to select its background from colorful and animated shapes. These backgrounds are supposed to represent the emotional content along two axes of arousal and valence.
5. ***Automatic recognition****:* emotional understanding of human behavior and emotions; almost the peoples with the autistic range disorder have difficulty to understand the

emotional states of others and expressing these states themselves [SB85] [WR09]. An automatic recognition of affect could assist autistic persons to express their own affective states [WR09], by permitting them to express exterior what is being felt in the interior. It could also be conceivable to build systems that assist these people better understand the affective states of others. Systems could be utilized to accelerate the current labor intensive, error prone and tedious task of labeling emotional data. One of interest work made by *M. Girard. et al.* [MG13], in which automated tools for facial expression analysis (Action Unit detection) are utilized to support and inform existing theories of depression.

In addition, affect synthesis is advantageous for the formation of believable virtual characters (avatars) [JC00], and robotic platforms as it permits these agents to act more like humans [DR11]. Systems that are able to analyze effect can frequently be used to synthesize it if generative models are used, hence affects synthesis would benefit from better effect analysis. *McDuff et al.* [DM13], proposed a work in determining if people liked certain advertisements and were likely to watch them again by analyzing their smiling behavior. This work was useful for advertising and marketing domains, where new evaluation metrics are constantly sought. The authors gathered a dataset in naturalistic situations by using the webcams of the guests to their website.

3.3 Human Facial Expressions of Emotions

One of the first scientists who stressed the importance of facial expression in the face to face communication of human beings was *Charles Darwin*. He defines the facial expressions of emotions (anger, fear, surprise, disgust, joy, sadness and other more complex emotions) and the body language “the language of the emotions” [AH15]. Facial expressions can also provide information about the cognitive state of a person, such as confusion, stress, boredom, interest, and conversational signal [MS03].

The basic emotions, are emotions that have been scientifically proven to have a certain facial expression associated with it. In 1980, *Robert Plutchik* built a wheel-like diagram of emotions visualized eight essential emotions, in addition to eight derivative emotions each composed of two basic ones [RP11]. The wheel of emotions, identifies eight basic emotions, joy, sadness, trust, disgust, fear, anger, surprise, and anticipation.

The wheel of emotion is linked to the color wheel in which the primary colors combine to form the secondary and complementary colors. These basic emotions, then mix and combine to form a variety of emotions. For instance, anticipation plus happiness might consolidate to form optimism (Figure 3.6).

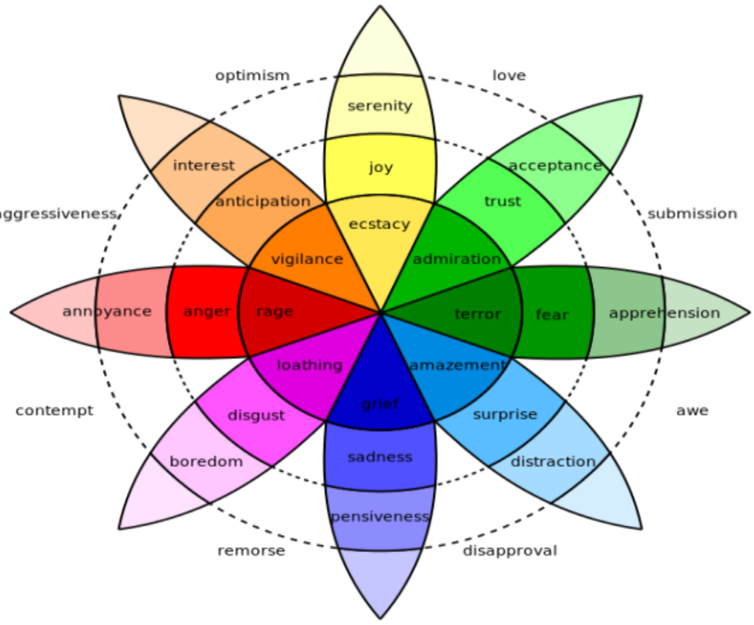


Figure 3.6. Robert Plutchik's wheel of emotions

Ortony and Turner [AO90] collected and examined an extensive variety of research on identification of basic emotions.

Table 3.1. Identifications of basic emotions from different theorists

Theorist	Basic Emotions
Plutchik	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
Arnold	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
Ekman, Friesen, and Ellsworth	Anger, disgust, fear, joy, sadness, surprise
Frijda	Desire, happiness, interest, surprise, wonder, sorrow
Gray	Rage and terror, anxiety, joy
Izard	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise
James	Fear, grief, love, rage
McDougall	Anger, disgust, elation, fear, subjection, tender-

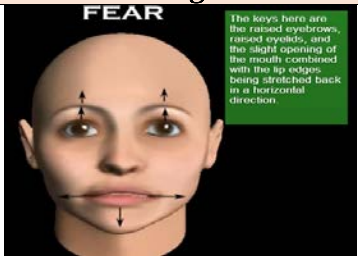
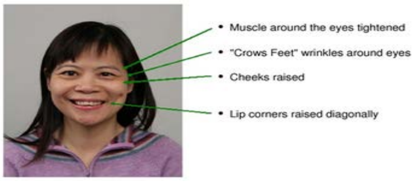
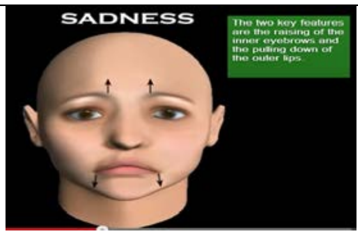
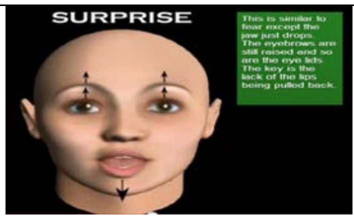
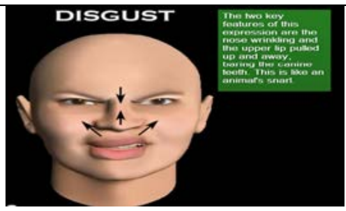
	emotion, wonder
Mowrer	Pain, pleasure
Oatley and Johnson-Laird	Anger, disgust, anxiety, happiness, sadness
Panksepp	Expectancy, fear, rage, panic
Tomkins	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
Watson	Fear, love, rage
Weiner and Graham	Happiness, sadness

Different researchers proposed that there are around six or seven basic emotions that are experienced in societies throughout the world. Psychologist *Paul Ekman* made what is known as the Facial Action Coding System (FACS) [PK02], a taxonomy that measures the movements of all the face's 42 muscles as well as the movements of the head and eyes. *Ekman* found that there are six facial expressions universal to people all over the world. These original six emotions he recognized were happiness, sadness, surprise, fear, anger, and disgust. Later he added a seventh emotion - contempt. These six emotions used in *Ekman*'s studies known as the "basic emotions". A few scientists said there are less than six basic emotions, and some said there are more (*Ekman* himself has now scaled up to 21); however, the idea remains as before: "Emotions are biologically innate, universal to all humans, and displayed through facial expressions" [JK14].

Facial Emotional Indicators

Ekman and Friesen's [PK03] research on the facial expression of emotion uses scores of photographs demonstrating emotions of surprise, fear, disgust, anger, happiness, and sadness. The authors explained how to distinguish these basic emotions effectively and how to advise when people try to cover, simulate, or neutralize them. Also presented several practical features exercises that help actors, teachers, salesmen, counselors, nurses, law-enforcement personnel and physicians and everyone else who deals with people to become adept, perceptive readers of the facial expressions of emotions. Table 3.2. shows the basic emotions and their facial signals [MI] [HU10].

Table 3.2. The basic seven emotions and their facial signals

Emotion	Facial signals	Image
Fear	Eyes wide, closed or pointing down; raised eyebrows; mouth open or corners turned down; chin pulled in; head down, white face.	
Anger	Eyes wide and staring; eyebrows pulled down (especially in middle); wrinkled forehead; flared nostrils; mouth flattened or clenched teeth bared; jutting chin, red face.	
Happiness	Mouth smiling (open or closed); possible laughter; crows-feet wrinkles at sides of sparkling *eyes*; slightly raised eyebrows; head level.	
Sadness	Eyes cast down and possibly damp or tearful; head down; lips pinched; head down or to the side.	
Surprise	Eyes wide open; eyebrows raised high; mouth dropped wide open with consequently lowered chin; head held back or tilted to the side.	
Disgust	Eyes and head turned away; nostrils flared; nose twisted in sneer; mouth closed, possibly with tongue protruding; chin jutting.	

Contempt	Eyes neutral, Lip corner pulled up and back on one side only	

3.4 EMOTION MARKUP LANGUAGE

The Human Markup Language provides a vocabulary which will permit a wide assortment of human-centric applications to be constructed. It will also allow for a greater depth of information about peoples to be assembled and used with existing applications at the circumspection of the individual concerned. A markup language is a language that shows text, so that the computer can manipulate the text. Most markup languages are human readable because the illustrations are written in a matter that can be distinguished from the text [ME].

There are distinctive markup languages [BR15] [TA08] [EX14] [AN07] [KO11] [MI]:

- HTML (HYPERTEXT MARKUP LANGUAGE): it’s a language of the web. All web pages are written in HTML. HTML characterizes the way that images, multimedia, and text are shown in web browsers.
- XML (EXTENSIBLE MARKUP LANGUAGE): XML is a language for writing markup languages. Extensible Markup Language (XML) is a markup language that characterizes a set of rules for encoding documents in a format that is both human-readable and machine-readable. There are also several standardized languages already created with XML: MathML for defining mathematics, SMIL for working with multimedia, XHTML, and numerous others.
- XHTML: stands for extended hypertext markup language. XHTML 1.0 is HTML 4.0 redefined to meet the XML standard. XHTML is written in lower case. While HTML tags can be written in UPPER case, MiXeD case, or lower case, to be correct, XHTML tags must be all lower case. All XHTML elements must have an end tag. Elements with only one tag, such as HR and IMG need a closing slash (/) at the end of the tag:

  <hr />

  <img />

XHTML requires that tags are nested correctly. If you open a bold (B) element and then an italics (I) element, you must close the italics element (</i>) before you close the bold ().

- DHTML: stands for Dynamic Hypertext Markup Language. Dynamic HTML, or DHTML, is utilized to make interactive and animated web sites by using a combination of a static markup language (such as HTML), a client-side scripting language (such as JavaScript), and a presentation definition language. DHTML permits scripting languages to change variables in a web page's definition language, which thusly influences the look and function of page substances after the page has been completely loaded and during the viewing process.
- Voice XML: Voice XML is utilized in Voice interaction between humans and computers, mainly in systems that enable the user to, for instance, check his credit card balance over the phone. The love- like dialogue management and speech recognition- is defined by voice XML.
- LaTeX: A document markup language utilized most part by mathematicians, authors, etc. to typeset their content. It is suitable for representing mathematical formulas.

The concept of emotion markup language itself is quite fascinating. Working with emotion-related states in technological contexts requires a standard representation format. Nowadays, emotion-oriented computing systems are a reality; a standardized way of representing emotions and related states is obviously clear [MS]. For real-world human-machine interaction systems, which typically consist of multiple components covering various aspects of data interpretation, reasoning, and behavior generation, it is evident that emotion-related information needs to be represented at the interfaces between system components.

W3C Emotion Incubator Group (EmoXG) [W3], defined an Emotion Markup Language (EML), as "a general-purpose emotion annotation and representation language, which can be used in a wide variety of technological applications where emotions need to be represented". The W3C has announced that "Emotion Markup Language (EmotionML) 1.0 has become a recommendation to represent emotion-related states in data processing systems" [EM14], four years after a draft of the new specification first emerged.

As indicated by the W3C, the development of the web in the last years has made the requirement for a new markup language which can be utilized in areas, for example, the automatic recognition of emotion-related states from user behavior, and the generation of emotion-related system behavior. The W3C stated on its official Emotion Markup Language page, “As the Web is becoming ubiquitous, interactive, and multimodal, technology needs to deal increasingly with human factors, including emotions”.

From a practical point of view, the modeling of emotion-related states in technical systems can be significant for two reasons.

1. To improve computer-mediated or human-machine communication. Emotions are a fundamental part of human communication and ought to, along these lines, be considered, e.g., in emotional chat systems or emphatic voice boxes. This includes specification, analysis, and presentation of emotion-related states.

2. To improve systems processing proficiency. Emotion and intelligence are emphatically interconnected. The modeling of human emotions in computer processing can help to design more proficient systems, e.g. utilizing emotional models for time-critical decision enforcement.

A standardized approach to markup the data required by such "emotion-oriented systems", can possibly support improvement fundamentally in light of the fact that:

- Data that was annotated in a standardized way can be interchanged between systems more effectively, thereby simplifying a market for emotional databases
- The standard can be utilized to facilitate a business of providers for sub-modules of emotion processing systems, e.g. a web service for the recognition of emotion from text, speech or multi-modal input [MS14].

According to the W3C, there are three wide use cases for EmotionML, which are:

1. Manual annotation of material, including emotionality, for example, annotation of videos, of speech recordings, of faces, of texts.

2. Automatic recognition of emotions from sensors, involving physiological sensors, talk recordings, facial expressions and from multi-modal combinations of sensors.

3. Generation of emotion-related system reactions, which may include thinking about the emotional implications of events, emotional prosody (mood, stress, and intonation of speech) in synthetic speech, facial expressions and gestures of embodied agents or robots, the decision of music and colors of lighting in a room.

The W3C said of the need for EmotionML, “Interactive systems are likely to include both analysis and generation of emotion-related behavior”, besides, systems are likely to benefit from the data that was manually annotated, be it as training data or for rule-based modeling. In this manner, it is attractive to propose a single EmotionML that can be utilized as a part of every one of the three contexts.

Specific applications of EmotionML include observing customer sentiment, fear detection in security systems, a new design in computer games, social robots, more expressive speech synthesis, richer transcription services, and in helping people with disabilities like autism.

The working group iteratively extracted requirements on the markup language from a number of 39 collected use cases [LI05]. Based on the requirements, the syntax for EmotionML has been produced in a sequence of steps [MS12].

The following snippet exemplifies the principles of the EmotionML syntax [SC13].

```
<sentence id="sent1">
Do I have to go to the dentist?
</sentence>
<emotion xmlns="http://www.w3.org/
2009/10/emotionml" category-set=
"http://.../xml#everyday-categories">
<category name="afraid" value="0.4"/>
<reference role="expressedBy"
url="#sent1"/>
</emotion>
```

The following properties can be observed.

- The emotion annotation is self-contained within an <emotion> element;
- All emotion elements belong to a specific namespace;
- It is explicit in the example that emotion is represented in terms of categories;
- It is explicitly from which category set the category label is chosen;
- The link to the annotated material is realized via a reference using a URI, and the reference has an explicit role.

It is possible to use EmotionML both as a standalone markup and as a plug-in annotation in different contexts. Emotions can be represented in terms of four kinds of descriptions taken from the scientific literature: categories, dimensions, appraisals, and action tendencies, with a single <emotion> element containing one or more of such descriptors [FB14]. EmotionML makes scientific concepts of emotions practically applicable. This can help potential users to identify the suitable representations for their respective applications.

3.5 APPROACHES TO FACIAL EMOTION RECOGNITION

Facial Action Coding System (FACS), was the earliest system that characterized the physical expression of emotions. It was developed in 1978, by *Paul Ekman*, along with *Wallace Friesen* [PK02] and is still widely used today. Their system is used to measure all visually distinguishable facial movements and for encoding how movements of facial muscles result in changes in the appearance of the face. *Ekman and Friesen* studied anatomy and found the associations between the action of muscles, and the changes in facial appearance (Figure 3.7). Some appearance changes are the outcome of movements of multiple muscles and some muscles can have more than one action. Because of this, they named the measurements of FACS action units (AUs). AUs are the actions performed by individual muscles or muscles in combination [PK02]. The Facial Action Coding System (FACS) is “a human-observer-supported system that has been modified to facilitate objective measurement of subtle changes in facial appearance caused by contractions of the facial muscles” [PK03]. Through using 44 action units, FACS is able to give a linguistic description of all visibly expressions.

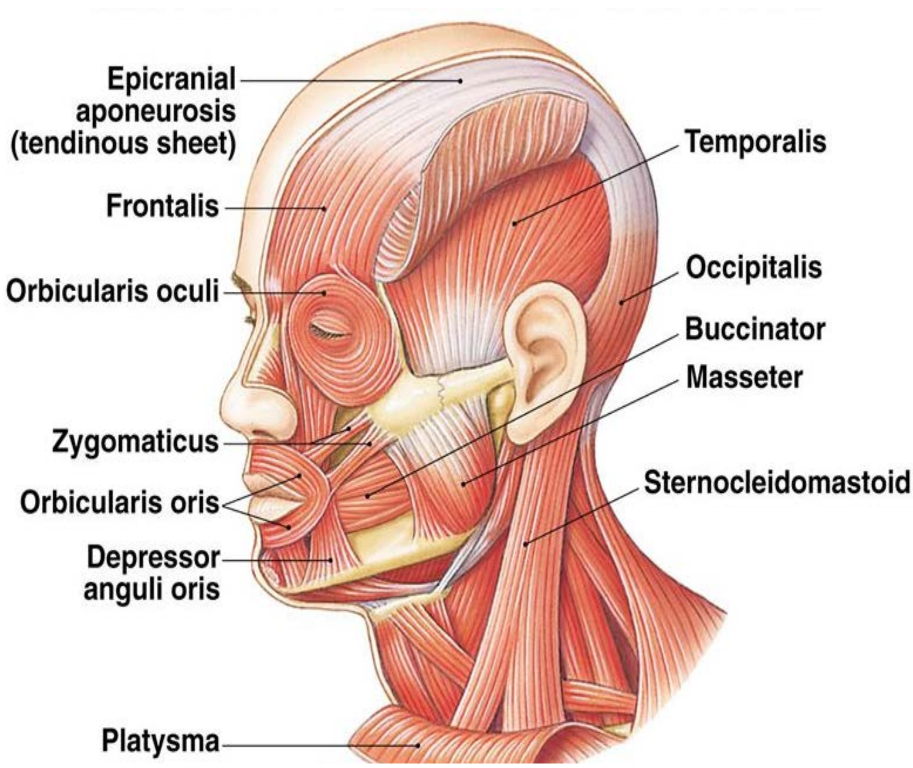


Figure 3.7. Muscles of the head and neck [ME63]

A facial expression recognition system performed by computers should consist of:

1- Locate the facial region in the scene like, image, video sequence (also referred to as face detection).

2- Extract the facial features of the detected face region (e.g., detecting the shape of facial components or describing the texture of the skin in a facial area), referred to as facial feature extraction.

3- Analyze and classify the information from the motion of facial features and/or the changes in the appearance of facial features into some facial expression- interpretative categories such as facial muscle activations like smile or frown, emotion (affect) categories like happiness or anger, attitude categories like, disliking or ambivalence, etc. (this step is also referred to as facial expression interpretation).

The six basic emotions (fear, sadness, happiness, anger, disgust, surprise) proposed by *Ekman* and *Discrete* emotion theorists are the most commonly used facial expression descriptors in message judgment approaches [DK00]. An automatic facial expression system can be applied to human-computer interaction, stress monitoring systems, low-bandwidth video conferencing, human behavior analysis, etc. [GW91].

An overview of the early works in facial expression analysis can be found in [AS92]. Later, the research of developing automatic facial expression recognition systems has attracted in a great deal of consideration from a wide range of fields, a more recent and complete overview is referred to [MP00].

The approaches to facial expression recognition can be generally divided into two categories [YT02]:

1- Geometrical feature-based approaches. This approach depends on the geometric facial features which exhibit the shapes and locations of facial components such as eyebrows, eyes, canthus, nose, mouth etc.

2- Appearance-based approaches. This approach used the entire face or particular regions in a face image to extract features by the mean of optical flow or some kinds of filters.

Some approaches concentrated on the separation of facial expression at the level of emotion models, some other approaches have the ability to segregate expressions at a fine-grained level by means of the recognition of action units [JF98]. Some approaches can completely automatically recognize expressions from image sequences, but others still need to manually mark some feature points before the recognition step. With few exceptions, most proposed approaches have used relatively limited data sets.

M. Pantie and L. J. M. Rothkrantz, [MP00], provided a review with comparisons of the existing approaches. The potential benefits from efforts to automate the analysis of facial expressions are varied and numerous and span fields as diverse as cognitive sciences, medicine, communication, education, and security.

In computer science and computing technologies, facial expressions provide a way to communicate basic information about the needs and demands of the machine. Where the user is looking (i.e., gaze tracking) can be adequately used to free computer users from the classic keyboard and mouse. Additionally, certain facial signs (e.g., a wink) can be connected with certain commands (e.g., a mouse click) offering an alternative to the traditional keyboard and mouse commands. The human ability to ‘‘hear’’ in noisy environments by means of lip reading is the premise for bimodal (audiovisual) speech processing (Lip-Movement Recognition), which can prompt to the realization of robust speech-driven user interfaces. To make a convincing talking head (avatar) representing a real person, recognizing the person’s facial signals and making the avatar respond to those using synthesized speech and facial expressions is essential.

The focus of the relative, recently initiated research area of affective computing lies in sensing, detecting and interpreting human affective states, for example, satisfied, disturbed, confounded, etc., and devising appropriate means for handling this effective information in order to enhance current HCI designs. The tacit assumption is that in many situations human-machine interaction could be enhanced by the presentation of machines that can adapt to their users and how they feel

As facial expressions are our direct, naturally preeminent means of communicating emotions, machine analysis of facial expression forms an indispensable part of the effective HCI designs. Monitoring and interpreting facial expressions can also provide important information to lawyers, police, security, and intelligence agents regarding person’s identity (research in psychology suggests that facial expression recognition is much easier in familiar persons because it seems that people display the same, ‘‘typical’’ patterns of facial behavior in the same situations), deception (relevant studies in psychology suggest that visual features of facial expression function as cues to deception), and attitude (research in psychology indicates that social signals including accord and mirroring – mimicry of facial expressions, postures, etc., of one’s interaction partner – are typical, usually unconscious gestures of wanting to get along with and be liked by the interaction partner).

Automated facial reaction monitoring may form a valuable tool in law enforcement as currently solely informal interpretations are typically used. Systems that can identify friendly faces or, additional significantly, recognize unfriendly or aggressive faces and inform the appropriate authorities, represent another application of facial activity technology [MP08].

Researchers concerned to analyze and recognize the facial expressions by utilizing different computer technologies like camera, video camera, Kinect sensor, also produced and developed many methods and languages, to recognize facial expressions [RZ08] [IK07] [PK06] [AK06] [AH15].

3.6 Real Time Facial Emotions Recognition using Kinect

3.6.1 Microsoft Kinect V2 sensor

The latest version of the Kinect V2 sensor was produced by Microsoft Company in September 2014, which has a modified technology than the older versions [HM]. This device brings to the users some of the latest achievements in human computing technologies and is also a facility that enables researchers to develop various applications by allowing the people to interact naturally with computers by simply gesturing and speaking [AH15].

a. Kinect specifications

Some of its specifications are, new active infrared (IR) capabilities, a new color camera with high resolution up to 1080 pixels in capturing and viewing, a new depth sensing with on mode 0.5 to 4.5 meters and a new microphone array [HD].

With this device, a development kit (SDK 2.0) was provided, with the following features and benefits [HH]:

1- Improving hand, body positioning and joint orientation; it has the ability to track six people and 25 skeletal joints per person.

2- Windows store support; users and developers can create Kinect-enabled applications by using familiar Windows runtime components.

3- Unity Pro support; users and developers can create and publish applications by using many kinds of software known by them.

4- Powerful tools; Kinect studio APIs enable the user to modify custom tools, to record and display body data using XEF files.

5- Advanced face tracking, with a resolution up to 20 times greater and a mesh up to 1.000 points, the applications with a face will be a more accurate representation.

A development kit (SDK 2.0) with new facilities, drivers, tools, APIs, device interface, and many sample code in C#, C++, and Java to help the application developers. With this new (SDK), face and body tracking is more stable.

b. Features and distances from Kinect

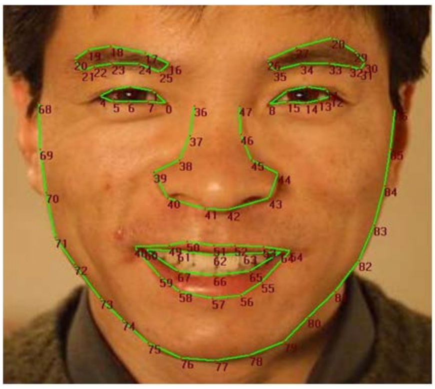
The Microsoft Face Tracking Software Development Kit (Face Tracking SDK), together with the Kinect for Windows Software Development Kit (Kinect for Windows SDK), helps us to build applications which have the ability to track human faces in real time.

Face Tracking SDK contains a face tracking engine, which can analyze the input from the Kinect camera, it can detect the head pose and face features depending on the points that can be tracked, and generate an information to the real time application. For instance, this information can be utilized in tracking person's head position. The Face Tracking SDK has the ability to track the 87 2D points, and 13 additional points that belong to the corners of the mouth, the center of each eye, the nose center, and for the bounding box around the head. Figure 3.8 shows the tracked points [AH15].

The 87 points are:

- 16 points for the eyes (0-15, 8 for the left eye and 8 for the right eye).
- 20 points for the brows (16-35, 10 for the left brow and 10 for the right brow).
- 12 points for the nose (36-47).
- 20 points for the lips (48-67,12 for the exterior lips, 8 for the interior lips)
- 19 points for the cheek (68-86).

These points are presented in an array and are defined in the coordinate space of the RGB image (640 x 480 resolution) returned by the Kinect sensor.



(a) (b)

Figure 3.8. (a) 87 2D points, (b) mesh for face tracking points for Kinect v2 sensor

3.6.2 RELATED RESEARCH TO FACE RECOGNITION AND FACIAL EMOTION EXPRESSIONS RECOGNITION USING KINECT SENSOR

Kinect sensor with its new feature for face tracking opened a new way to build systems which have the ability to track and recognize faces and facial emotions. Many papers in the field of computer vision have reported about systems of automatic face recognition and facial expression recognition, through using Kinect sensor device that gives a greater result than the camera or video camera. In this section, we present some of these papers.

Many researchers used Kinect sensor for face recognition. *Billy Y.L., Li1 Ajmal S., Mian Wanquan Liu and Aneesh Krishnal* [BI], presented an algorithm that uses a Kinect sensor to recognize faces under different conditions. In their experiments, they used a commonly reachable database which contains over five thousand facial images (RGB-D) that have a varying pose, expressions, illumination, acquired using the Kinect sensor. The recognition rates were 96.7% for the RGB-Data and 88.7% of noisy depth data alone.

Gaurav, G., Samarth, B., Mayank, V., and Richa, S. [GG], described a face recognition algorithm based on RGB-D images captured from a Kinect sensor. Their results demonstrate that using RGB-D information can improve face recognition performance compared to existing 2D and 3D approaches.

Also, for facial emotion expression recognition, *P. Lemaire, L. Chen, M. Ardabilian and M. Daoudi* [PL13], proposed in their work an approach to 3D facial expression recognition based on differential mean curvature maps and histograms of oriented gradients. The aim of their work, like many other works, was to classify the face emotions from the 6 primary emotions (happy, surprise, sadness, anger, fear, and disgust).

Facial expression analysis systems which are using 3D data can be characterized as static or dynamic. In dynamic systems, time is the fourth dimension, for this reason, they are sometimes called four-dimensional (4D) [GS11]. The techniques for detection vary greatly and include the use of Gabor wavelets [ML98], SIFT descriptors [SB11] and quadtree decomposition [GS12].

G R. Vineetha, C. Sreeji, and J. Lentini [GR12], presented a method for facial expression recognition by using MS Kinect in 3D from the input image. They used Microsoft Kinect sensor for the Xbox 360 video game console with its technique, described by MS Kinect for human face detection. After the human face was detected, edge detection, thinning, and token detection are performed. The user has to give the input threshold value for the detection of tokens. It is a difficult task to decide the best threshold value to generate the tokens. Their results showed that the expression of sadness and disgust were more difficult than the others to recognize.

A. Youssef, S. F. Aly, A. Ibrahim, and A. Lynn [AY13], proposed a system that attempts to recognize facial expressions using Kinect sensor. They constructed a training set containing 4D data (time is the 4th dimension) for 14 different persons performing the 6 basic facial expressions and used it with both SVM and k-NN classifiers. For individuals who did not participate in training the classifiers, the best accuracy levels were 38.8% (SVM) and 34.0% (k-NN). When considering only individuals who did participate in training, however, the best accuracy levels that they obtained raised to 78.6% (SVM) and 81.8% (k-NN). The authors also described the potential to use such a system for treatment of children with autism spectrum disorders (ASD).

Mihaela Puica, focused in her Ph.D. thesis [MP13], on emotion recognition from facial expressions. The main tool for doing this was a Microsoft Kinect sensor with the Face Tracking SDK. She used 58 points that define the brows, eyes and mouth, returned by the Face Tracking SDK, to measure 18 distances between face elements. These distances were used as inputs for a feed forward back propagation neural network, with 3 output neurons, each one grouping two emotions from the 6 basic emotions. The second experiment was by issuing directly the 58 coordinates to a neural network with 7 output neurons, one for each emotion, plus one for a neutral state. The accuracy of emotion recognition with data outside the training set was off 80%.

3.6.3 Our proposed approach to recognizing facial emotion expressions using Kinect V2 sensor

The purpose of the proposed system is to recognize facial emotion expressions of the patients after a stroke, in the initial phase of their recuperation. We have made experiments with this system using different kinds of persons, in order to establish its accuracy in recognizing emotions of different persons, in real time.

The steps of emotion recognition

Kinect V2 sensor with its new features, have been used in the system. Figure 3.9 illustrates the steps of our proposed facial emotion expression recognition system.

1. For step 1, we developed an interface application with Microsoft Visual Studio 2013 (C++), to track a person's face, and saved the 23 face animation feature values in a matrix. Figure 3.10 shows the 23 face animation values.

   These are: headpivx, headpivy, headpivz, jawopenn, jawsliderightt, left cheekpufff, left eyebrow loweredd, left eye closedd, right eye brow loweredd, righteyeclosedd, lip cornerepressed left lipcornerdepressed rightt, lip corner pulled leftt, lip cornerpulled rightt, lip spuckeredd, lip stretchleftt, lip cornerstretch rightt, lower lipdepressed leftt, lower lipdepressedrightt, right cheekpuffedd, facepitchh, faceroll, and faceyaww.

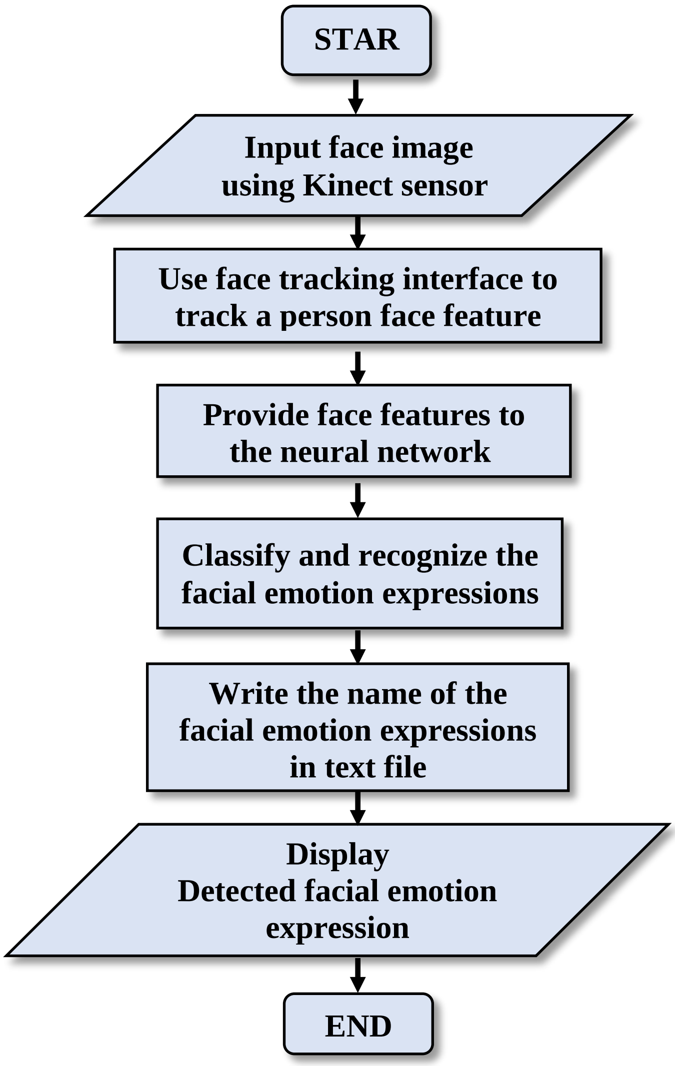


Figure 3.9. Flowchart of our proposed method

2. For step 2 we have built a Matlab engine to save the face features distances in a .MAT file and send it to the neural network.
3. For step 3, we defined a Neural Network, using MATLAB, to classify and recognize the emotion expression, based on 17 from 23 Animation Units.
4. The detected emotional expression name was written to a text file. The interface application reads the text file and displays it in real time on the screen.

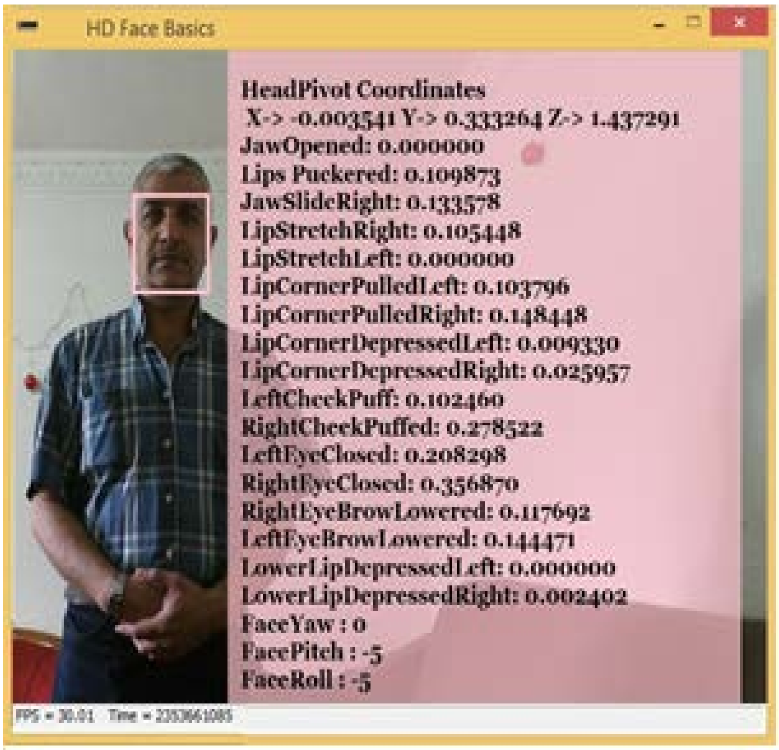


Figure 3.10. The 23 HD face animation values

Building the database

The Kinect sensor camera gives us 30 frames in one second and a set of 23 face animation unit values for each frame, all the values for the face animation units were taken from the subject and updated on each frame. 8 expressions were recorded, and 30 samples of each expression were taken, for each 14 persons. The person looks at the Kinect camera and makes any expression; also snapshots for all the 8 expressions (30 for each) were taken. The total number of frames for each expression is 240 and the total number of frames for all persons is 3360 frames. All this data were saved in .Mat files, which together represent the database.

The Neural network

After a few tests, the best neural network was decided, which gave a high performance; it was a multilayer feed-forward neural network that can be implemented using the simple NN tool present in Matlab.

70% of the database was used as training data, 15% for validation, and 15% for testing. In Matlab 2014, there is an option to save a neural network in the form of a standalone function. In this approach, NN was used to classify and recognize the expression in each frame. With 30 frames per second, it would have been extremely time-consuming if we had opted to train the NN in each frame. To find the best results for the number of hidden layers used in NN, that gives minimum mean square error (mse), a different number of layers were tested with the calculation of the mean square error for each case. The best neural network, which gives the best result was with 15 hidden layers. More layers could use, but a pretty good model was obtained at 15 layers. A standalone NN function was made, which gives us greater freedom even with 15 hidden layers, which makes the model very advanced.

System testing

The face features were provided to the neural network in a 1x23 matrix, which contains all the 23 values stated before, the first 3 values are the head pivot x y z, and last three values are the roll pitch and yaw. These 6 values are not used in expression recognition. The result of the Matlab neural network engine is passed to the 'outputs' variable, in the form of an 8x1 numeric matrix. For example, if the 8 values are: 0.0000, 0.0000, 0.0000, 0.0000, 0.1000, 0.9000, 0.0000 and 0.0000, their meaning is:

0% sure that the expression is happy
0% sure that the expression is sad
0% sure that the expression is disgust
0% sure that the expression is angry
10% sure that the expression is fear
90% sure that the expression is a surprise
0% sure that the expression is contempt
0% sure that the expression is neutral

Then, the maximum surety of the expression is "Surprise", because the highest confidence value lies at position 6, which corresponds to surprise.

Experimental results

- The system was tested on many persons, the facial features for them were stored in the database. We've got a very good result and the identification rate was more than 96%. Figure 3.11 shows the snapshot of our automatic real-time facial emotion recognition.

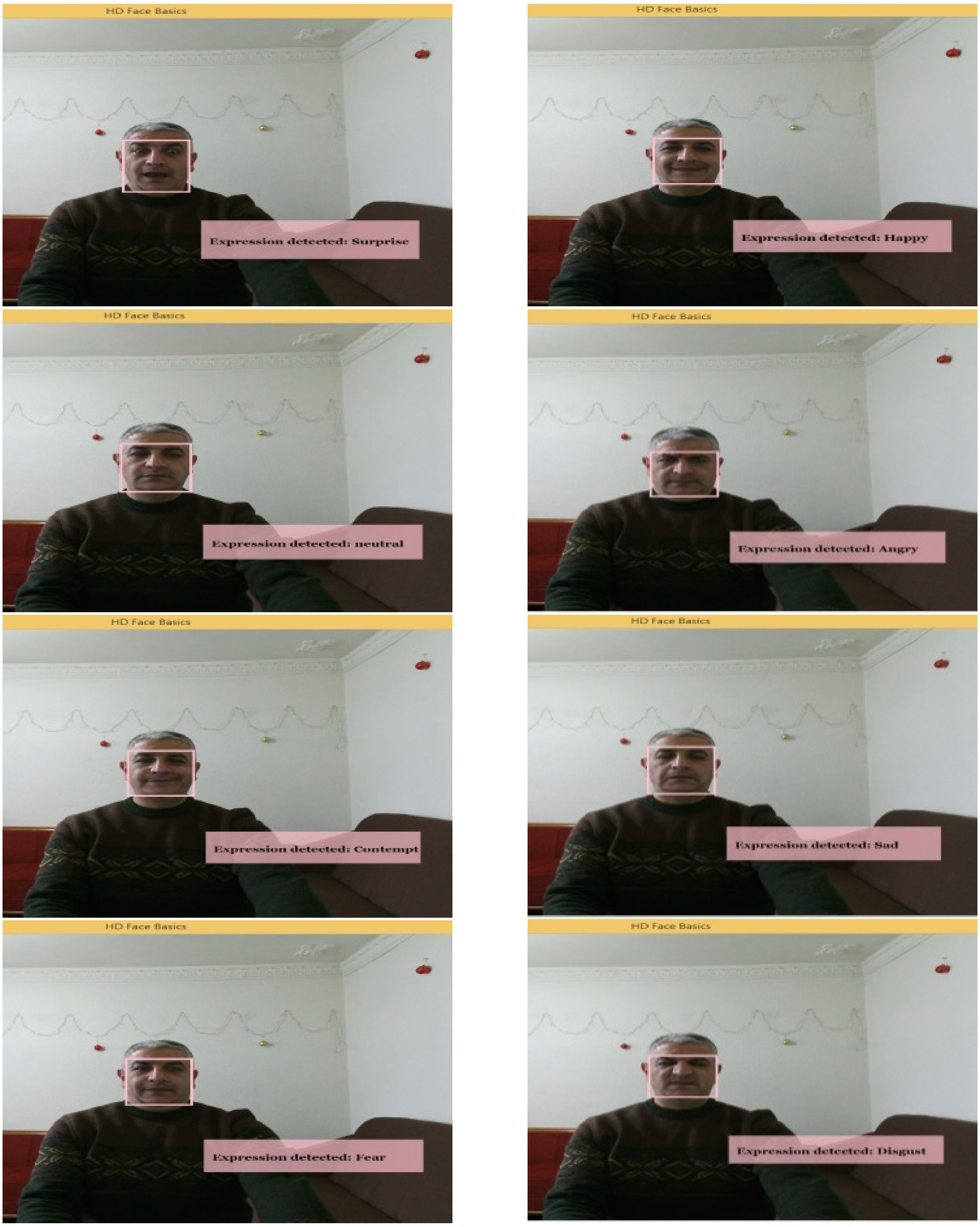


Figure 3.11. Automatic real-time facial emotion recognition; from top to bottom, left to right: surprise, happy, neutral, angry, contempt, sad, fear and disgust; the labels are written in each screenshot

- The system was tested with another model of a neural network with different layers, this didn't give the same error histogram for 15 layers which were so close to 0 (Figure 3.12).

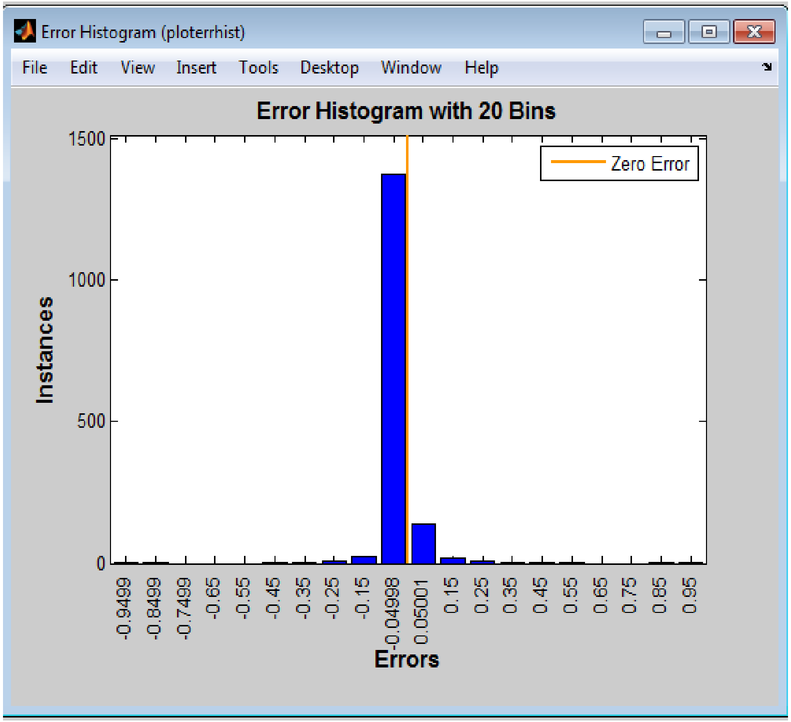


Figure 3.12. Error histogram: the error close to 0

- By using this simple NN, makes the system simple and efficient. This is a plus point for the proposal that avoided any complex NN methods and gives the mean square error equal to 0.17 (Figure 3.13). The true positive rate is very high and false positive rate is very low.
- Also, the system was tested on other persons (not from the database). We calculated the identification rate individually for each expression, by using this simple formula: *IR=1- (N/T)*, where *N* represents the number of false cases and *T* represents the number of trials in each case. The identification rate (*IR*) for all expressions was 93%.
- 17 features from the face were used to recognize eight emotion expressions to reduce the time and storage.

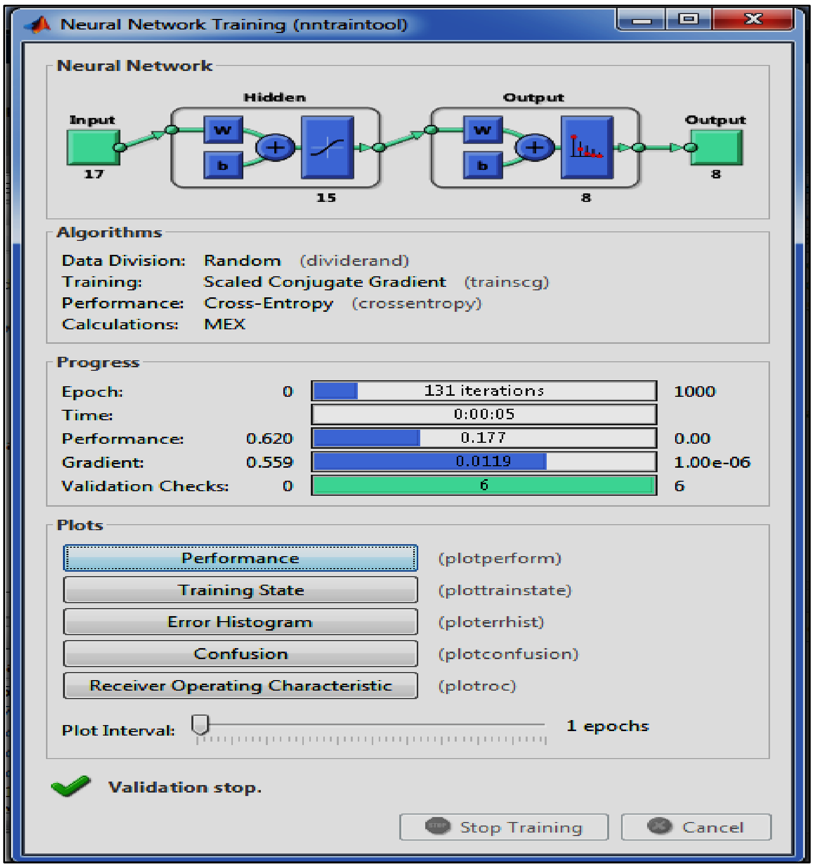


Figure 3.13. Running the used neural network

Some problems raised with Kinect sensor:

1. This kind of Kinect needs a specific requirement: a physical dual-core 3.1 GHz - or faster processor, USB 3.0 controller dedicated to the Kinect for Windows v2 sensor, 4 GB of RAM, a graphics card that supports DirectX 11, Windows 8 or 8.1, or Windows Embedded 8.
2. It is not compatible with all kinds of USB3.0 even we tried to use USB3.0 switch hub with high power.
3. Sometimes the frame per second becomes a little bit slow (24 FPS), this also happened with the older version of Kinect, V1.8.

CONCLUSIONS

In this chapter, we presented a study of human emotion expression, its definitions, the fifth main theories of emotions and a study of various approaches and existing methods for facial emotion expressions recognition.

We presented some explanations about emotion markup language (EML), some specifications and features related to face tracking capabilities of the last version of the Kinect V2 sensor and study of various approaches which used the old Kinect sensor to recognize face recognition and facial expression.

We proposed a simple psychological real time approach to recognizing eight facial emotion expressions referred by *Paul Ekman* and other scientists. The main application was implemented in Visual Studio 2013 (C++). The face of the person was acquired, tracked by the Kinect V2 sensor. The recorded features coordinate values were provided to the Matlab Engine neural network. In Matlab2014, the neural network classified and recognized the facial emotion expressions and passed back the result to the main application, which displays it on the screen**.**

This approach can be helpful to understand the facial emotion expressions for a patient who suffer from a disorder or stroke in the initial phase of his recuperation. But, these results can be extended for the monitoring of patients with other diseases of the brain, such as Alzheimer's or dementia**.** The reason of using only 17 features to recognize 8 emotion expressions was to reduce time and storage. Also, the rest six features of the head were not removed because they could be useful for next researches **.**

Some difficulties appeared when building the database: the persons should learn how to do each expression and this have taken too much time to populate the database.

CHAPTER 4. FACIAL EMOTIONS AND BRAIN ACTIVITIES

One of the important non-verbal human communication along with another form like postures and gestures is an emotional facial expression. Facial expression carries a huge amount of information which represent the emotional feelings of humanity, and by observing people’s facial expressions with its cunning and moment movements can help us to understand what the face tell us because it reflects an immediate indicator of affective dispositions in other people. All the action and the reaction in the human body is controlled by the brain, and because of the major emotional significance of facial expression, many studies have used emotional faces to identify neural substrates of emotional processing. These studies have found that brain regions commonly included in the processing of emotional information, are also activated when the facial emotion happened.

4.1 HUMAN BRAIN AND EMOTIONS

Since the 1970's, the research in the emotional field has been almost based on *Ekman* and *Wallace Friesen* theory, that, when an emotion appears on the face, a series of electrical impulses stemming from the emotion center in the brain, leads to particular facial expressions and other physiological changes, for example, increased or decreased heart rate or increase blood pressure [AH15].

The question arises, why do we have emotions? If we were happy, we relax, if we feel afraid, we try to escape from the danger, if we were disgusted, we may feel sick. Our emotion affects our behavior. Our ancestors adopted on their emotions for survival, but in these days in our life, we use our emotion to conduct the lifestyle decisions simply staying alive. Emotions make us respond to circumstances, as an example, when we feel happy this will make us smile, when we feel fear or anger this will increase the heart. We need to understand which area of the brain controls emotions. *Victoria J. Bourne* [VI10], showed that there are two contrasting hypotheses that attempt to explain how emotion perception might be organized in the brain. One suggests that all emotions are lateralized to the right hemisphere, whereas the other suggests that emotions may be different lateralized according to valence.

4.2 BRAIN PARTS ASSOCIATED WITH EMOTIONS

The human brain contains many areas that control our actions; the region of the brain that deals with showing, recognizing and controlling the reactions to emotions is known as the limbic system. The limbic system responds to sensory stimuli detected from the body surface.

The limbic system, as a result of research, is known to involve a number of structures in the control of emotions, behavior, drive, and also important to memory. Anatomically, the limbic system consists of: the subcallosal, the cingulate and par hippocampal gyri insular cortex, orbital frontal cortex, the hypothalamus, mammillary bodies and anterior thalamic nucleus. Also, it includes these pathways; the fimbria, alveus, fornix, stria terminalis and the mammillothalamic tract [SN10]. Not all of these paths are responsible for emotions, there is the anatomical description of some of the most important parts involved in emotions.

Figure 4.1 shows the limbic system. It’s also associated with pain and pleasure as well as many emotional aspects of behavior [EA13].

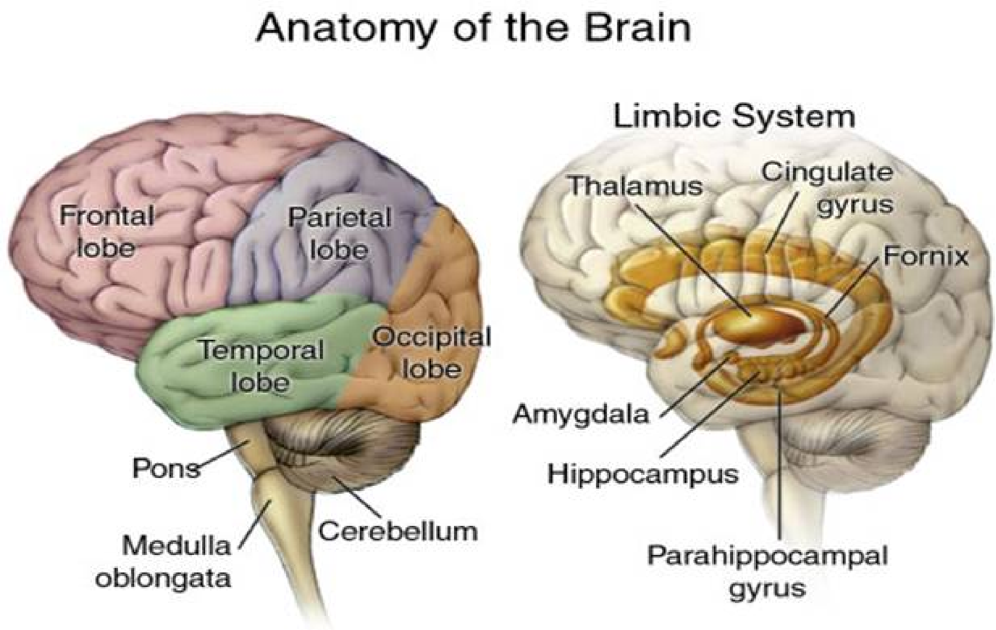


Figure 4.1. Limbic system [CO06]

.

The limbic system is the center of emotions [JU]. Limbic is a Latin word which means border. Like the familiar word “limbo”, it means an intermediate or transitional state, which is a border. In this case, the border is between the neocortex and the subcortical structures (diencephalon) [PR09]. We describe some parts of the limbic system and other brain regions that have an effect or related to emotions:

1. ***AMYGDALA*****:** "Amygdala is the integrative center for emotions, emotional behavior, and motivation" [PR09]. It takes its name from its resemblance to almond, it is located anterosuperior at the tip of the inferior horn of lateral ventricle, and it is about an inch long, it is strongly involved in emotions, memory, and motivation. It consists of many nuclei: basolateral group and cortico-medial group. Also, it is connected to other brain regions and receives messages from different body parts and organs [SN10]. In complex vertebrates, together with humans, the amygdalae perform primary roles within the formation and storage of reminiscences related to emotional events. Analysis indicates that, during the fear conditioning, sensory stimuli reach the basolateral complexes of the amygdalae, notably the lateral nuclei, where they form associations with memories of the stimuli [DA12].
2. ***INSULAR CORTEX*****:** is a group of convolutions located deep in the lateral fissure and can't be seen only after separating the borders of the fissure widely. It has a widespread connection with other brain parts. The insular cortex is split into 2 parts: the larger anterior insula and the smaller posterior insula during which quite a dozen field areas are identified [GX13]. The anterior insula processes a personality's sense of disgust each to smells [WI03] and to the sight of contamination and mutilation [WI04]. In social experience, it is involved in emotional processing [PH05]. *Damasio et al*. [DA12], proposed that this region plays a role in mapping visceral states that are associated with emotional experience, giving rise to conscious feelings.
3. ***HYPOTHALAMUS*****:** a part of the limbic system that feeds information into the amygdala [LI11]. It occupies the ventral diencephalon, it's composed of various nuclei. It's connected with the other prosencephalic areas and the mesencephalon. Lesions of the hypothalamic nuclei interfere with many vegetative functions and some of the so-called motivated behaviors, like thermal regulation, sexuality, aggressiveness, hunger, and thirst. The hypothalamus is additionally believed to play a task with emotion. Specifically, its lateral regions appear to be committed pleasure and rage, whereas the median region is like to be involved with aversion, annoyance and an inclination to uncontrollable and loud laughing. However, in general, the hypothalamus has additional to do with the expression (symptomatic manifestations) of emotions than with the genesis of the affective states. Once the physical symptoms of emotion seem, the threat they create returns, via the hypothalamus, to the limbic centers and, hence, to the pre-

frontal nuclei, increasing anxiety [PR09]. Shippensburg University states that the hypothalamus acts as a regulator of emotion, controlling levels of sexual desire, pleasure, aggression, and anger [LI11].

4. ***HIPPOCAMPUS*****:** an another region of the limbic system [LI11], it lies medially in the temporal lobe figuring the medial wall of the lateral ventricle [SN10]**.** The hippocampus contains two parts, U-shaped interconnected parts; the dentate gyrus and hippocampus proper [HR15]. It's responsible for sending information to the amygdala, also its one of the memory processing centers of the brain, the hippocampus communicates with the amygdala when a person has memories with emotional ties. The Canadian Institutes of Health Research includes that the association between the hippocampus and amygdala "may be the origin of strong emotions triggered by particular memories," which explains emotional responses to traumatic memories [LI11].
5. ***PREFRONTAL CORTEX*****:** the prefrontal cortex, situated close to the front of the head, is involved in decision making in response to emotions. The Canadian Institutes of Health Research expresses that the prefrontal cortex controls what decision a person makes when confronted with a passionate response, furthermore regulates anxiety [LI11].

6. ***FRONTAL LOBE*****:** the frontal lobe, which comprises a right and left lobe or hemisphere, is the center hub of "who you are"; emotions and personality [BR11]. Left frontal lobe manages language abilities (the logical thinker) while the right frontal lobe is generally more concerned with non-verbal aspects of communication, for example, the awareness of emotions in person’s facial expressions. The right frontal lobe is likewise accountable for picking sound signs like the tone of voice when someone is angry, sad, or scared. The right frontal lobe is more included with negative emotions while the left frontal lobe is more included in positive emotions. The left side of the frontal lobe known as the left prefrontal cortex is more active when people feel happy [BR11]. The right side of the frontal lobe –the right prefrontal cortex - is more active when people feel sad. Both fun and social interaction affect the function of the brain. Left prefrontal cortex simulates many activities like, enjoying pleasurable, doing something that seems to make time stand still, spending time with loved ones, and celebrating accomplishments. A ventral tegmental area is additionally included in emotions and adoration, especially in how a person perceives pleasure. Dopamine pathways are situated in the ventral tegmental area: dopamine is a neurotransmitter involved in mood, and expanded levels elevate the person's level of pleasure [MO13].

From the above studies, the following are the most important emotions and how they are interpreted by the brain, but it is essential to know that there is no specific or single part of the brain responsible for specific emotion, these parts work together, but some contribute more than others.

1. ***Happiness*****:** it has been found that happy and unhappy states are associated with regions of the prefrontal cortex as the left and right middle frontal gyri (LMFG & RMFG). High levels of electrical activity in the left middle frontal gyrus were found in those feeling happy and enthusiastic. Those were found to be more able to appreciate the good experience is life and more able to face the life's difficulties and make a happy life, compared to those who have a high level of electrical activity in the right middle frontal gyrus [PE13]. *Aristotle* has thought of happiness to be composed of (hedonic and eudaimonic) which are referred to as pleasure and meaning, and recently scientists used these definitions in measuring happiness. Depending on these definitions hedonic or pleasure mechanisms are found deep in the brain (nucleus accumbens ,brain stem) others in the cortex (orbitofrontal, medial frontal, insula and cingulate Gyri) these parts contain the hotspots of pleasure [MO10].

2. ***Sadness*****:** the result of a research on brain activity during transient sadness showed a diffuse activation of different parts of the limbic system, including the medial prefrontal cortex, left lateral prefrontal cortex, bilateral anterior cingulate and insular gyrus [MA95]**.** In a *PET* study of film induced emotions, sad films induced amygdala. Also, it has a role in the increasing intensity of sad facial expression. The comparison between Rt. and Lt. Amygdala showed that the left one is more activated [PH05].

3. ***Fear*****:** the main brain part associated with fear processing is the amygdala especially the lateral, central and basal nuclei in addition to the intercalated cells (ITCs). The lateral nucleus works as the input gate or sensory gate which receives information from visual, auditory, olfactory, tactile and gustatory stimuli through the thalamus and cortex. The central nucleus is believed to be the output gate; it sends projections to regions of brainstem controlling the expression of fear responses (behavioral, autonomic and endocrine). The lateral and central nuclei are interconnected directly and indirectly through the basal cell and the intercalated cell groups of the amygdala (ITCs) neurotransmitters involved in fear processing include Norepinephrine, Dopamine, Serotonin, and Acetylcholine [JA09].

4. ***Disgust*****:** functional brain imaging studies found that insula responds selectively to facial expressions of disgust in a study of the brain response to two different types of disgust (contamination and mutilation). Significant activation of the insula was achieved [PW04]. The resulting research on obsessive-compulsive disorder (OCD) showed brain activation in the insula in response to induction; also disgust induction activated other parts including the inferior frontal gyrus, the caudate nucleus in the basal ganglia, the parahippocampal region and the primary sensory cortex [NA03].
5. ***Anger*****:** the neural circuits that orchestrate anger or rage emotion runs from medial amygdaloidal areas through stria terminalis downward to the medial hypothalamus then to different parts of the periaqueductal gray matter (PAG) of midbrain. This circuit is organized hierarchically so that anger aroused in higher centers is dependent on lower center, but not vice versa, so that the rage response of the evoked amygdala will highly diminish due to lesions in the medial hypothalamus [GE11].
6. ***Surprise*****:** surprise can be defined as a reflection of the ex-ante possibility of a specific outcome, so in an experiment manipulating the degree of regret and rejoice, the regions that showed increased activity to both regrets and rejoice were designated surprise- related where the lateral orbitofrontal cortex was the most strongly activated during surprise. In another study and using fMRI the posterior parietal cortex, especially the superior parietal lobule showed a significant activation in response to surprise [OR03].
7. ***Contempt*****:** Prof. Jorge Martins de Oliveira [JU], in his research, showed that Amygdala is the part of the brain which associated with contempt emotion.

4.3 EMOTIONAL BRAIN-COMPUTER INTERFACES

BCIs systems empower the users to exchange information with the environment and control devices by using brain activity, i.e., without using the neuromuscular output pathways of the brain [WO02]. Brain signals can be acquired by invasive or non-invasive methods. In the past, electrodes were implanted directly in the brain. Later, the signal is acquired from the scalp of the user. In spite of the fact that the presence of a few techniques to acquire brain signals, the most utilized technique is the electroencephalogram (EEG) in light of the fact that it is non-invasive, compact, inexpensive, and can be utilized mostly in all environments [BR10]. In addition, low-cost and increasingly compact EEG equipment has been produced in the most recent years.

Wang et al. [WA07], *Blankertz et al.* [BL05] and many other researchers, showed that BCIs systems have been utilized in rehabilitation, speller systems, neuroscience, observing consideration systems and cognitive psychology, treatment of attention-deficit hyperactivity disorder. BCIs systems have been examined recently in recognizing emotions and are seen as a promising technique in this area because the emotions are created in the brain.

In spite of the fact that BCI models have just been produced recently, the fundamental thoughts were at that point set forward in the 1970s. There are a several difficulties in using BCIs systems for recognizing emotions, like the choice of selecting which technique to use and the channels of acquisition of brain signals which provide the best information. Many BCI's systems recently progress and future prospects regarding the emotional state of the individual as well as processing techniques so as to achieve a decent accuracy in the emotions recognition [TA13].

BRAIN-COMPUTER INTERFACE (BCI)

A brain-computer interface (BCI) is a system with an equipment (H/W) and programming communications (S/W) enabling cerebral activity alone to control computers or external devices. A brain-computer interface (BCI), also indicated as a brain machine interface (BMI), allows humans to interact with their surroundings, without the inclusion of peripheral nerves and muscles, by utilizing control signals generated from the electroencephalographic activity. BCI creates a new non-muscular channel for relaying a person's intentions to external devices, for example, computers, talk synthesizers, assistive apparatuses, and neural prostheses. That is especially attractive for people with extreme motor disabilities. Such an interface would enhance their personal life, in the meantime; reduce the cost of intensive care [LU12].

Brain-computer interface (BCI) could be a collaboration between the brain and tools that allow signals from the brain to direct some outside activity, the communication channel between the human brain and the computer is formed by utilizing electrodes to discover electric signals within the brain, which are sent to a computer, the computer then interprets these electric signals into kind of data, which are used to control a computer or tools connected to a computer.

A BCI is an artificial intelligence system that has the ability to recognize a number of certain patterns in brain signals following consecutive stages: signal acquisition, preprocessing or signal enhancement, feature extraction, translation, and the control interface

[KH09]. User received feedback reflecting the outcome of the BCI's operation, and that feedback can affect the user's subsequent intent and its expression in brain signals as shown in figure 4.2 [JR00].

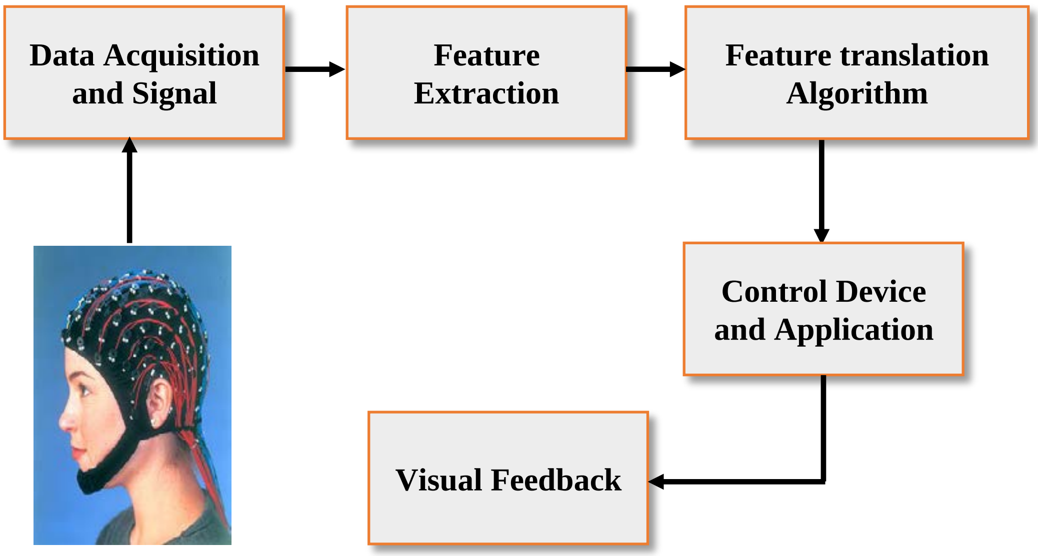


Figure 4.2. Block diagram of a BCI

The BCI's commonly contains the following stages:

a. Data Acquisition and Signal

Signal acquisition in a BCI helps within the measure of brain signals using a sensor modality. The sensor is essential, a device implanted in the brain, generally multi-electrode arrays that record the signals directly related to the movement. The signals can be amplified to levels appropriate for electronic processing. Even more, they can be subjected to filtering to eliminate electrical noise or undesirable signals. After enhancement and filtering process, the signals can be digitized and transmitted to a computer.

b. Feature Extraction

Feature extraction in Brain-Computer Interface (BCI) is the procedure of examining computerized signals to extract signal features (characteristics) and represent them in a compact form, suitable for translation into output commands. These extracted features should have good correlations with the user's intent.

c. Feature Translation Algorithm

Resulting signal features are passed to the feature translation algorithm, which changes the features into the commands to the output device (i.e., commands that achieve the users need).

d. Output Device

The commands from the feature translation algorithm operate the external device of the Brain Computer Interface (BCI), providing functions, for example, cursor control, letter selection, robotic arm operation etc. The device operation then provides feedback to the user, thus completing the closed loop of Brain-Computer Interface (BCI).

TYPES OF BRAIN-COMPUTER INTERFACES

The primary goal of BCI research is to provide communications assistance to many people with disabilities who are definitely paralyzed or 'locked in' by neurological neuromuscular disorders, for example, amyotrophic lateral sclerosis, brainstem stroke, or spinal cord injury [LU12].

BCI also extends to the fields of neurobiomimetics and complex hybrid neurogenic systems. Advances in neuroscience, computational technology, component miniaturization, biocompatibility of materials, and sensor technology have led to a much-improved workability of advantageous BCIs drift engineers, neuroscientists, effectual scientists, and behavioral and social scientists can develop as a large-scope team effort [TH08].

Many different types of brain-computer interfaces were produced. The main purpose of them is to intercept the electrical signals that pass between neurons in the brain, and then translate them to a signal that is sensed by external devices, figure 4.3 shows BCI types [AN12].

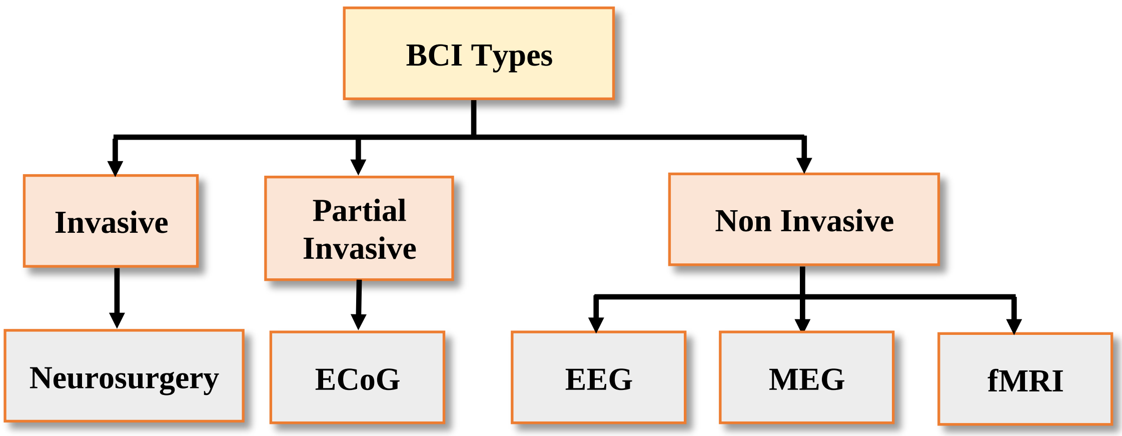


Figure 4.3. Types of BCI

A. Invasive Brain Computer Interfaces

An invasive system requires physical instruments placed directly into the brain to get the highest quality signals, making it possible to measure single neurons or every local field potentials [ER11]. Invasive Brain Computer Interface devices are used to provide functionality to paralyzed people. Invasive BCIs are also used to restore vision by connecting the brain with external cameras and to restore the use of limbs by using brain-controlled robotic arms and legs [LU12]. The main advantage of invasive BCI methods is that they produce a very strong signal than noninvasive BCI. This permits for many effective BCI systems, in terms of the user having more consistent and precise control of the external mechanisms of the BCI system. Invasive BCI systems have some obvious disadvantages because it must be implanted surgically, which can introduce risks associated with any type of surgery. Scar tissue fabricate can even cause the moderately stronger signals to weaken or be lost altogether [RI13]. Much of the current neural stimulation research is based on invasive probes [LA12].

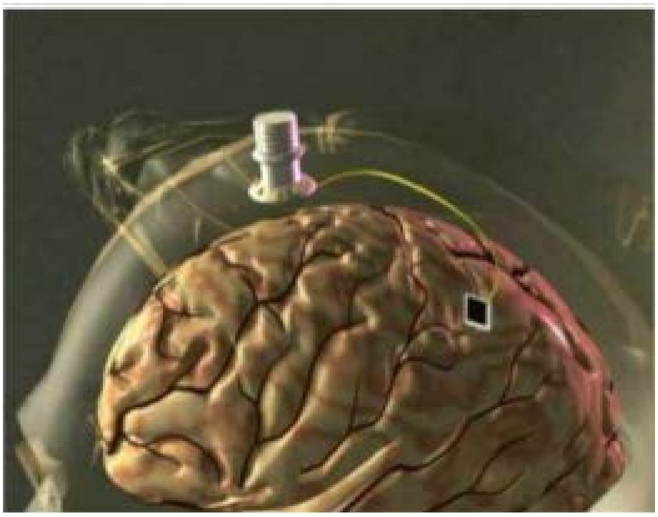


Figure 4.4. Invasive physical devices implanted directly into the brain [ME14]

B. Partially Invasive Brain Computer Interfaces

Partially invasive BCI system interface devices are neuroprosthetics that are implanted on a permanent basis within the skull itself, but only onto the surface of the brain. They spread out electrode arrays over the surface instead of burrowing inside. The benefit of going under the skull is that the skull acts as a significant brainwave dampener. Also, not cutting into a living brain. The accuracy of the signal strength is better than none- invasive systems, but not as high as invasive systems and they produce better resolution signals than non-invasive BCIs. Also, it has less risk of scar tissue formation when compared to invasive BCI [SC08]. One of partially invasive brain computer interfaces is ***Electrocorticography (ECoG)***, it measures the electrical activity of the brain taken from beneath the skull. The partially invasive BCI electrodes are embedded in a thin plastic pad that is placed above the cortex, underneath the dura mater (the thick membrane that encompasses the brain and spinal cord) [SE03]. Nowadays, research on light reactive imaging BCI devices is taking place. This BCI device, for the most part, includes embedding a laser inside the skull. Researchers of Carleton University in Canada, believed that the same interface of light reactive imaging BCI could form the basis of a brain-controlled password system [AN12].

Figure 4.5. Partially invasive physical devices implanted on a permanent basis within the skull directly into the brain [ME14]

C. Non-Invasive Brain-Computer Interfaces

A kind of interface used when only a temporary connection to the brain is required. It is the least accurate of the neuroprosthetic techniques, dealing with general brainwaves that are dampened by passing through the skull. However, it is sufficiently delicate to perform general tasks and gather non-particular information.

Non- invasive brain-computer interface has the minimum signal clarity with regards to communicating with the brain (skull distorts signal); however, it is thought to be extremely most safety when compared to other kinds. This kind of device has been found to achieve success in giving a patient the ability to move muscles implants and restore partial movement [RI13].

The technique for this system contains a medical scanning device or sensors mounted on caps or headbands, which help to scan signals from the brain. In this system, because the electrodes can't be placed specifically on the wanted part of the brain, its signals clarity are less than of the invasive system. EEG or electroencephalography is one of the most popular kinds of these systems [AN12].

Non-invasive neuroprosthetics typically has the longest training curve, because they detect brainwaves at the point outside the skull, where they have run together, and greatly weakened [FG08].

There are many kinds of non-invasive BCI systems:

1. ***Electroencephalography (EEG).*** EEG is a very simple type of non-invasive BCI. In the 1980s, *Lawrence Farwell and Emanuel Donchin* [FA88], developed an EEG-based brain-computer interface. The system works via an array of electrodes adhesive to the scalp or grasped in place by a cranial device. It is easy to use, cheap, portable and produces relatively good general brain readings. Generally, in a BCI system, EEG signals are preprocessed and extracted feature that they represent the best details of the signal are used to train the classifier which discriminates the features [SN12].

   Various methods have been presented to design a BCI system based on EEG signals, such as event-related synchronization [LL05] and event-related desynchronization [LI03]. Some methods use emotion recognition with brain activity using EEG signals and the analysis based on various emotions such as happiness, surprise, and anger. Many EEG-based BCI systems have been developed recently in which patterns of EEG in different mental states can be discriminated for information transmitted by feature extraction and classification algorithms [DC10] [PC10] [SN11] [HP12].

   Currently, correct EEG-based recognition of artificially evoked emotion is only about 60%, but much research shows the suitability of EEG for this kind of task [WE09]. This field of research is still relatively new, *Hyunjin Yoon et al*. [HY13], presented emotion recognition of serious game players using a simple BCI, and there is still much to be done to improve on existing elements in BCI, but also to discover new possibilities.

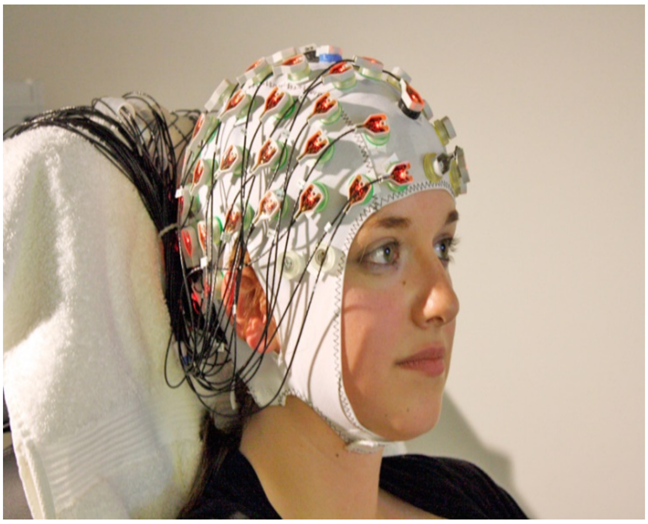


Figure 4.6. Non- invasive EEG scanning devices are mounted on caps or headbands [ME14]

2. ***Magnetoencephalography (MEG)*** is a non-invasive neuroprosthetics technique which functions by measuring changes in the tiny electromagnetic field that surrounds the brain. MEG detects the little magnetic fields created as individual neurons "fire" inside the brain. It can pinpoint the active region with a millimeter, and can follow the movement of brain activity as it goes from region to region in the brain. The electromagnetic field in MEG is weak compared to the normal background noise. Therefore, to minimize this ratio, a special highly sensitive detector called “superconducting quantum interference devices” or “SQUIDS” are used. Moreover, they can only detect the field itself, special algorithms are used to determine the rough location in the brain.

   MEG is a rapidly developing and unique tool for the study of brain function [PR14]. The researchers used MEG combined with other non-invasive to decode or classify brain representations of external events, emotional states, and movement plans [TH08].

   MEG provides the most information about brain activity after some time and in addition, connections between cortical regions. The MEG let us know about brain activity as EEG, indicating the activity of neural networks in real time; however, it gives more information than the EEG about deeper structures. Coherence examination of EEG or MEG tells which parts of the brain are associated with each other by analyzing similarities in brain activity patterns. Combining information from these and different sources provides an additional complete portrait of brain functioning than has ever been possible [AN12]. More related research can be found in [JS05] [EM08] [EM09] [DP10] [WH11] [JZ11] [GU11] [NE12].

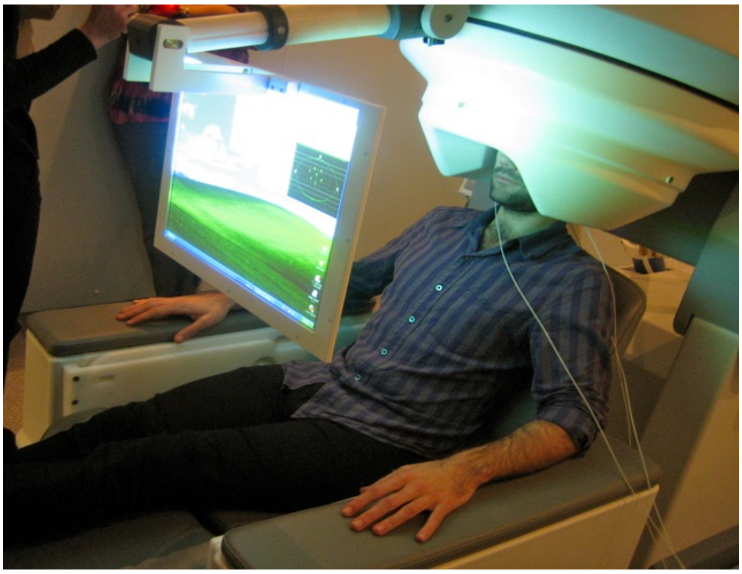


Figure 4.7. Non- invasive MEG system [ME14]

3. ***Functional* Magnetic resonance imaging *(fMRI)*** is a relatively young brain activity imaging technology. It supports detection of the dynamic regulation of blood flow within the brain. Medically, this can be termed the hemodynamic response; however, it is basically tracking brain activity based on increasing and decreasing demand for oxygen and glucose in the hemoglobin of the blood in the brain [BR11].

Functional MRI (fMRI) presents information about the location of major brain activity during a behavior, including not only in the cortex but also structures further down in the brain. The operator can see how and where the brain activates in response to learning new information, recognizing a face instead of simply seeing a face, or learning new languages.

Also, permit us to figure out precisely what changes in the brain when a few types of information are learned, or when we perform distinctive tasks, for example, talking, or when we are sick. Unlike detection of electrical signals, it can be performed non-invasively with practically no loss of data. However, the outcome doesn't track in real-time, with fMRI data sometimes taking as much as a minute to determine. This delay can abundance the results, on everything except long-term cognitive activity [AE09]. Further research can be found in [HT04] [RA07] [TH08] [SE13].

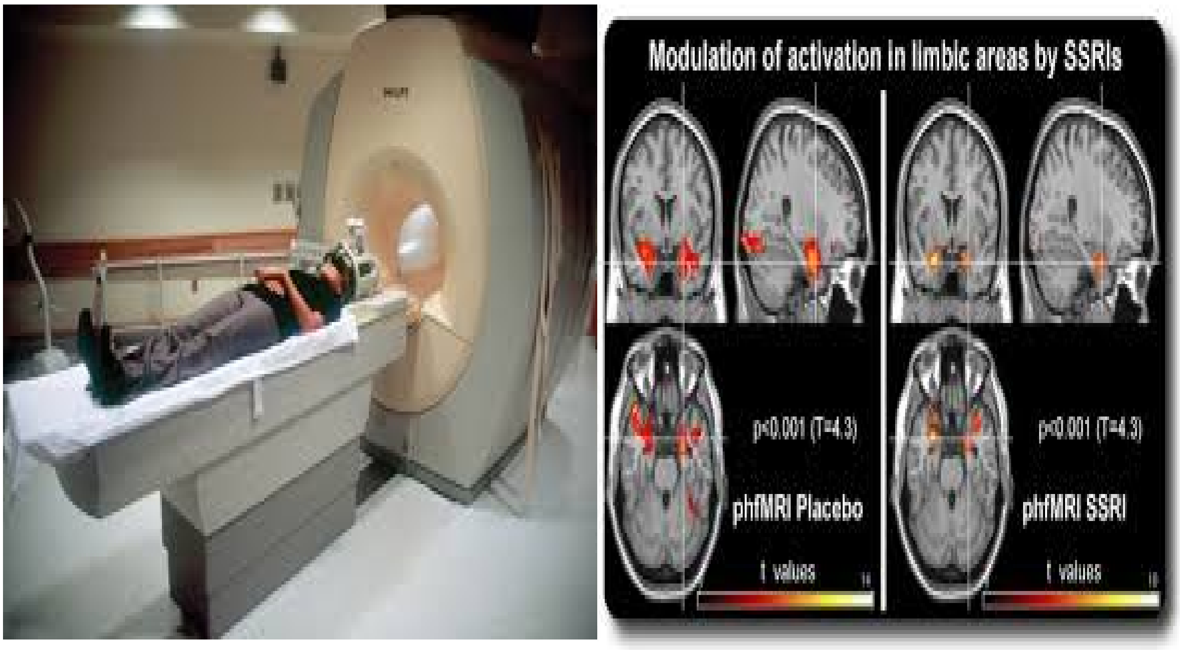


Figure 4.8. The fMRI system with its modulation of limbic areas [ME14]

One of the BCI system applications is face recognition and facial emotional expression recognition. Researchers used EEG, MEG, AND fMRI technologies or combinations between them to have the best results.

Panagiotis C. and Leontios J. [PA12], in their research presented a new feature extraction methodology for a user-independent emotion recognition system, namely, HAF-HOC, by using electroencephalograms (EEGs). They developed a new filtering procedure, namely, Hybrid Adaptive Filtering (HAF), for an effective extraction of the emotion-related EEG-characteristics by applying genetic algorithms to the empirical mode decomposition-based illustration of EEG signals. The presented HAF-HOC scheme incorporated 4 distinctive classification strategies to achieve powerful emotion recognition performance, through a continuum of facial expression image projection, as a Mirror Neuron System-based emotion elicitation process. EEG data associated with six basic emotions (happiness, surprise, anger, fear, disgust, and sadness) have been acquired from sixteen healthy subjects utilizing three EEG channels.

Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu [XI14], in their research, presented an emotion recognition system supported electroencephalogram (EEG) signals. They experimented using movie elicitation were designed for acquiring subject's EEG signals to classify four emotional states, joy, relax, sad, and fear. After pre-processing the EEG signals, they investigated different kinds of EEG features to produce an emotion recognition system. For classification, they used multilayer perceptron and support vector machines and for the evaluation for the classifier, they used the k-nearest neighbor (*k*NN) algorithm. For extracting common critical features across subjects, they used a minimum redundancy maximum relevance. Their results show that an average test accuracy of 66.51% of classifying four emotional states can be obtained by using frequency domain features and support vector machines.

John M. Allman, Atiya Hakeem, Joseph M. Erwin, Esther Nimchinsky and Patrick Hof [JO06], proposed that the anterior cingulate cortex is a specialization of neocortex rather than a more primitive stage of cortical evolution. The evidence from single-neuron recording, electrical stimulation, EEG, PET, fMRI, and lesion studies indicates that the anterior cingulate cortex has an important role in emotional self-control as well as focused problem-solving, error recognition, and adaptive response to changing conditions. These functions are central to intelligent behavior. They are juxtaposed in this structure and probably are intimately interconnected.

Rafael Ramirez [RR12], in his work, demonstrated that the investigation of emotions in human-computer interaction expanded lately, trying to address new client needs. In the meantime, it is conceivable to record brain activity in real-time and find patterns to relate it to emotional states. He described a machine learning approach to detect emotion from brain activity, recorded as electroencephalograph (EEG) with the Emotive Epoc device, during auditory stimulation. The features were separated and extracted from the EEG signals in order to characterize states of mind in the arousal-valence 2D emotion model. He used these features and applied machine learning techniques to classify EEG signals into high/low arousal and positive/negative valence emotional states. The used classifiers, categorize emotions such as happiness, anger, sadness, and calm based on EEG data.

4.4 OUR PROPOSED APPROACH TO RECOGNIZE BRAIN ACTIVITY BASED ON FACIAL EMOTION EXPRESSIONS

To find out the brain activity that goes on behind the scenes, during the facial expressions, we have taken into account the information and studies that we have introduced in section 4.2 and 4.3, related to the brain parts that will be active when the emotions occurred and the computer brain interface used to recognize emotion, in addition to many researches which used EEG and fMRI [HU05] [HA10] [MA11] [MI11] [JA14].

The physiological signals have the benefit that they can hardly be deceived by voluntary control and are available all the time, without requiring any further action of the user. The insufficiency of utilizing these signals is that the user needs to wear some measurement equipment which could be extremely straightforward when measuring the heartbeat, but EEG and fMRI measuring devices have a tendency to be more requesting with a very high cost.

We have associated for each facial emotion expression a specific brain area which is active when this emotion occurred. This approach can be used with the patients suffering from a stroke, and with another disease like Alzheimer in the initial phase of their recuperation, by checking if the patient recognizes the visitor or not.

Table 4.1. presents the associations between facial emotion expressions and brain activity that we used in our experiments.

Table 4.1. Emotions and its associated active brain area

Emotion	**Active brain area**
Neutral	Dorsal Anterior Cingulate
Happiness	Left middle frontal gyrus / nucleus accumbens
Sadness	Prefrontal cortex / Amygdala
Angry	Amygdala /media hypothalamus / PAG
Disgust	Insular Cortex (right frontal)
Contempt	Insular Cortex / Amygdala
Fear	Amygdala (lateral, central and basal nuclei)
Surprise	Posterior parietal cortex, lateral orbitofrontal Cortex

To get brain activity information using facial expressions, we also incorporate this information in our system presented chapter 3, so that the system tells us:

1. Which emotion our patient is experiencing?
2. Which center in his/her brain is active corresponding to this emotion?

The hippocampus deals with the formation of long-term memories and spatial navigation. In diseases such as Alzheimer's, the hippocampus is one of the first regions of the brain to become damaged and this leads to the memory loss and disorientation associated with the condition [AN]. The hippocampus is a small region of the brain, which activates when a person recognizes someone.

However, if a person does not recognize anyone, any of the 3 cases will arise:

1. The person face will remain expressionless. For example, if a visitor approaching the person, his / her (person) face will not showing any emotion (neutral), so the dorsal anterior cingulate will be active.
2. Or the person will look expectantly at the approaching visitor, with no facial expression other than the raising of the inner eyebrow. In this case frontalis muscle (pars medialis) will be contracted.
3. Or if the person wants to be polite and attempts a smile towards the approaching visitor, the smile will not be a heartfelt one (it will be a voluntary smile). It will be due to a contraction of the zygomatic major muscle alone. Figure 4.9 shows the screenshots of the developed system.

In conclusion, even when there is a facial expression like, smiling, this does not necessitate that an emotion had occurred, where in fact there is no brain activity , the smiling was the result of the muscle contraction in order to be polite. Meaning that the person didn't recognize the approaching person.

Figure 4.9. Automatic real-time facial emotion expressions and brain activity recognition, from top to bottom, left to right: neutral, happy, angry, fear, disgust, contempt, sad and surprise.

CONCLUSIONS

In this chapter, we presented a study of human brain and emotions, we described some parts of the limbic system and other brain regions that related to emotions. Also, we presented different brain-computer interfaces (BCI's), which can be used for emotions recognition, and some approaches based on these technologies.

We are not psychologists nor doctors, but from these studies, we tried to find the relation between the brain regions and each emotion, so we build a table that associates an active brain region to each specific 8 emotions and included it in the system for facial emotion expressions recognition.

The system can be used with patients who suffered from a stroke or Alzheimer's to analyze their behavior and if they recognize their visitors or not.

CHAPTER 5. HUMAN MOTION TRACKING APPROACHES AND TECHNOLOGIES

The subject of human motion capturing and tracking received a considerable attention over the last decades. There is a wide range of applications that use human motion tracking while the industry provides a novel constantly movement in the area tracking systems, which have a great accuracy and high performance. Human motion tracking is getting expanding consideration from computer vision scientists. This consideration is motivated by wide uses, for example, athletic performance analysis, man–machine interfaces, video conferencing and others.

5.1 MOTION TRACKING AND ANALYSIS SYSTEMS

Motion tracking began as pictorial examination instruments in the field of biomechanics exploration in the 1970s and 1980s, the expansion was in education, games, sports training. Nowadays, with the growth of technologies, motion tracking is used in various fields, such as computer vision, animation, rehabilitation, and virtual reality applications. At the beginning of the 20th century, the individual needed to wear markers (sign) near every joint to determine the positions of the movements or angles between the markers. Acoustic, inertial, LED, magnetic signs, or a combination of any of these, are tracked, optimally at least twice times the frequency rate of the required action. A human motion tracking system has a motion capture (mocap) process which can be defined, as “a process for recording the movement of objects or people”. This process can be used in various fields like, sports therapists, medical rehabilitation applications, military applications, video games, and for robotics [DN0] [JC11].

Motion capture systems research started from a long time. A survey of vision-based technologies can be found in [TB06] [RB07]. At the beginning of the 21th century, given the rapid growth in new technology, new methods have been developed. The most modern systems extract the shape or the graph of the performer from the background. All joint orientations or joint angles are computed by implementing them in a mathematical formula that can be used for analysis and comparison.

Human motion tracking methods used devices that permit the tracking of many motions like waving, standing, sitting, walking, running and even facial movements. There is a wide range of methods used for capturing human motion in virtual environments, which utilized optical motion capture systems [VI]; other systems used combined multiple sensor types [DV07]. Motion capture systems are used to capture human motions on computers with the aim of translating these motions to animated or simulated characters [WI].

Nowadays, almost all capture and track systems require specific hardware and software, the costs of which are expensive. In addition, these systems need long times and complicated steps to install [KY13]. Figure 5.1 represents the various kinds of motion capture systems.

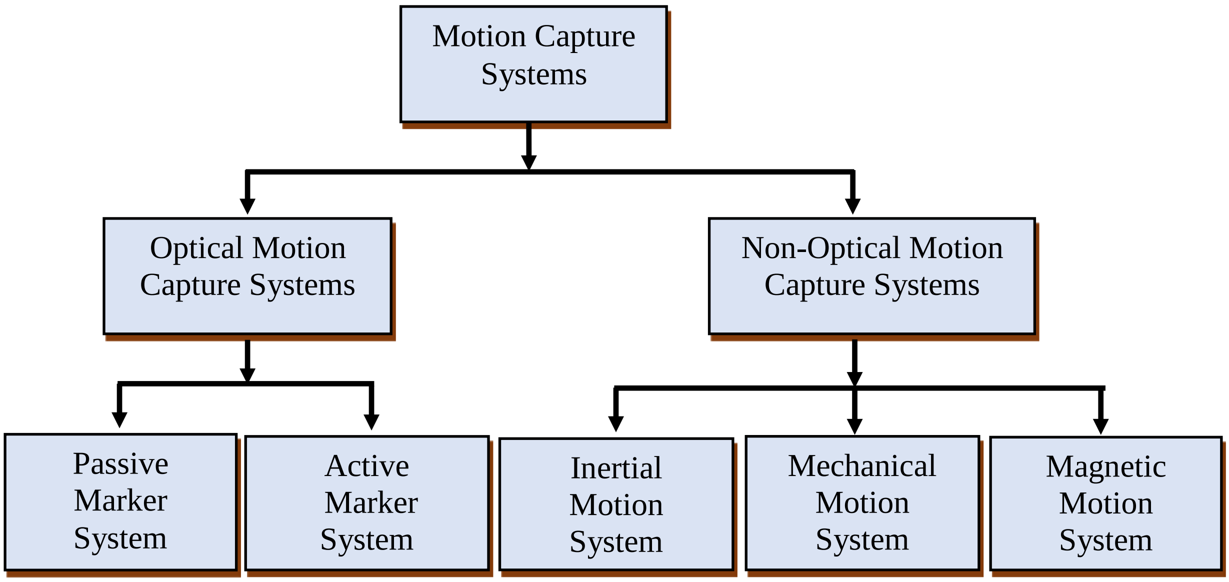


Figure 5.1. Classification of motion capture systems

5.1.1 Optical Marker systems

The majority of modern motion capture systems are optical, because the subjects are “watched” by cameras.

With these systems, the performer wears a particularly designed suit, covered with reflectors that are placed in their main articulations. At that point, high-resolution cameras are strategically positioned to track those reflectors during the performer’s movement. Each camera produces the 2D coordinates for every reflector, obtained by mean of a segmentation step. The software is then used for analyzing the data captured by the cameras to generate 3D coordinates of the reflectors. These systems are the most costly, because of to their cutting-end technological nature, for example, the high-resolution cameras and sophisticated proprietary software [PN11].

The benefits of using these systems are higher sampling rate, which empowers the capture of quick movements, for example, martial arts, acrobatic exhibition, and gymnastics, and others. Another benefit is the flexibility offered by these systems, there are no wires or restricted workspace, also, the reflectors pose no restraint or cumbersome effect on the performer.

The disadvantages of this systems are the occlusion of a few transmitters, particularly in small objects, for example, hands or intently interacting objects, an issue that can compromise the entire process if the occluded data is unrecoverable. Another issue is the absence of interactivity since the data gathered must be processed (and sometimes undergo filtering and noise reduction) before it is usable [LD09].

Today, these techniques are widely accepted and are used by researchers for motion capture and record in all over the world.

Figure 5.2. Optical motion cameras emitting Infrared Radiations [OP15]

1. PASSIVE (REFLECTIVE) MARKER SYSTEMS

These systems use markers coated with retroreflective materials that reflect the light back to the cameras, that must first be calibrated so just the markers are distinguished and ignoring other materials. Through using many cameras, they calibrate an object with reflectors for which the positions of these are best known. Waving a “wand” imbued with a number of reflectors, over the catch volume, is the means by which this calibration procedure is normally performed.

Regularly, a system will fuse anywhere in the range of 6 to 24 cameras, yet a few systems with more than 300 cameras exist to reduce marker swap or confusion issues in complex catches. It is the most adaptable and basic technique utilized as a part of the industry [PN11].

Figure 5.3. A performer wearing passive marker [PN11]

2. ACTIVE MARKERS

The active marker is an optical real-time body positioning and movement tracking technique whereby a network of LEDs everywhere throughout the body, situated along bones, and between joints, are illuminated up in a steady progression, at a rate excessively quick for the human eye to capture [VI15]. Active marker 3D motion tracking/capture systems are a trendy technology, it has a high resolution, speed, and capture range requirements. This kind of system uses LED'S for capturing instead of reflecting the light emitted by the high-resolution cameras, which transmit their own particular light, being powered by a small battery.

The power to each marker can be provided sequentially in phase with the capture system providing a unique identification of each marker for a given capture frame at a cost to the resultant frame rate. This ability is useful in real-time applications [ME]. PhaseSpace and Phoenix technologies are the major players in this category and normally provide active markers of both these types [PH].

Figure 5.4. WPI's motion capture suit when activated [ME]

a. TIME MODULATED ACTIVE MARKER

Time modulated active marker MoCap or motion capture, is actually simply a refinement to active marker MoCap. Time modulation, rather than doing each one at a time, strobes the LEDs. A few are turned in the meantime, but each flashes at different frequency rate, permitting them to be individually ID. This type of unique marking radically will increase the resulting capture frequency as instead of cycling through markers each one at a time, they're done in staggered simultaneous batches [VI15].

The trade-off is, obviously, a far higher process overhead than the standard active marker. Systems with 12-megapixel spatial resolution modulated indicate more subtle movements than 4 megapixel, optical systems by having both higher spatial and temporal resolution [OP15].

Computer processing permits fewer hands cleanup or filtered results for lower operational prices. This higher precision and resolution requires more processing than passive technologies, however the extra processing is done at the camera to enhance resolution through a sub-pixel or centroid processing, giving both high resolution and high speed. This motion capture system is generally under $50,000 for an eight camera megapixel spatial resolution 480-hertz system with one on-screen character [MH12].

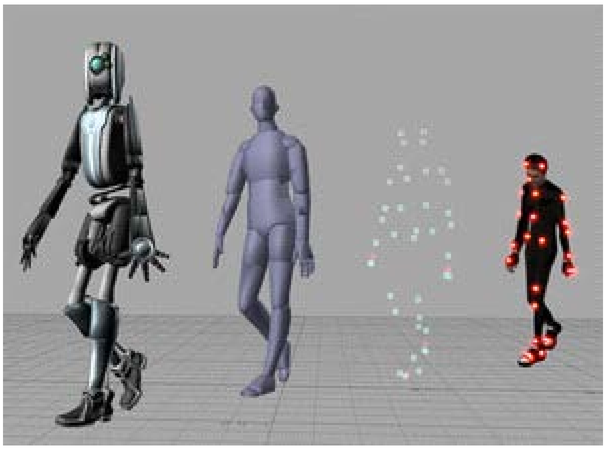


Figure 5.5. Showing different stages of motion capture technology [MH12]

b. SEMI-PASSIVE IMPERCEPTIBLE MARKER

These systems use cheap multi-LED fast projectors that optically encode the space. The system utilizes photosensitive marker labels to decode the optical signals. By attaching labels with photograph sensors to scene points, the labels can compute their own locations of every point, as well as their own orientation, incident illumination, and reflectance.

These tracking labels that work in natural lighting conditions can be vaguely implanted in clothing or different items. The system supports a boundless number of labels in a scene, with each label uniquely identified to wipe out marker acquisition issues. The labels additionally provide incident illumination data, which can be utilized to match scene lighting when inserting synthetic elements. The technique is, in this manner, perfect for on-set motion capture or real-time broadcasting of virtual sets [OP15].

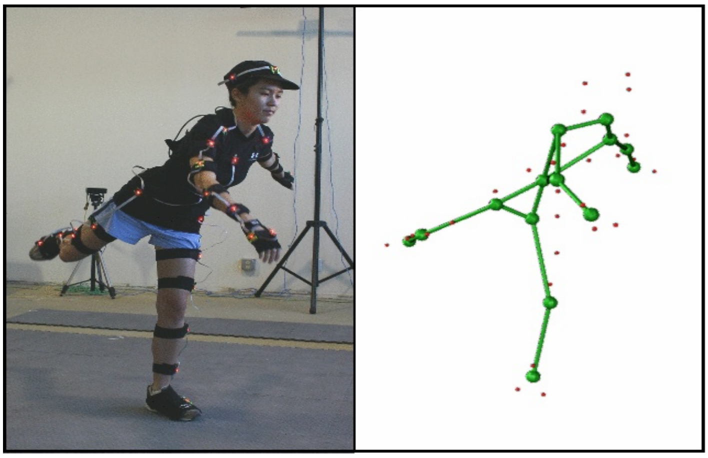


Figure 5.6. Semi-passive imperceptible marker [CO15]

From the previous presentations of capturing / tracking systems, the tracking computer is responsible for capturing and tracking images, then analyzing these images to extract target position and controlling the mechanical tracking stage to follow the target.

Several difficulties arise. Firstly, the tracking computer must have the ability to capture the image at a relatively high frame rate, this post a necessity on the transfer speed of the picture catching equipment. Secondly, the system must have an image processing program with the ability to separate the target image from its experience and compute its position.

Several algorithms for image processing were intended for this assignment, however, each has its own particular constraints; this issue can be improved if the tracking system can expect certain attributes that are regular in every one of the targets that will track. Another issue is controlling the tracking stage to follow the target; this is a typical control system design problem rather than a challenge, which includes modeling the system dynamics and designing controllers to manage it. This turn into a challenge in highly dynamic applications, if the tracking stage is not designed for real-time, in which case the tracking software has to compensate for the mechanical and software imperfections of the tracking stage. Software that runs such systems is customized for the relating equipment components. One example of such software is Optic Tracker, which controls computerized telescopes to track moving objects at great distances, such as planes and satellites.

5.1.2 NON-OPTICAL MARKER SYSTEMS

Non-optical marker systems are classified into three types:

1. **Mechanical Motion capture system**

This system uses exoskeleton technology to capture the motion, every joint connected to an angular encoder. The computer record the value of movement of each encoder (position, rotation, etc.), knows the relative position encoders (joints) and have the ability to rebuild these movements on the screen [ME]. Mechanical motion capture systems are made out of potentiometers and sliders that are placed in the desired articulations and empower the showcase of their positions (Figure 5.7). These systems have few benefits points that make them quite attractive. One of its benefits is that they possess an interface that is similar to stop-motion systems that are very popular and used in the film industry, thus permitting an easy transition between the two technologies. The other benefit is that they are not influenced by attractive fields or undesirable reflections, not requiring a long recalibration process, which makes their utilization simple and gainful [PN11].

This technique offers high accuracy, the capture is restricted by mechanical imperatives identified with the execution of the encoders and the exoskeleton. The exoskeleton uses wired associations with the interface to connect the encoders to the computer. The flexibility is somewhat constrained because the exoskeleton is heavy. Something else, the exactness of reproduction of the movement relies upon the position encoders and modeling of the skeleton while the big disadvantage comes from the coders themselves because if they are of great precision between them it cannot move the object to capture in a so genuine. At long last, to animate each object, it requires an exoskeleton over it, is quite complicated to measure the interaction of many exoskeletons [MM].

Figure 5.7. Mechanical motion capture using Exoskeleton [BL01]

2. Magnetic motion capture

This technology uses an arrangement of receptors that are placed in the actor’s articulations; it’s possible to measure the position and orientation of the articulations relative to an antenna. Magnetic motion capture systems using this technology is not extremely costly compared to other systems for motion capture. The workstation utilized for data acquisition and processing is cheap as well and the accuracy of data is very high. With an average testing rate of around 100 frames per second, magnetic systems are ideal for simple movement capture.

The weaknesses of these systems include that it has a large number of cables that connect to the antenna, which decrease the flexibility degrees of the actor. Some systems that do not require the utilization of cables are at present being modified which effectively eliminated this weakness [HK09]. The possible interference in the magnetic field caused by different metallic objects and structures represents a restriction to the surrounding material, which can be of some gravity. Some systems are exceedingly sensitive even to the building's own structure showing some interference, making this a critical flaw in magnetic mocap systems. These days, an exertion by these product manufacturers, which are investing in new and improved versions of their systems that exhibit a decreased powerlessness to these issues [LD09].

Figure 5.8 Magnetic field transmitter source [SO]

3. Inertial Motion Capture

This technique does not require cameras except as a localization instrument. Inertial sensors are worn by the subject and the data from the sensors is transmitted wirelessly to a computer. Measurement sensors, for example, accelerometers and gyroscopes are ordinarily applied for motion tracking. The motion data from these sensors is detected remotely by a computer software system and recorded. Tracking movement utilizing inertial sensors can be troublesome as the data recorded can be ambiguous and this is the reason it is ideal to create models of human motion as an essential for being able to obtain the most precise readings on human movement. The combination of both accelerometers and gyroscopes to measure movement is controlled by applying an algorithm. The gyroscope in this movement sensor network measures the orientation of the sensor, data that reflects the gravitational acceleration. The accelerometer can calculate the initial position by subtracting the gravitational acceleration from the sensor frame [AZ15] [HO10].

5.1.3 MARKERLESS SYSTEMS

Markerless system utilizes a technique which doesn't require markers to be worn and instead relies on software to track the subjects' movement. For motion capturing, a wide range of software and camera systems that permit the user to capture and analyze motion without the need to use of trackers. These systems are suitable for biomechanical analyzes to be performed on animals. However, the accuracy of the mark less tracking devices depends on the accuracy of selecting the points on each frame of the video, the efficiency of the automatic tracking algorithm implemented on the system and various other factors.

Emerging techniques and research in computer vision are leading to the rapid development of the markerless approach to motion capture. Markerless systems, for example, those modified at the Stanford University of Maryland, MIT and Max Plan Institute do not require subjects to wear special equipment for tracking. Among the very few companies, Noraxon-Inc [NO] and Organic Motion [OR] are the commercially successful systems [HO10].

Figure 5.9. Markerless motion capture system by organic motion [HO10]

5.1.4 3D MOTION CAPTURES SCANNERS

A 3D scanner creates a point cloud data of geometric samples on the surface of the subject. These points can then be utilized to extrapolate the shape of the object. The colors on the surface of the subject can be determined if color information is collected at each point.

3D scanners digitize real objects, which can then be utilized for a virtual examination and visual applications like facial motion capture. This technology is currently being used to drive facial animation in real-time to provide user feedback, by using the features of the face, for example, the nostrils, the corners of the lips and eyes, and wrinkles and then track them [AH15].

Vision-based approaches additionally can be able to track human movement, eyelids, tooth occlusion on the lips and tongue, which are a clear problem in computer animated features. Typical limitations of vision based approaches are resolution and frame rate, both of which are decreasing as issues as high-speed, high-resolution.

CMOS cameras become available from multiple sources, structured light scanners like PONTOS from G-O-M [GO] (Gesellschaft für Optische Messtechnik) are used to capture exact position, motion and deformation calculation of structures and components. For a more detailed survey of digital 3D scanners, readers are referred to Kannan [SK08].

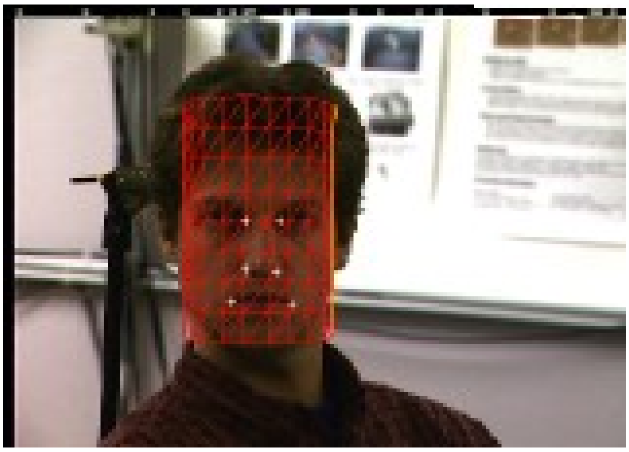


Figure 5.10. 3D Head motion capture by a 3D scanner [CA]

5.2 HUMAN MOTION TRACKING USING KINECT

Most of the motion capture systems presented before require quantity of time, effort, and cost much money related assets to realize the motion capture processes that are considerable and generally not extremely possible on a large scale.

Reducing the expense and effort for precise motion capture was a major accentuation in related research. The need for a very meticulous setup was paralyzed mainly by the lack of depth sensor feasible technologies, allowing for translating three-dimensional space to the computer in real time. Further research was also trying to eliminate the need for multiple camera setups and the need for specialized markers on the actors.

The research was additionally attempting to take out the requirement for different camera set-up and the requirement for particular markers on the performers. Energizing new advancements includes the Xbox Kinect system from Microsoft, which, by its style, may motivate client movement to vitality consumption levels, which will extra intently emulate daily exercise recommendations [ER06].

Kinect sensor with a dual-camera allowed for three-dimensional body tracking without the need for multiple cameras or a meticulous marker setup for the users. The unique functionality of the Kinect sensor, coupled with a fast and uncomplicated setup, an easy development framework, and access to a lot of online learning resources made a huge impact in the research community.

Microsoft Kinect V2 sensor with its Software Development Kit (SDK) is an improvement of technology from older versions [MI14], a developed motion-sensing device that provides the users a facility to interact with computers and game consoles through many ways like gestures or spoken commands. This technology allowed many researchers and companies to develop beyond the original scope of gaming, many real-time applications in various fields like healthcare, sport training, facial emotion detection, airport security, law enforcement, three-dimensional reconstruction, motion recognition, and even more, like sign language recognition, robotic control, voice and gesture recognition, as well as 3D reconstruction, wonderful for the 3D printing [BO11] [RA12] [RL15]. Figure 5.11 shows the major application domains that benefit from the Kinect sensor device’s technology.

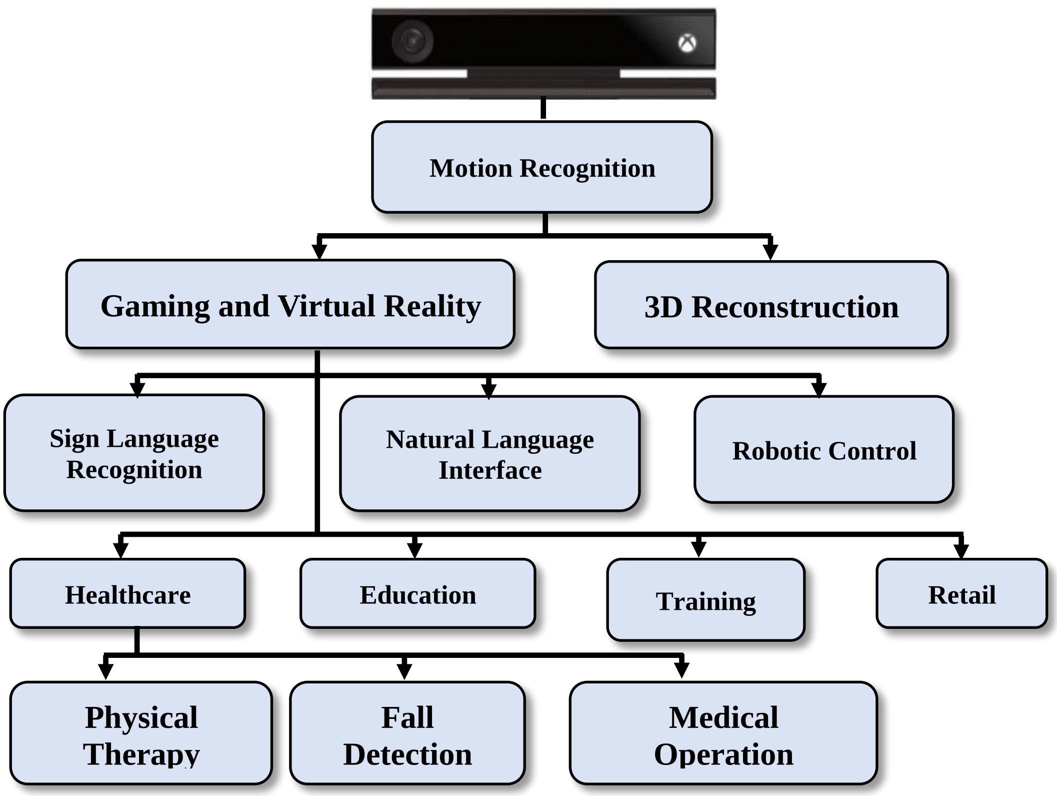


Figure 5.11. Kinect major applications

5.2.1 THE KINECT SENSOR AND JOINT TRACKING TECHNIQUE

Kinect sensor allows us to extract the location and body position of a person, without requiring additional computer vision techniques. It is easier to acquire data, and we can easily extend it to other environments as well. The sensor is not bound to specific lighting conditions. Even when the lighting conditions are continuously changing or when there is no light at all, the infrared technology ensures that we can always track people [MJ13].

Kinect V2 sensor brings to the users some of the latest achievements in human computing technologies, and it is also a facility that enables researchers to develop much more applications by allowing the user to interact with computers by gesturing and speaking. Its specification is mentioned in section 3.6.1, further details can be found in [KI14].

We focus in this part on Microsoft Kinect sensor V2 tracking capability. The most powerful Kinect feature is tracking, face tracking and body tracking.

For body tracking, Kinect can track a total of 25 skeletal joints instead of 20 joints in the older version. The new joints that were added are handy tips, thumbs, and shoulder center. All the skeleton joints are presented in figure 5.12 with the coordinates provided by the Kinect v2. These 25 joints, improved the understanding of the soft connective tissue and body positioning, so we can get more anatomically correct positions for crisp interactions, more accurate avatars, and avatars that are more lifelike. This improvement for the body tracking feature opened many prospect applications.

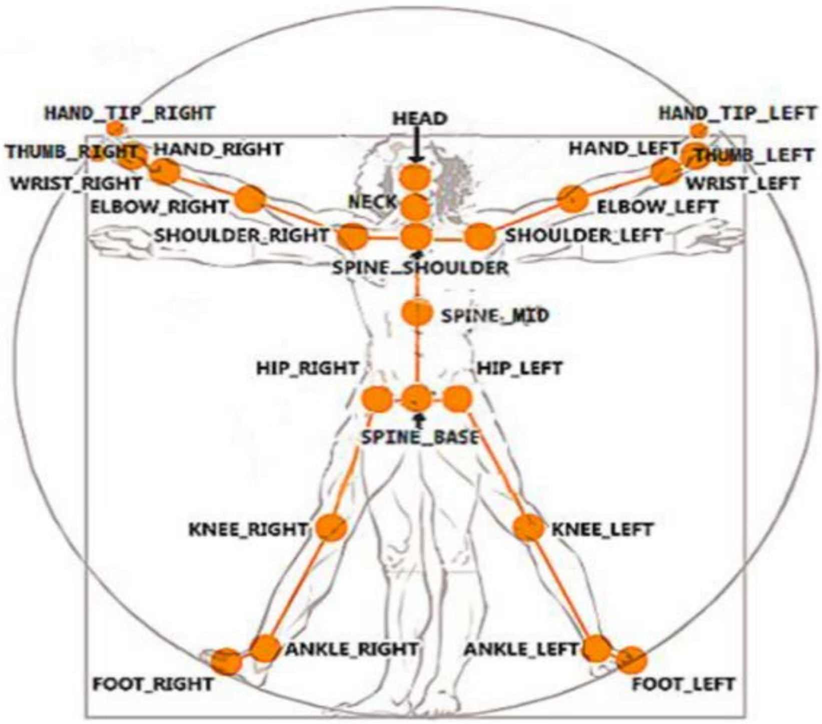


Figure 5.12. Skeleton joints elements in Kinect V2 sensor [JO14]

5.2.2 RELATED RESEARCHES ON MOTION AND GESTURE RECOGNITION

Using Microsoft Kinect, we were able to detect and track human motion easily, bracing the new technology, researchers can focus on the intelligent recognition of activities, instead of having to deal with computer vision first. The use of a logic-based activity recognition approach enables us to reason about the activities of humans based on low-level observable actions. A more complex activity or eventually a behavior can be described in terms of simpler activities.

Many kinds of literature generally focused on the specific field and environments for Kinect's applicability. Extending from analysis of dance performance and golf aids, to online combative technique showing stages and enhanced physical recovery, there is a wide range of focuses and uses cases to be found for this new innovation, some of which will be briefly presented.

Clark, et al. [CL12], has conducted a study about the validity of Kinect's postural control. They did a comparison for twenty various subjects performing three unique and distinct movements between skeleton tracking data acquired from Kinect's sensor and skeletal tracking with an established kinematic assessment tool using 3D camera-based motion analysis. The experimental results of their study show the capability of the Kinect for providing valid anatomical displacement data, compared to the 3D camera-based motion analysis system. They conclude that the Kinect sensor is, therefore, an effective, reliable and marker-less alternative to a lot of elaborate markers, bound 3D camera-based systems for conducting potential anatomical positioning analysis.

Chang, et al. [CH12], presented another comparable research about the feasibility of Kinect. The comparison was made between the use of Kinect against an expert multi-camera setup, to prove the performance and the capability for rehabilitation purposes in clinical and home situations. Their outcomes, assess Kinect as an effective option compared to more costly and prohibitive systems. While they mentioned that the restriction to one camera angle puts certain limitations on the Kinect sensor, it fares remarkably well in most experiments and movement comparisons.

Another approach was presented by *Shotton, et al.* [JS13], as one of the significant research teams and software developers behind Kinect skeletal tracking ability and specifically funded by Microsoft for their contributions to Kinect's technology. Using a

Kinect for skeleton recognition to show the Kinect reliability, they described a new approach to rapidly and precisely forecast 3D positions of body joints from a Kinect single depth image without using temporal information.

From above three studies, Kinect can be considered a potential innovation to give precise joint information to modeling purposes compared to more costly and extensive systems.

On the applicability of the Kinect field, the hardware of the Kinect-focused on its ability to detect, compare and analyze different movement patterns. While these studies are mostly aimed at providing software that supports automated teaching and training methods for a certain set of movements, these studies give valuable insights into different movement comparison and segmentation techniques.

Bo, et al. [BO11], presented a methodology by connecting Kinect's skeletal tracking and internal sensors (accelerometers, gyro-meters) to enhance the accuracy and quality of self-physical treatment and rehabilitation. They are utilizing the combination of both systems to balance out the sometimes significant estimation errors of the inertial sensors. These sensors comprised of accelerometers and gyro-meters appended to a patient's knee and ankle, yet require constant re- calibration. Subsequently, the authors want to utilize Kinect's simple setup and joint tracking capabilities for calibration reference of these sensors, by including 3-D joint angle calculations in the initial setup and initialization process. Their experimental results show that the combination of internal sensors (accelerometers, gyro-meters) with external motion capturing hardware (Kinect) yields easier and more reliable initialization procedures and better visualization capabilities.

D. Alexiadis et al. [DA11], presented an approach to the use of quaternions as a comparison model, by using the Kinect sensor to align the performance dance compared to a teacher, provided a general and an instant score for the student performance, and comments on the performance. They modified C++ OpenNI-supported skeleton tracking software and a MATLAB engine to capture and record the body positions of 17 joints of the student performance (Head, Neck, Torso, Left and Right Collar, L/R Shoulder, L/R Elbow, L/R Wrist, L/R Hip, L/R Knee and L/R Foot) for each frame in a 3D vector signal. They captured dynamic movement for the dancer from the output of a 3D skeleton tracked character joints which were calculated from the convolution of the discrete time position signals with a first order Derivative of Gaussian.

M. Ether, H. Wan and J. Lee [ME13], presented a mechanism to use the human body to control a virtual or real robotic arm by using the Kinect sensor and its SDK. They built a program with C# to get the skeleton and the joints in the human motion from the Kinect sensor, measured the distance between two joints of the skeleton and then generated eight simple commands, UP, DOWN, LEFT, RIGHT, FORWARD, BACKWARD, OPEN, and CLOSE. By using a socket communication, they passed these commands to either a virtual robotic arm or a real robotic arm which implemented the commands.

A.Shingade and A.Ghotkar [AS14], presented a survey of motion, skeleton tracking techniques and different depth cameras. Their system for skeleton recognition included separating the foreground from the background and segmented the human body into many regions. They used 20 joint positions for the skeleton depending on *Fitzgibbon et al*. research [AF13]. They used an open source software (MakeHuman) to build a 3D character and applied a rigging process for the attached skeleton to a human character by applying the algorithm which was proposed by *I. Baran* and *J. Popovic* [IB07]. A database set was built for each gesture with its related joints in the system.

J. Lee, M. Hong, and S. Ryu [JL15] proposed a system that used Kinect V2 sensor, instead of other attached professional medical devices, to monitor the sleep movement, posture, and to extract the sleep information. They proved that the Kinect sensor can provide significant sleep-related information from the human body. Their system was implemented by using C++ with the SDK for the sensor and used the OpenCV library for detection and tracking the movement of the sleeping human body. The Kinect sensor required a distance between 0.5 meters to 4.5 meters, so they located the Kinect at 2 meters on the body. A critical 19 joints from 25 joints were selected which were related to the sleep movements. The values of the movements were calculated every 0.5 seconds using Euclidean Distance between the former position of joints and the current position of joints from image sequences. The system continues to accumulate the values and make the comparison with 5 references sleep postures to determine user's current sleep posture until the wake up of the sleeper.

Lin, et al. [LI13] used Kinect as a golf training tool for beginners. They utilized the Kinect sensor to identify six distinct types of commonly made mistakes during the golf swing movement. These positional mistakes range from shoulder and knee disposition moves in the focal point of gravity of the golfer. Keeping in mind the end goal to recognize and dissect these mistakes, they proposed a mathematical, 2D coordinate system evaluation technique that is specifically trained to identify these six dispositions. They experimented the system by comparing the system's results in all six mistakes with the analysis of the movement by an expert. Their outcomes demonstrate that the system is reasonably accurate in identifying instance mistakes and position of novice players.

Zhang, L. et al. [ZH13], presented another exploration to enhance and improve Kinect's practicality for golf swing analysis. They built an automated system utilizing the Kinect sensor that have the ability to segment golf swing recordings and classify them based on an exclusively modified scoring system. For analyzing and classification, they utilized the extracted positional data and applied a Support Vector. They classified the positions in four various classes based on classification dataset. The results demonstrate that their methodology is accurate in extracting and classifying test swings accordingly, with an average exactness of 84%.

Yao-Jen Chang, Shu-Fang Chen, and Jun-Da Huang [YA11], presented a system based on Kinect for physical rehabilitation, by studying done on a pilot for young adults with motor disabilities.

These studies are mostly aimed at providing software that supports automated teaching and training methods for a certain set of movements, these studies give profitable experiences into diverse movement comparison and segmentation techniques.

5.3 OUR SYSTEM FOR VIRTUAL SPORTS TRAINING BASED ON HUMAN MOTION TRACKING USING KINECT V2 SENSOR

5.3.1 THE APPROACH

Virtual reality is an innovation which permits a user to collaborate with a PC reproduced environment, whether that environment is a recreation of this present reality or a fictional universe. Most of the present virtual reality environments are essentially visual encounters, showed either on a PC screen or through unique or stereoscopic showcases, yet a few reproductions incorporate extra tactile data, for example, sound through speakers or earphones [WH15].

Virtual reality uses advanced technologies, including PCs and different interactive media peripherals, to create a simulated environment that clients imagine as similar to real-world objects and events. With the guidance of extraordinarily planned transducers and sensors, users can work with showed images, moving and controlling virtual things, and performing different activities in a manner that brings on a sentiment genuine vicinity in the reproduced environment [AL11]. One of the cardinal elements of virtual reality is the procurement of a feeling of the real vicinity in and control over the reenacted environment. This element is accomplished to more noteworthy or lesser degrees in the different utilizations of virtual reality, contingent on the objectives of the specific application and the expense and specialized intricacy its engineers are eager and ready to accept [KO04].

Virtual reality simulated in a 3-D image that can be investigated conjecturally at a PC, generally by controlling keys or the mouse so that the substance of the picture moves in some bearing or zooms in or out. More modern endeavors include such methodologies as wrap-around showcase screens, real rooms enlarged with wearable PCs, and haptic devices that let the user feel the presentation pictures [WH15].

The use of Virtual Reality technology for developing tools for rehabilitation has attracted significant interest in the physical therapy arena [BE11]. Sports training and rehabilitation have recently achieved a major enthusiasm for our life, so such a large number of researchers attempt to build interactive software applications, assisting the users with learning how to do the right sports rehabilitation movements or various trainings.

Virtual reality is not characterized or restricted by any technological methodology or equipment setup. The making of a VR client experience can be accomplished using mixes of a wide assortment of interactive gadgets and sensory presentation frameworks and the design of content presented in a computer-generated graphic world. Progressive advances have happened in the fields of VR and intelligent computerized innovative technology, and this has bolstered innovative work that has focused on essential societal-level medicinal services challenges in ways unrealistic in the past [BE11].

Sports training systems are based on human motion capturing and tracking. The subject of human motion capturing and tracking received a considerable attention over the last decades. There is a wide range of applications that use human motion tracking while the industry provides a novel constantly movement tracking systems, which have a great accuracy and high performance. Human motion tracking is receiving increased attention from computer vision researchers. This interest is motivated by a wide spectrum of applications, such as therapists, athletic performance analysis, military applications, video games, video conferencing, validation of computer vision and robotics [LC11].

Athletes are always searching for additional thoughts to assist them to perform in their sport. *Pierre Beauchamp and Jocelyn Faubert* [PI11], in their paper asked, “How do we create effective simulations for training purposes?”. The answer was by using virtual reality. In virtual reality, users will move their heads, eyes, and limbs explore multisensory 3-D integration whenever they will act with objects [AL07]. Sports organizations, researchers have explored the potential of virtual reality environments with the exception of baseball batting gloves [AN93] and table tennis [TO97]. The system presented by them outline a training technology for enhancing perceptions-cognitive skills of athletes achieved through training with a 3D ”cave‟ environment [PI11].

Our approach aims to record and display on the screen the sequence of sports or medical exercise, performed by a professional sportsman or a medic and captured using Kinect V2 sensor, and then represent them on the screen using a 3D avatar. The trainee person will need to imitate these exercises in front of a Kinect and the system will record the imitative actions performed and will make a comparison with the trainer’s movements. A real-time correction will be shown on the screen through messages and body parts highlights, assisting the trainee to correct his movements. Thus, the trainee can self-learn the sequence of exercises by following the real-time visual feedback. Toward the end of the considerable number of movements, a synopsis of his execution and a general aggregate score will be shown on the screen. In this methodology, information of the body skeleton will be used (joint pivots, positions, and angles between joints) which are required to make the comparison. Our proposed system covered the following steps:

1. Record the sequence of a sports exercise, performed by a professional sportsman or a medical doctor (the trainer).
2. Display on the screen the recorded sequence, using a 3D character (Prof_Action).
3. Record the imitative actions performed by the trainee user (User_Action).
4. Compare the trainees performed action with the professional action (Prof_Action vs. User_Action).
5. Display Score and Feedback (body part highlight and text suggestions)

Figure 5.13 shows the flowchart of our proposed system processing.

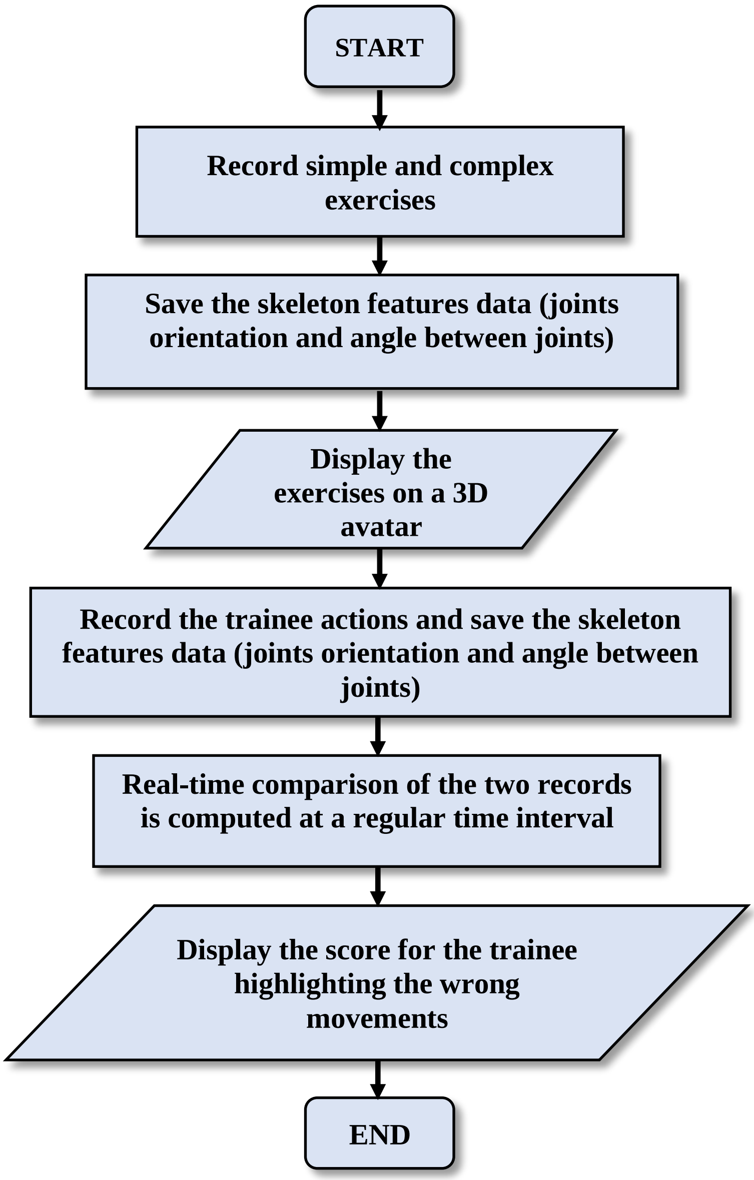


Figure 5.13. Flowchart of the proposed approach

5.3.2 SYSTEM IMPLEMENTATION

Unreal Engine 4 (UE4) software supports Kinect V2 sensor with all its 25 joints and it is an extremely powerful graphics engine for building and developing games or other graphical applications, starting from 2D mobile games to full 3D immersive applications. Unreal Engine 4 provides all that we needed to begin building our project, so we used it to deliver a prototype very fast.

We also have chosen UE4 engine because of its great scripting language: Blueprint - a visual scripting tool that hastens a lot the development process and helps to focus just on the actual functionality and quick results. The Blueprint system represents components (functions, variables, events, etc.) as nodes that can be related (connected) with other components through virtual wires (Figure 5.14).

Unreal Engine innovation powers several diversions and also real-time 3D movies, training reproductions, representations, and more other things. A huge number of people and groups and have manufactured professions and companies around abilities development utilizing the engine.

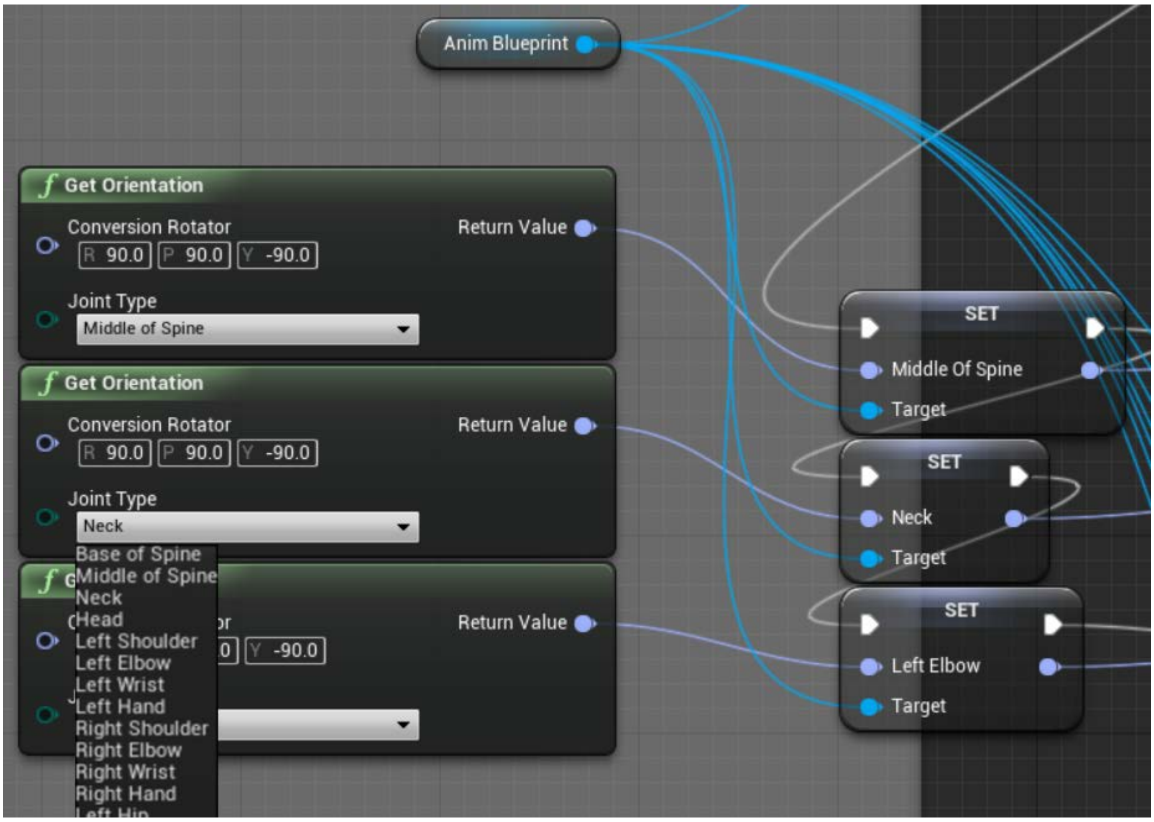


Figure 5.14. Kinect 4 Unreal plugin Blueprint example

Unreal Engine 4 has numerous features like:

a. Rendering and Graphics: the rendering in Unreal Engine 4 is an all-new, DirectX 11 incorporated conceded shading, worldwide illumination, lit translucency, and post processing and additionally GPU molecule simulation using vector fields

b. Unreal Motion Graphics UI Designer (UMG): a visual UI authoring tool which can be used to create UI elements such as in-game HUDs, menus or other interface related graphics you wish to present to your users.

c. Skeletal Mesh Animation: (UE4) animation system takes into consideration a profound level of control for the characters and Skeletal Meshes by mixing the skeleton-based deformation of meshes with morph-based vertex deformation to allow for complex animation.

d. Audio and Sound: Unreal Engine 4's audio system provides instruments and components to shape the sounds in the diversion to give them the desired feel. This is imperative in light of the fact that a perfect form of the sound can be produced once in an external application, imported, and then crafted within made inside of the engine to create the suitable result.

e. Environment Query System: The Environment Query System is a feature of the Artificial Intelligence system in Unreal Engine 4 for collecting data on the environment, asking questions of the data through Tests, then returning then one Item that best fits the questions asked.

f. Many other features like physics simulation, open world tools, landscape outdoor terrain, Foliage instanced meshes, level streaming, matinee and cinematic, media framework, performance and profiling and packaging and cooking games [DO].

Also, UE4 has support for Kinect V2 SDK through several middleware plugins. One of them, Kinect 4 Unreal (sometimes abbreviated to K4U), permits Unreal Engine 4 developers to use the Blueprint visual scripting system to access the full functionality of Kinect 2 for Windows as seamlessly and as easily as possible. The key outline goal of K4U is to engage artists and designers by exposing everything through Blueprint so that the development group has the ability to focus their exertion on developing their thoughts without a devoted coder. Presenting more than 30 original nodes, all with broad documentation and pre-assembled Avateering systems, K4U exposes all that the Kinect has to offer to UE4 developers. The plugin primarily interfaces through a component-based system where, once activated through the plugin menu, any Blueprint can have the Kinect Interface Component added to it. The Kinect Interface Component gives developers access to a wide number of Blueprint nodes, each granting them access to some aspect of the Kinect functionality [OQ].

For all of the functionalities, we implemented an easy to use graphical interface in UE4 (Figure 5.15). A user can easily record and name an exercise, and play a record from a list.

Both the recording and the play start with a counter and have a stop and return button.

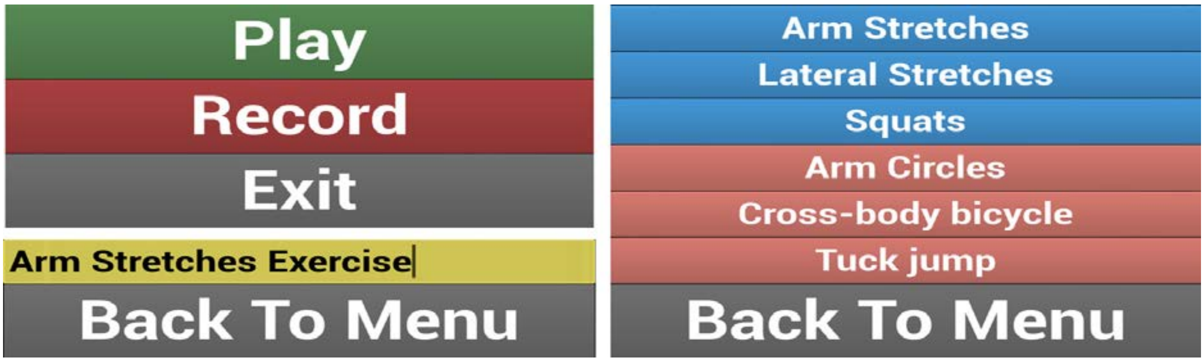


Figure 5.15. Main menus

1. Recording Stage

We used the features of UE4, along with the K4U plugin and the Blueprint scripting to record motion using the Kinect V2 sensor and then display it on the screen using a 3D avatar. First of all, we built a humanoid 3D skinned avatar with a skeleton animation and imported it in the graphics engine. In UE4, the joint structure for a rigged skeletal mesh is represented as a simple tree list as shown in Figure 5.16.

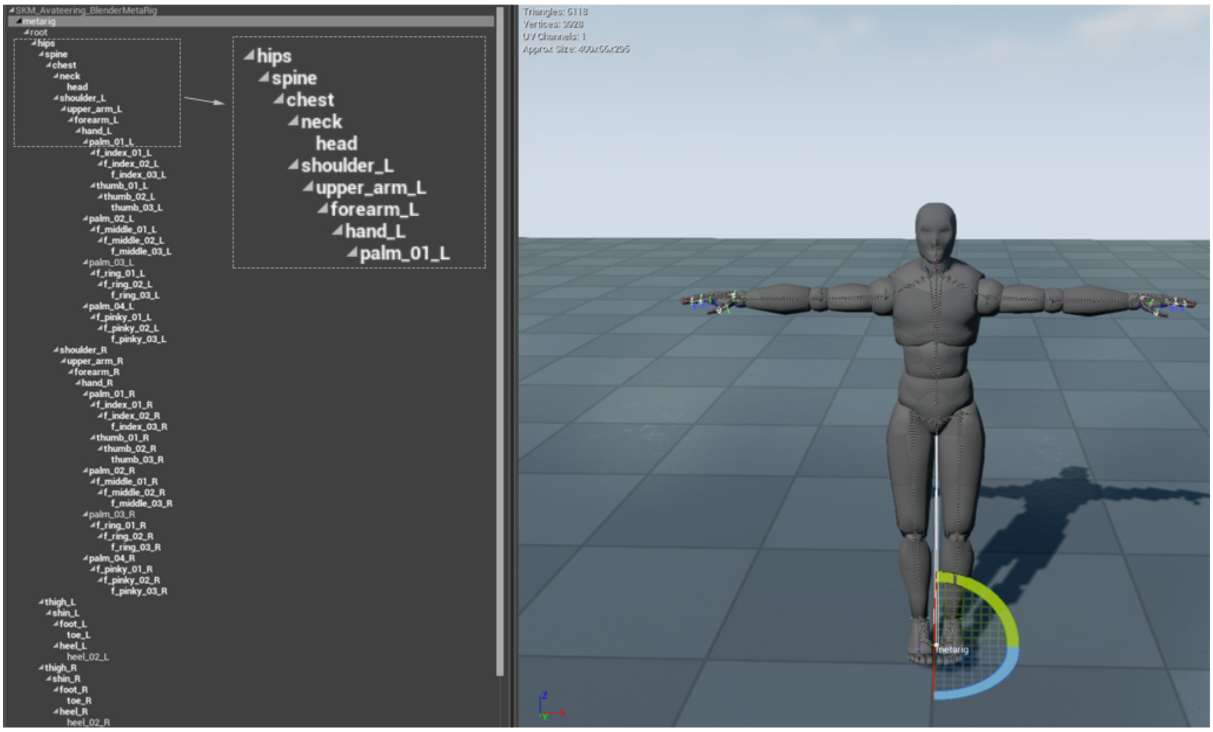


Figure 5.16. Skinned avatar with joint list in UE4

Then, by using the K4U provided functions in the Blueprint system, we were able to capture the Kinect data for all the 25 available joints and combine them to animate our skinned avatar and then display the avatar in real-time on the screen. The skeletal animation in UE4 consist of Bone Transforms that are composing a final animation pose, so we used the data from the Kinect to update the bone transforms (Figure 5.17).

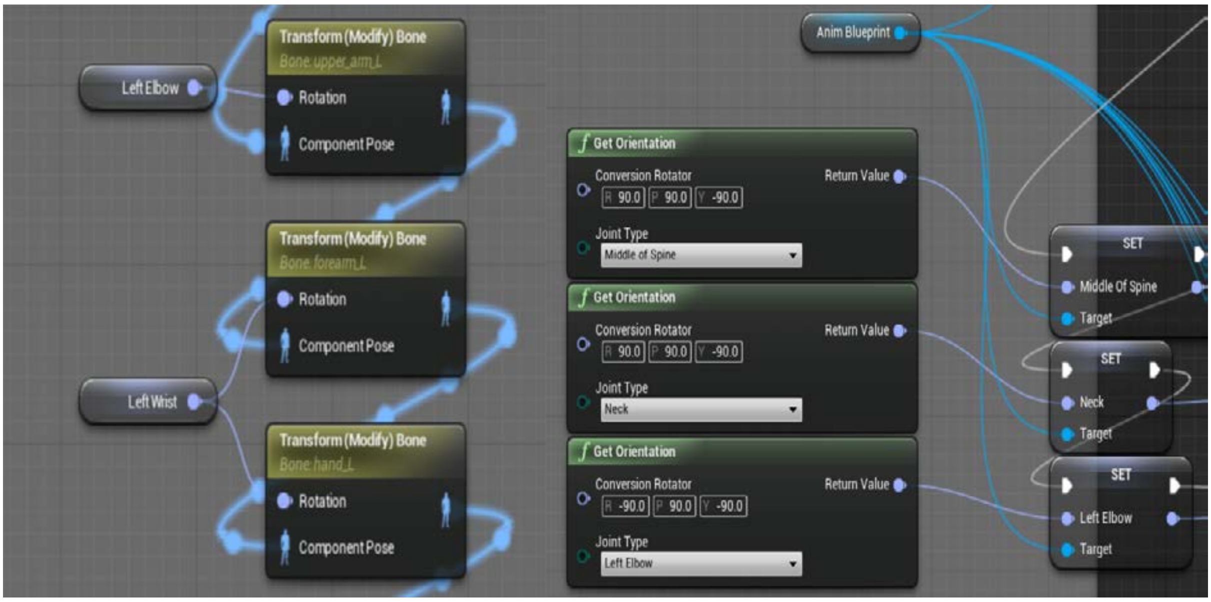


Figure 5.17. UE4 Composed Avatar Animation Pose updated from Kinect V2

In order to be able to save the data as a recording and then play it back by a trainee, we took advantage of UE4 that has already a basic saving system feature that can save any data. We defined a custom data structure with joint orientations (all the 25 available joints) and we used it to save each frame of the recording. This structure is then saved to disk in a specially named slot using the UE4 saving system and can be loaded anytime just by referencing the saved slot name (Figure 5.18). This way, we can have unlimited records, but currently the menu for playing the records is limited to several entries.

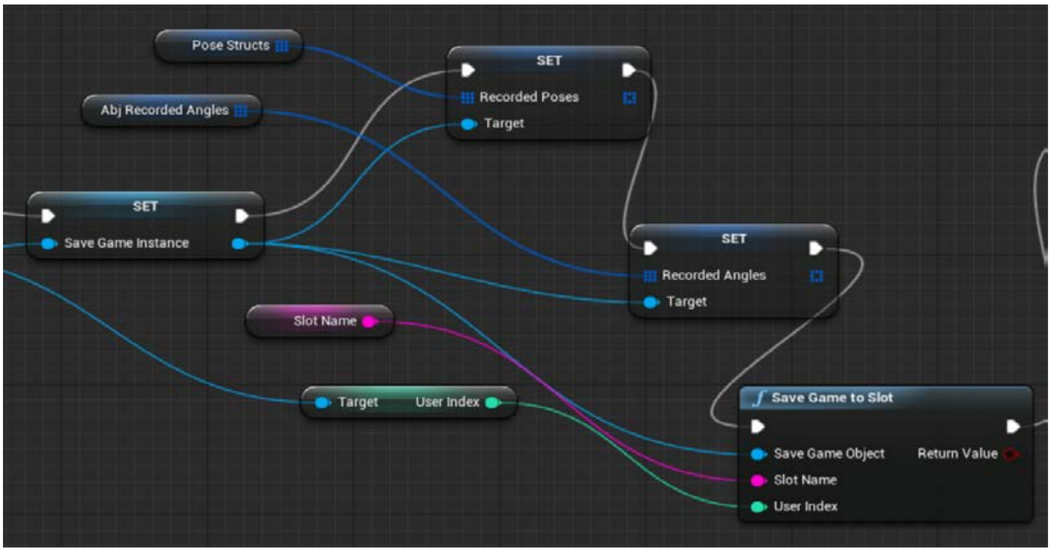


Figure 5.18. UE4 saving system using Blueprint

2. Comparison Stage

After the trainee selects the desired exercise, the system will first start a counter for the trainee to be able to prepare himself. After the countdown has finished, the recorded exercise will start to play so the trainee will have to imitate the record. Both the record and the trainee are represented on the screen using 3D avatars, one playing the training record to imitate and the other one with the real-time captured motions of the trainee. The system will capture the trainee movement data similar to the recording stage and will run real-time comparison algorithms.

We experimented with two different comparison algorithms: simple joint orientation and angles between joints [HA12] providing real-time feedback side by side.

The simple joint orientation algorithm captures joint orientations each frame and compares the values of the record and the live trainee for each of the 25 joints, returning the total matching percentage. We implemented this function programmatically and then we exposed some Blueprints functions to be able to easily interact with them (Figure 5.19). We considered a threshold of a 20-degree angle, and the comparison result is computed each second, based on the maximum result from every frame comparison set computed in that second.

```
UComparePoseStructures::UComparePoseStructures(const class FObjectInitializer& ObjectInitializer)
  : Super(ObjectInitializer)
{

}

float UComparePoseStructures::ComparePoseStructure(const FKinectCPose& CRecordedFrame, const FKinectCPose& CBodyFrame)
{

  float result = 123;
  bool results[25];
  int joints = 25;
  results[0] = compareFrameJoint(CRecordedFrame.AnkleLeft, CBodyFrame.AnkleLeft);
  results[1] = compareFrameJoint(CRecordedFrame.AnkleRight, CBodyFrame.AnkleRight);
  results[2] = compareFrameJoint(CRecordedFrame.ElbowLeft, CBodyFrame.ElbowLeft);
  results[3] = compareFrameJoint(CRecordedFrame.ElbowRight, CBodyFrame.ElbowRight);
  results[4] = compareFrameJoint(CRecordedFrame.FootLeft, CBodyFrame.FootLeft);
  results[5] = compareFrameJoint(CRecordedFrame.FootRight, CBodyFrame.FootRight);
  [...]

  int count = 0;
  for (int i = 0; i < joints; i++) {
    if (results[i] == true) count++;
  }

  return (1.0 * count / joints) * 100;
}

bool UComparePoseStructures::compareFrameJoint(FRotator RecordedJoint, FRotator BodyJoint)
{

  return BodyJoint.Equals(RecordedJoint, 20);

}
```

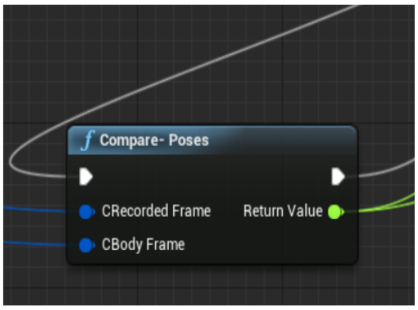


Figure 5.19. Simple joint orientation comparison algorithm exposed to Blueprint

The angles between joints algorithm is implemented directly in the UE4 Blueprint system using multiple custom defined functions. The algorithm proposes to define vectors that join two joint points and then compute the angle between sets of two vectors so the first step that we did was to implement this function in the Blueprint system (Figure 5.20).

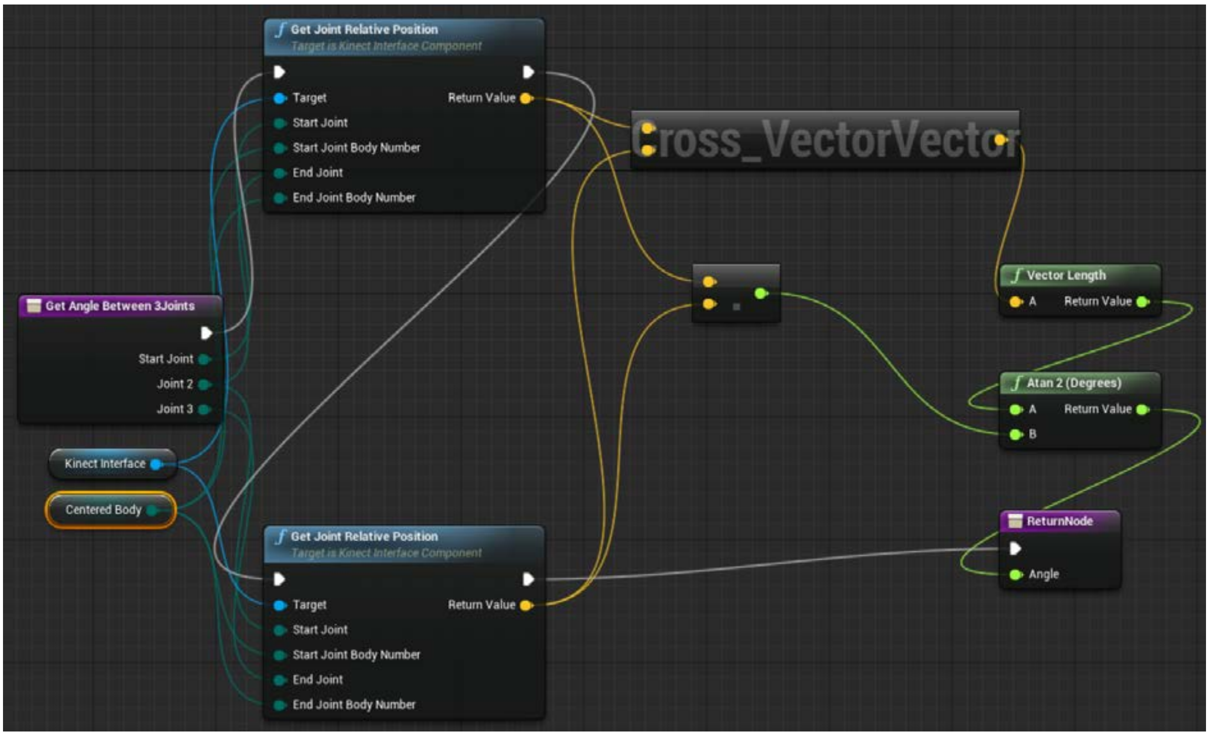


Figure 5.20. Function to compute the angle between joints implemented in Blueprint.

The next step was to define multiple sets of vectors to cover the entire body parts and to compute the angle for each of them. In order to use all this data for comparison, we had to define another custom data structure for the angle between joints and save each frame in the same way as we did for the joint orientations using the UE4 saving system (Figure 5.21). This way, we can capture and compare movements using a more natural approach (for example, we would rather compare the angle between the arm and the forearm or the shoulder and the arm, then just comparing the elbow joint or the shoulder joint orientation). For our implementation we used a number of 21 vector sets as follows:

- Head-SpineShoulder-SpineMid
- RightWrist-RightElbow-RightShoulder
- RightElbow-RightShoulder-SpineShoulder
- RightAnkle-RightKnee-RightHip
- LeftWrist-LeftElbow-LeftShoulder
- LeftElbow-LeftShoulder-ShoulderCenter

- LeftAnkle-LeftKnee-LeftHip
- RightHip-SpineBase-RightKnee
- LeftHip-SpineBase-LeftKnee
- RightWrist-RightHand-RightElbow
- LeftWrist-LeftHand-LeftElbow
- RightAnkle-RightFoot-RightKnee
- LeftAnkle-LeftFoot-LeftKnee
- SpineBase-SpineMid-LeftHip
- SpineBase-SpineMid-RightHip
- SpineShoulder-SpineMid-SpineBase
- Neck-Head-SpineShoulder
- Head-RightShoulder-Neck
- Head-LeftShoulder-Neck
- SpineShoulder-Neck-SpineMid

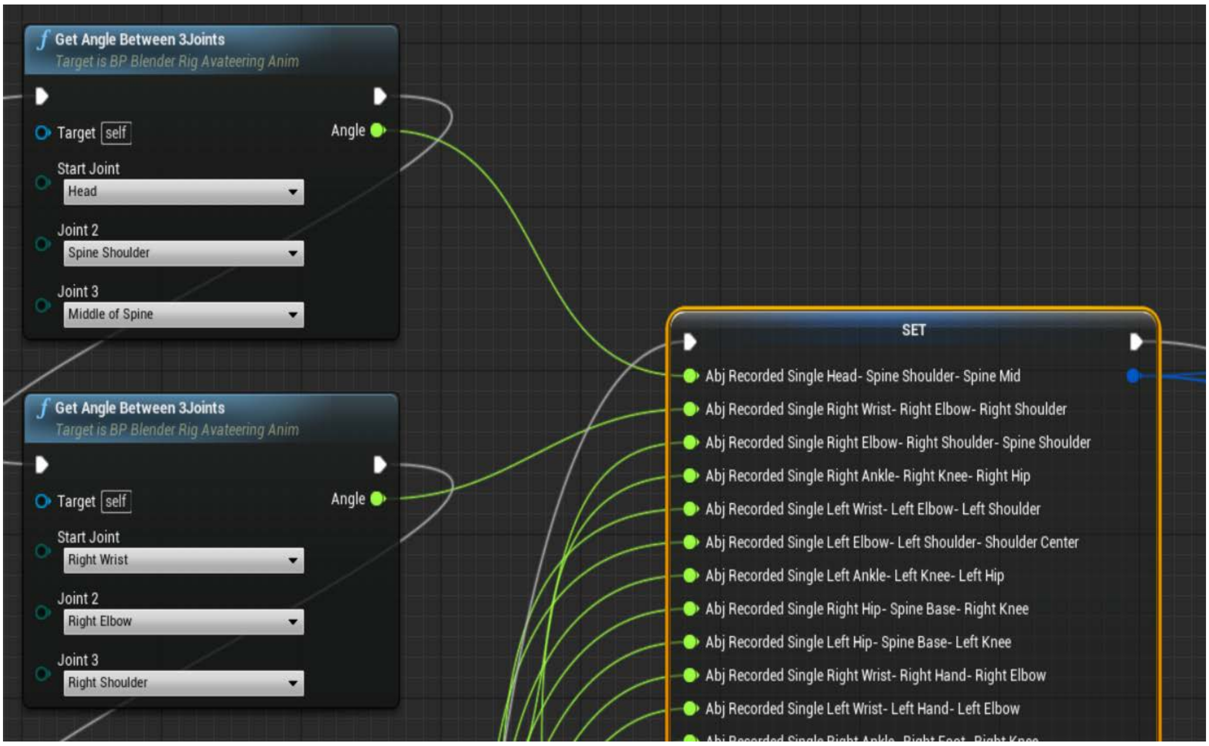


Figure 5.21. Blueprint function to save angle between joints sets for each frame

The comparison of the recorded and the trainee imitation is computed, as with the other algorithm, each frame but with the result computed at the 1-second interval, based on the maximum result from every frame comparison set computed in that second (Figure 5.22).

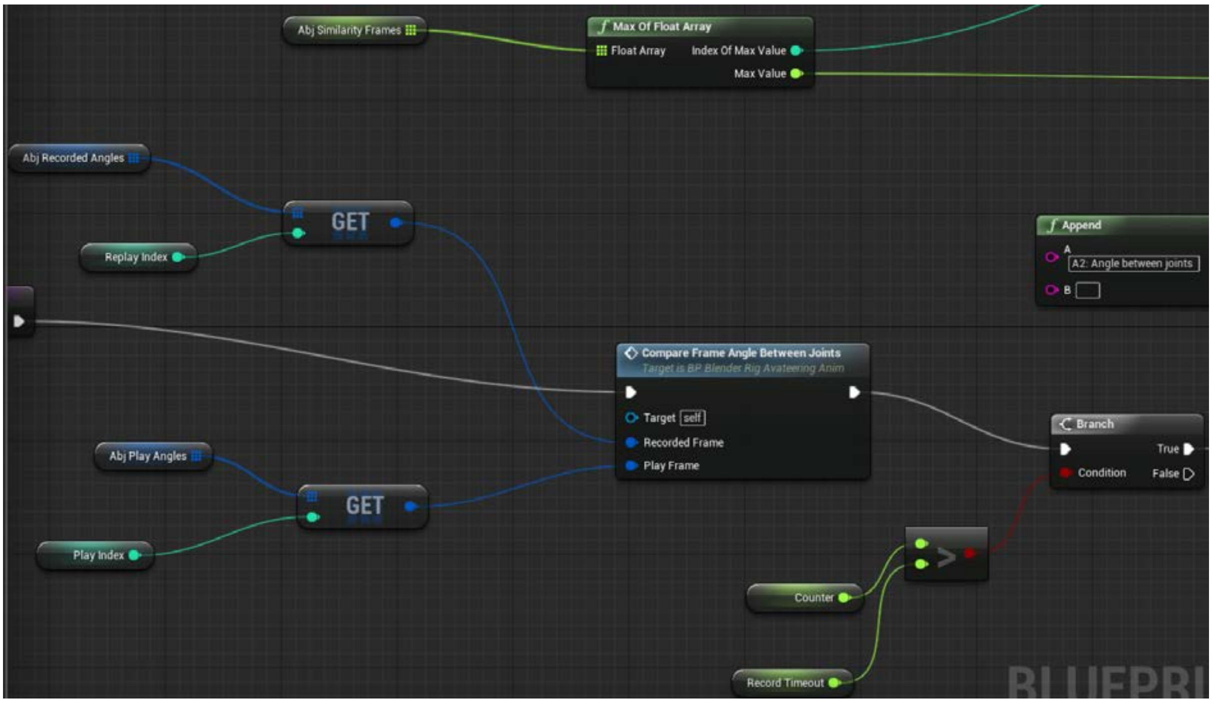


Figure 5.22. Angle between joints algorithm overview in Blueprint

For each frame, we compare the angles for each set of vectors using another custom function defined in Blueprint (Figure 5.23). The function also includes the use of a threshold for the comparison, experimentally configured to a 12-degree value.

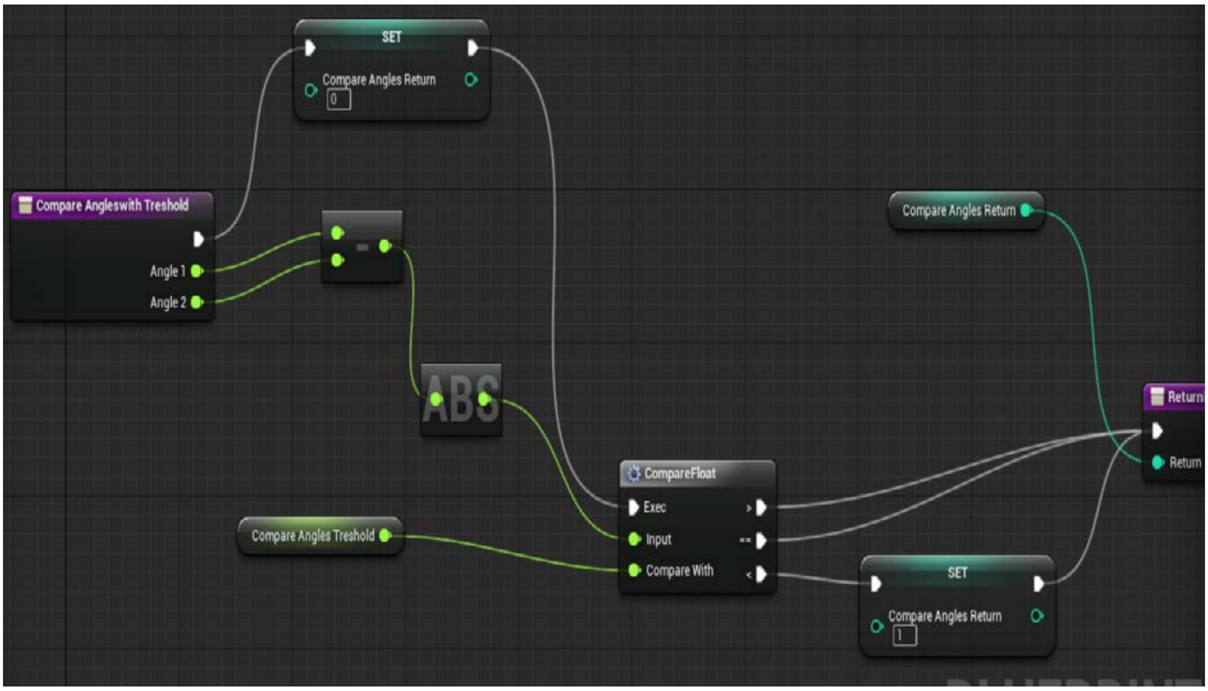


Figure 5.23. Blueprint function to compare angles with threshold

For both algorithms, we considered a movement matched if it was within the defined threshold.

The primary results of the comparison are the total score and the visual feedback on the avatar body using colors.

In the first iteration, we colored the whole body with red if the movement was wrong, yellow if the movement was partially matched and green if it was considered matched (Figure 5.24).

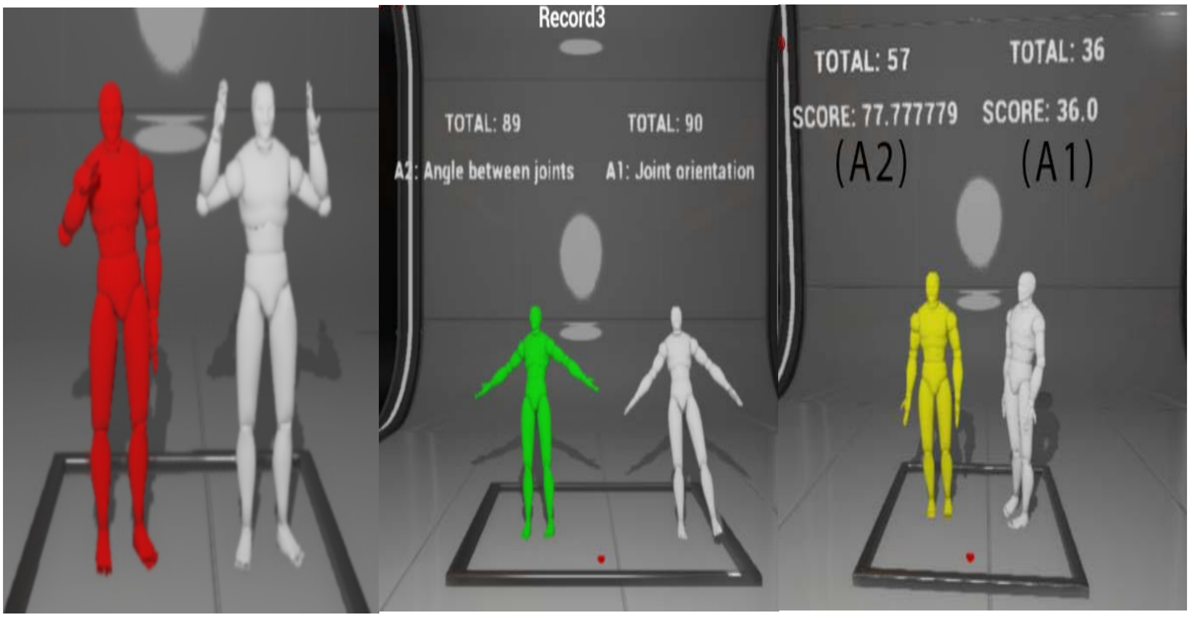


Figure 5.24. The first implemented version for movement highlighting

Still, the visual feedback was far from benefit to the trainee, because it didn't highlight the wrong part of the body, and a trainee could hardly understand what he was doing wrong. For example, in the first part of the above figure, the movement is marked as wrong, but only the hands are in a wrong position, the rest of the body is being correctly positioned. For this reason, we implemented a body part highlighting algorithm.

In order to implement body part highlighting, the avatar body mesh must be composed of multiple body parts, so we need to have an individual mesh object for each part that we want to be able to highlight. Separating a body mesh into multiple parts is quite easy in Maya, but skinning them to the same skeletal rig and animation is a complicated task. So, using Autodesk Maya we had to separate the body from the skeleton, and for each body part to skin again with the same skeletal rig (Figure 5.25).

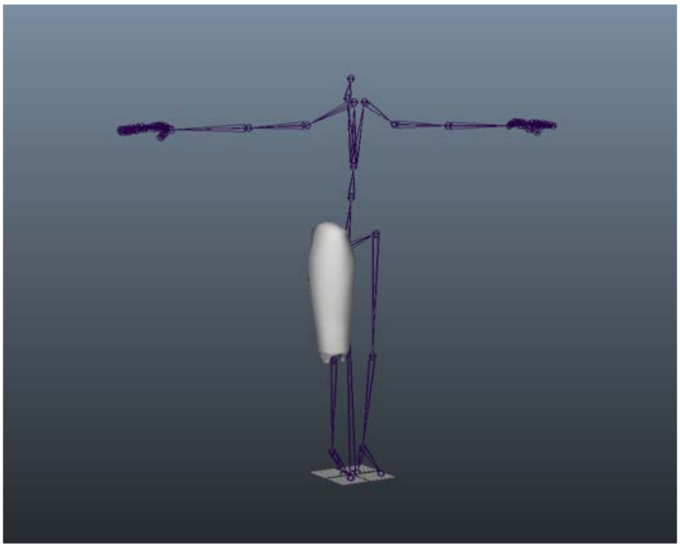


Figure 5.25. Rigged body part mesh in Maya

Combining each individual body part into a single body pose sound hard, but UE4 great features helped a lot. As it is mentioned above, UE4 keeps the skeletal metal rig as a simple tree list, so if the body parts have the same tree list the only thing you must do is to set a master pose component (for example the head) and to bind each other part to it. Obviously, all body parts should use the same skeletal animation and should be placed accordingly. For the actual highlighting, we used the movement results from the angle between joints algorithm as seen in Figure 5.26. Given the fact that a single body part is included in more than one vector set, we created an algorithm to score each body part combining all the vector sets results.

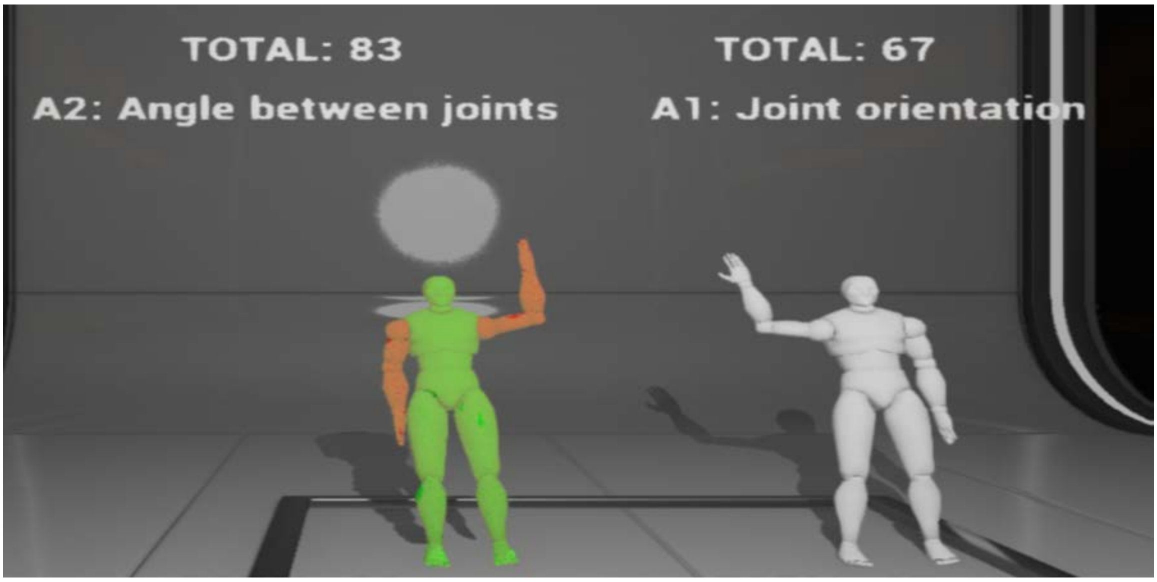


Figure 5.26. Real-time movements highlight feedback. (A) – Trainee Real-Time Movement (B) – Trainer Record

Still, the score was not very relevant for simple exercises, because all the joints had the same relevance. For example, if we consider the arm stretches exercise, the score difference from perfect hand moves and wrong moves would not be so high (the trainee will still get more than 70% matched movement if the other body parts are standing still).

In order to improve the relevance of the score we created an algorithm to measure each joint relevance in the exercise and compute the score based on that relevance. First of all, we needed the relevance of each joint in a particular movement so we assigned each joint with a weight. We computed the total movement by adding the angle between joint differences for each frame relative to the previous one, then the final joint weight was computed by dividing the total movement to the total sum of the joint movements (Figure 5.27). The result was a relevance percentage for each joint that will be used to compute a weighted average for the final score. This way, each joint will be scored according to how much it has contributed to the whole movement. Notice that this relevance algorithm is applied only to the angle between joints comparison algorithm.

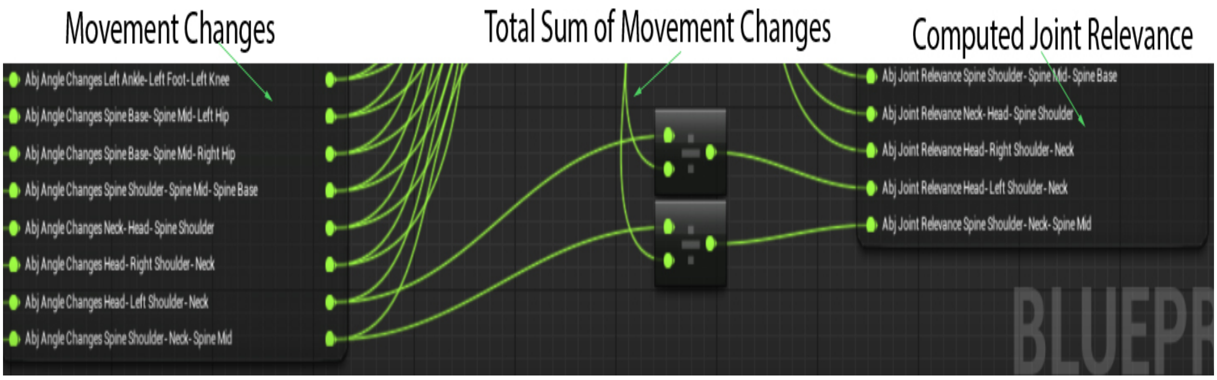


Figure 5.27. Joint relevance percentages algorithm

The last step in our proposed approach was to display text feedback correction in real-time (for example, if the trainee has the hand too low he will receive a text message in order to move his hand upper). This kind of feedback will complete the correction system by providing real-time instructions on how to correct the movements. In order to accomplish this, we captured, for each joint, the angle difference between the trainee movement and recorded one. Based on this result, we defined some templates to display: “Move your <body-part> <lower/upper>”. So, for each body part, if the movement was not matched, we compared the resulting angle and displayed the messages accordingly (Figure 5.28).

The last problem was that there could be too many wrong body parts and the trainee would be overflowed with feedback messages. In order to solve this problem we chose a basic solution: display only one body part text correction at a time, so if the trainee corrects that body part he will receive feedback for the next wrong body part and so on. Of course, the algorithm could be improved to use a relevant order also in this textual feedback.

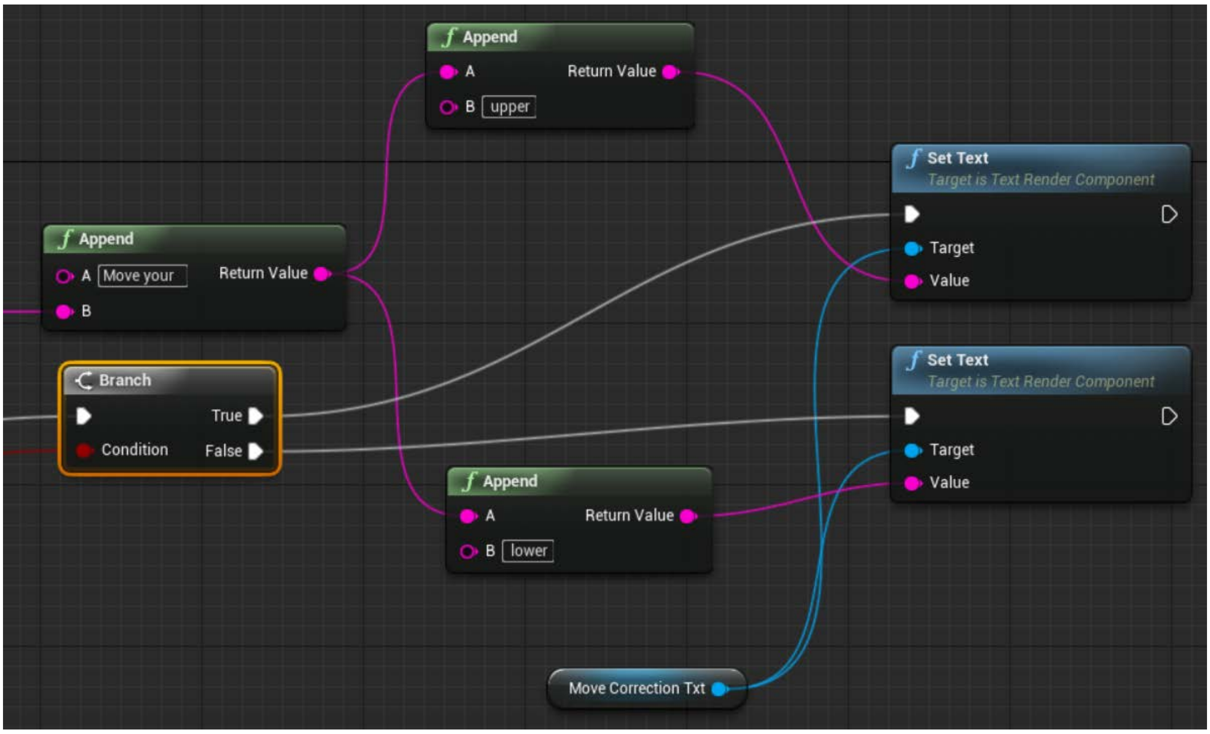


Figure 5.28. Text feedback display algorithm using templates

5.3.3 Experimental Results

To assess the proposed system we experimented by using several types of exercises. We started with some simple stretch movements and then continued with some more complex exercises. Each recorded exercise was imitated by two types of trainee, one with sports background (U1) and another a common person (U2). The following results are measured using the basic proposed system without our final improvements: body-part highlighting, score relevance and text feedback. The system only highlights the performance in real-time, providing full-body color feedback for each movement and computing a simple general score for each presented algorithms: A1-joint orientation and A2-angle between joints (Figure 5.29).

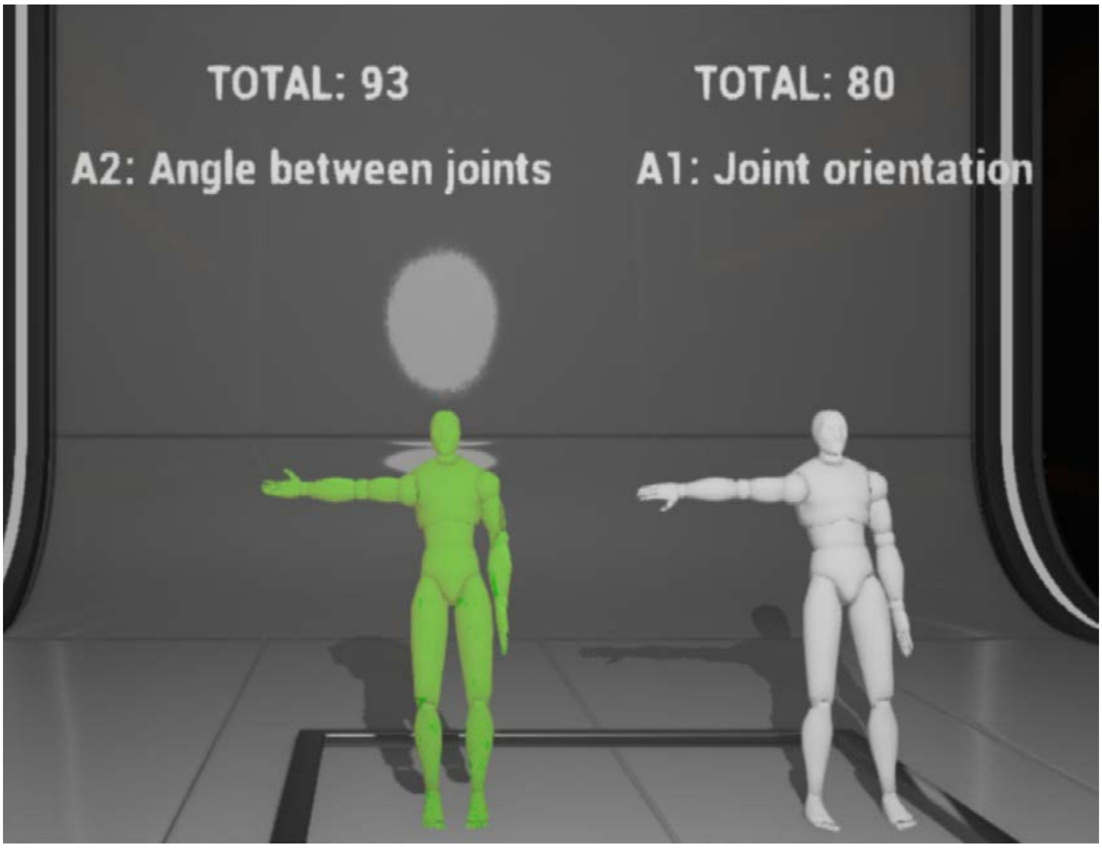


Figure 5.29. Imitation of a model with computed total score and basic movement highlighting

Tables 5.1 and 5.2, showed the experimental results of the used exercises (simple and complex exercises). With each exercise we associated the joints which have an effect on these exercises.

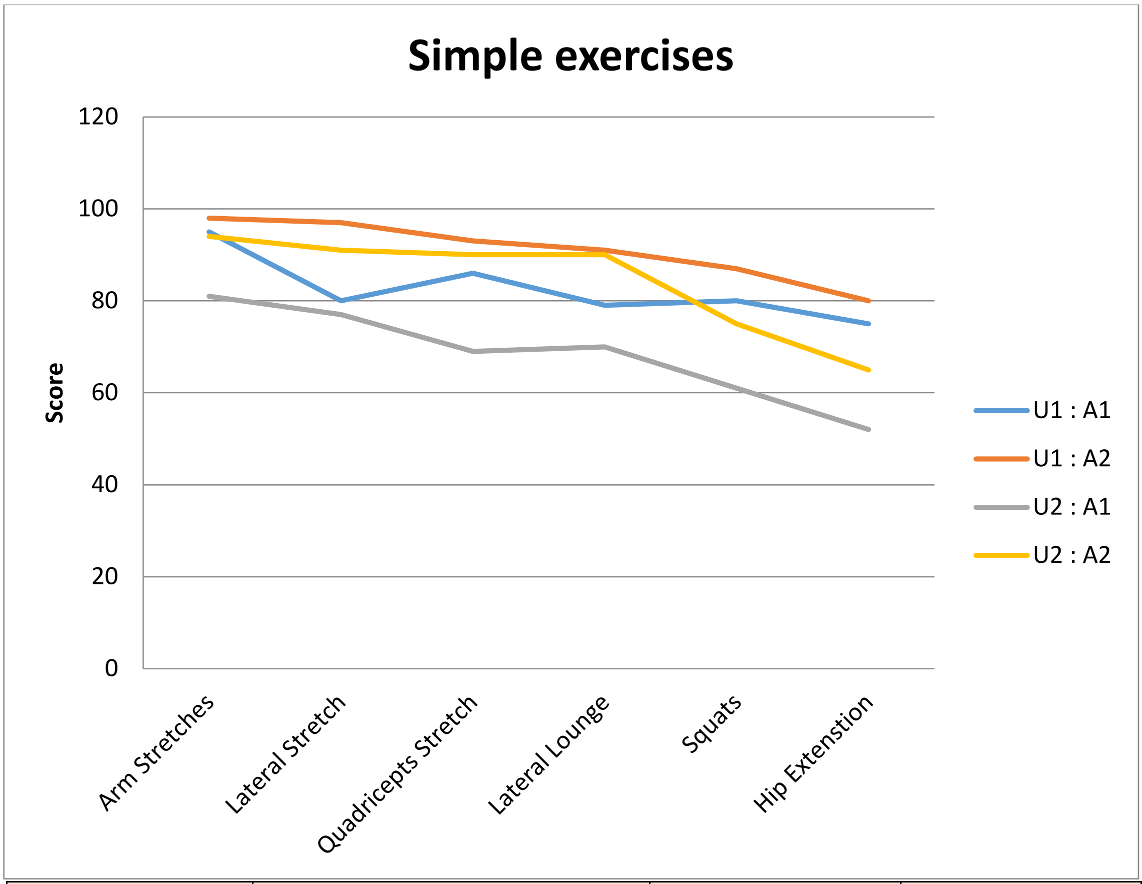
From the results, we notice that the score is decreasing proportionally with the increasing complexity of the exercises. This is somehow expected because the most complex is an exercise the harder is to imitate it, even for an experienced trainee. Still, there are a couple of remarks, resulted from this experiment, to take into account regarding the system.

First of all, even if it has greatly improved since the last version, the capture from the Kinect V2 sensor is not perfect, so smoothing should be manually applied to the captured values to lower the noise.

The countdown function was a very good addition. The system also shows the first movement of the training during this countdown, so the trainee has time to prepare and get in the correct position for training.

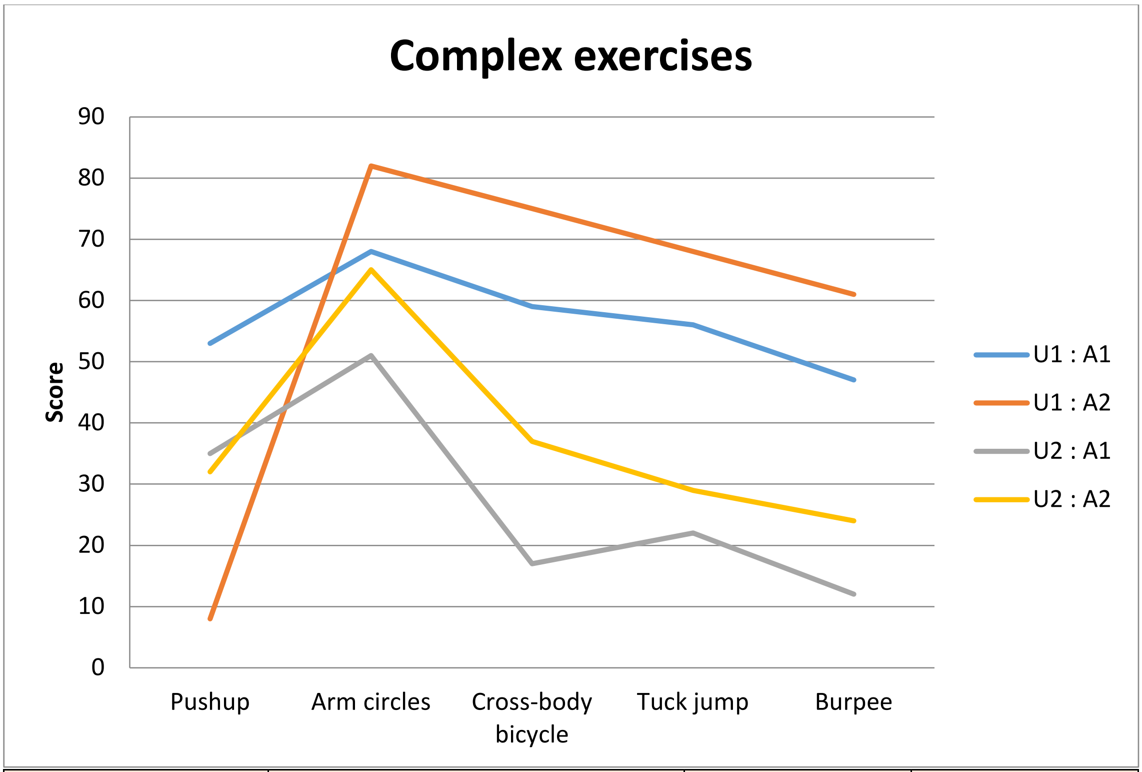
The synchronization threshold, comparing the movement each second (so it can cover up to 60 frames differences) is also very important as the user can perceive and follow usually just one movement in one second.

Table 5.1. Simple exercises



Exercise	**Relevant Joints**	**Score A1**		**Score A2**	
		U1	**U2**	**U1**	**U2**
Arm stretches	Shoulder, Elbow, Wrist	95	81	98	94
Lateral Stretch	Shoulder, Elbow, Spine (shoulder, mid, base), Hip	80	77	97	91
Quadriceps Stretch	Knee, Foot	86	69	93	90
Lateral Lunge	Hip, Knee, Foot	79	70	91	90
Squats	Hip, Knee, Foot, Spine (mid, base)	80	61	87	75
Hip Extension	Hip, Knee, Spine (base)	75	52	80	65

Table 5.2. Complex exercises

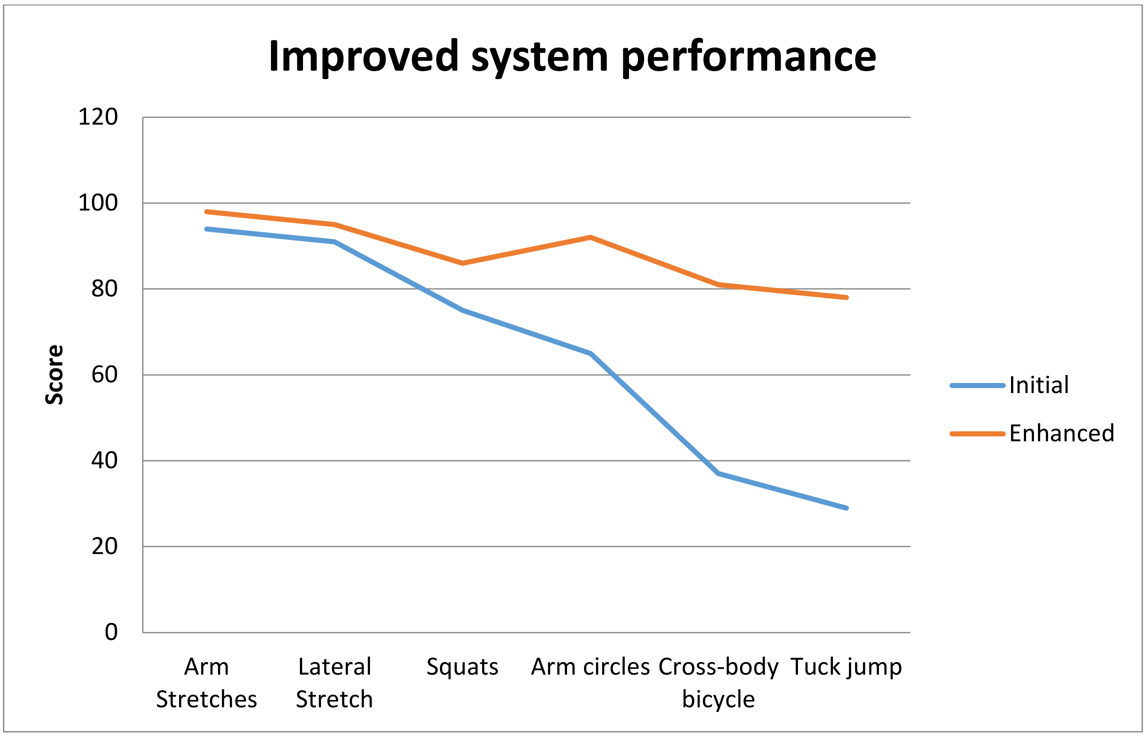


Exercise	Relevant Joints	Score A1		Score A2	
		U1	U2	U1	U2
Pushup	Shoulder, Elbow, Spine (shoulder, mid, base), Foot	53	35	48	32
Arm circles	Shoulder, Elbow, Wrist	68	51	82	65
Cross-body bicycle	All	59	17	75	37
Tuck jump	All	56	22	68	29
Burpee	All	47	12	61	24

Regarding the comparison algorithms, we can see from the experimental results that the A2 – angle between joints algorithm is more stable and has better results.

In order to see if our enhancements are improving the trainee performance and the overall system functionality we had another set of experiments with only several exercises from the original set. The results are compared with the initial results only for the angle between joints algorithm (Table 5.3).

Table 5.3. Improved system performance



Exercise	Relevant Joints	Initial Score	Score with improved system
Arm stretches	Shoulder, Elbow, Wrist	94	98
Lateral Stretch	Shoulder, Elbow, Spine (shoulder, mid, base), Hip	91	95
Squats	Hip, Knee, Foot, Spine (mid, base)	75	86
Arm circles	Shoulder, Elbow, Wrist	65	92
Cross-body bicycle	All	37	81
Tuck jump	All	29	78

We can see from the table above that the results are much more stable and the system is much more helpful for a basic trainee. The results proves that the individual body-part feedback and the text correction messages (Figure 5.30) are really great improvements in our system.

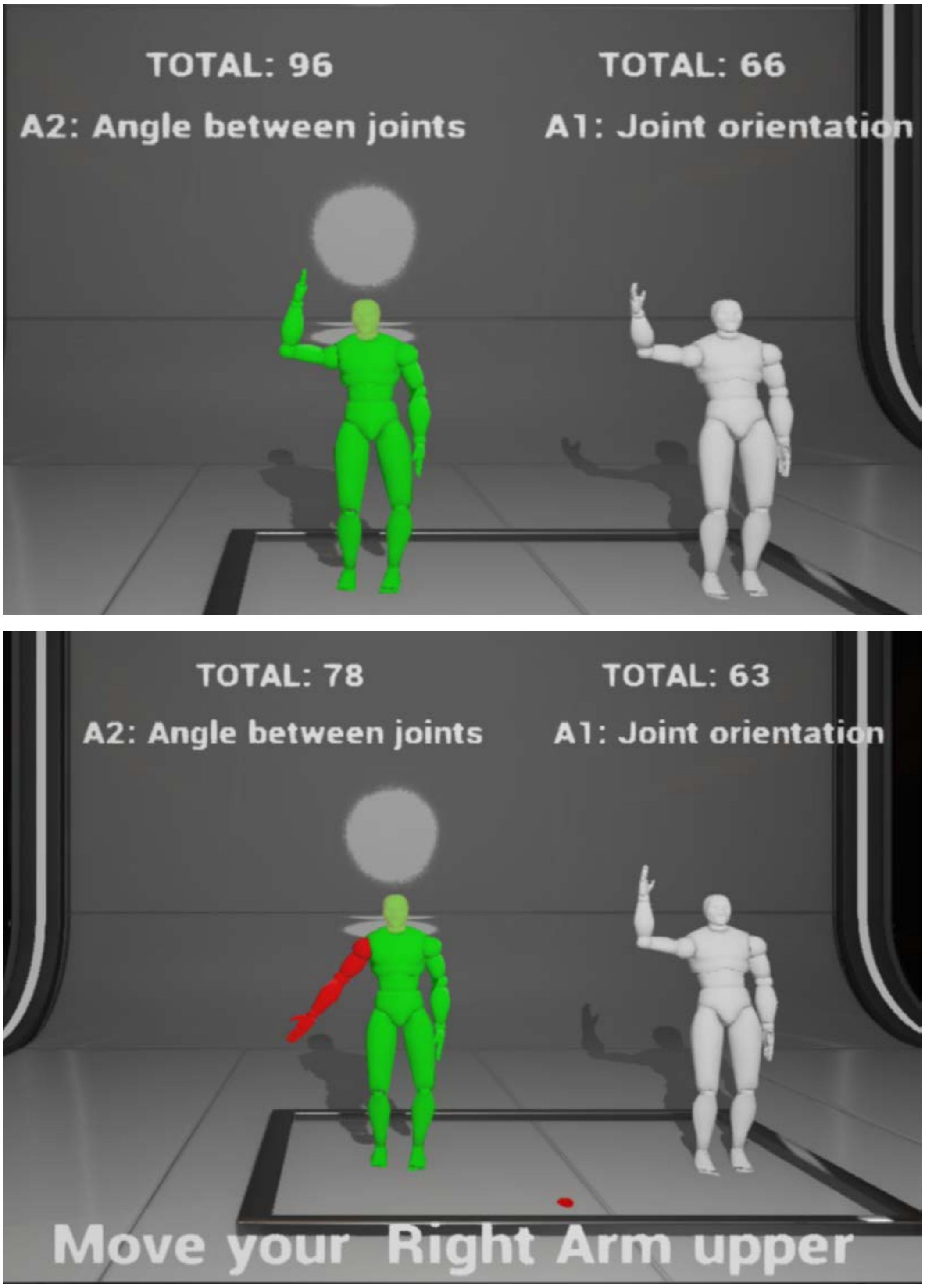


Figure 5.30. Imitation of a model with improved relevant total score and body-part movement highlighting and real-time text correction feedback a) matched movement b) wrong movement with body part highlight and text correction message

5.3.4 CONCLUSIONS

In this chapter, we presented most of the human motion tracking techniques and systems currently used, with its advantages and disadvantages. A brief explanation about the Kinect V2 features for body tracking (joints and orientation) and some related approaches used for body tracking in different applications are also presented.

We proposed an advanced virtual sports training system using the new Kinect V2 sensor and we implemented and presented a working prototype with promising experimental results. For building our system, we described a novel and simple approach by using the Unreal Engine 4 graphics engine alongside the Kinect 4 Unreal plugin. The experimental results showed that the system can be further improved, but it validates the proposed approach for a 3D virtual reality sports training or medical rehabilitation application that can be used easily at home and removes the need for a personal instructor each time.

Our sports training system is currently built for static physical training, in a closed environment, but can also be extended to other kinds of sports in a very simple and flexible way, and then just record any kind of sports exercise by an expert sportsman and save it in the system so that a trainee can simply select which sports he wants and enjoys the training course.

CHAPTER 6. CONCLISIONS AND FUTURE WORK

6.1 THE ORIGINAL CONTRIBUTIONS OF THIS THESIS

This thesis sets out to build and develop four main software prototypes.

1. **In chapter two of the thesis, a real-time face detection system is proposed**. From the system previous simulation results, some conclusions related to the behavior and performance of the proposed face detection system could be drawn. A face detection algorithm for color images has been presented using a skin-tone color model and a feature invariant approach that is based on skin and edge information for face detection. Our method first used light correction steps to adjust the illumination of the input image. **Our main contribution was the use of a combination between the result of segmentation in RGB and HSV color spaces, the result of segmentation using the Elliptical model in YCbCr color space and edge information.** It overcomes the difficulty of detecting the low Luma and high-Luma skin tones by applying a nonlinear transform to the YCbCr color space. The result of the face detection based on skin color depends on the color model or models used for skin pixels classification. In our method of skin region segmentation, we use three color models for the detection of skin pixels: RGB, HSV, and YCbCr. The combination of these color models is more robust to the variety of lighting conditions, leading to better results than those obtained with only one color model. Our proposed approach is a very good algorithm for face detection from single face images. The algorithm can be used in real-time applications with indoor and outdoor images and for advanced useges like face recognition. The results of this research was published in [HA14].
2. The subject of the emotion expressions recognition by computers is becoming increasingly popular. **Chapter three contains a study of the Real Time Facial Emotion Recognition issues and proposes an approach to recognizing facial emotion expressions using Kinect system**. We published this study and the developed system prototype in [AH15].
3. Scientists agree on the fact that the human brain is the main source of emotion. However, the precise role of the brain is not clear. Many theories exist about the exact involvement of the brain. **Chapter four contains a study of human emotion expressions in connection with brain activity.** We presented main theories of emotions, various

approaches, methods for facial emotion expressions recognition, human brain and emotions, some parts of the limbic system and other brain regions that have an effect or are related to emotions.

We integrated in the system for facial emotions recognition, described in chapter 3, the capability to recognize the brain activity based on the facial expression. We used in our experiments an association between 8 emotional expressions and a corresponding brain activity. **Our system enables tracking, recording and analyzing facial features, to recognize in real time eight emotion expressions and associate each of them to the brain region that is active when this emotion appeared on the human face**. The system can be used with patients who suffered from a stroke or Alzheimer's, to analyze their behavior and if they recognize their visitors. We published the results of this research in [HE15] [HE16].

4. Chapter five presents the most human motion tracking techniques and systems currently used, with their advantages and disadvantages. Also, a brief explanation about the Kinect V2 features for body tracking (joints and orientation) and some related approaches used for body tracking in different applications.

   We have built a virtual sports training system capable of tracking, analyzing, comparing the biomechanics movements of trainee with two types of exercises, simple and complex, and helping the trainee in real time to improve his movements. The system is described in subchapter 5.3 of the thesis. It was designed to be a Virtual Reality based sports training software, which evaluates the trainee movements and compares them with the trainer movements. A number of requirements and criteria were formulated that the software would be evaluated against:

   1. Accurate joint orientation and angles between joints tracking data
   2. Robust 3D avatar recording and displaying
   3. Perfect 25 joints orientation and angles between joints position analysis and comparison algorithms
   4. Adaptability to variable movement patterns.

The latest version of the Kinect sensor is a real improvement from the first one, and can provide good support for real-time evaluation of medical rehab or physical training exercises. **We described a novel and simple approach to using Kinect V2 with the Unreal Engine 4 in sports fields that can also be used in rehabilitation.**

We presented a working system functioning in a 3D virtual reality with promising experimental results, validating the proposed approach for sports training or medical rehab exercises in a virtual reality environment, removing the need for a real instructor each time. The professional trainer will only record the movements and then supervise the users in their training or rehabilitation.

This prototype would allow expanding current research into sports training and medical rehabilitation systems by providing additional biomechanical data, adding more information to the comparison algorithms. The accuracy of estimations and results can lead to more efficient and viable predictive models. The results presented in chapter five were published in [HA15].

6.2 FUTURE WORK

We think of future research related to the system presented in the subchapter 5.3. Improvements must be considered for the body comparison and highlighting. For example, most relevant joints should be taken into consideration when displaying the text correction feedback and a better smoothing algorithm for joints could be implemented. Also, the skeletal animation could be further refined including more details like finger thumbs and could also provide dimensions recognition (for example, the avatar could have the exact height and shape as the trainee).

Although we have implemented a Kinect movement-based menu interaction, the system could be improved by using voice recognition (for many rehab trainees movements-based menu interaction could be hard to handle). Other enhancements could also be made to improve the user interaction with the system.

Medical rehabilitation is a special medical field that, unlike normal physical training, requires a more careful approach. The trainees in such conditions are usually partially impaired and have movement restrictions and other issues. All these facts must be considered when experimenting and building a system for rehab. Therefore, special experiments with partially impaired trainees should be made in order to further calibrate and improve the comparison algorithms and also the interface or the avatar animations.

References

[AA09] A. Ashraf, S. Lucey, F. Cohn, T. Chen, Z. Ambadar, M. Prkachin, and P. Solomon, ”The painful face: Pain expression recognition using active appearance models”, Image and Vision Computing, 27:1788–1796, 2009.

[AE09] A. Eklund, H. Ohlsson, M. Andersson, J. Rydell, A. Ynnerman, and H. Knutsson, “Using Real-Time fMRI to Control a Dynamical System by Brain Activity Classification,” Lecture Notes in Computer Science, Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI), vol. 5761, pp. 1000–1008, 2009.

[AF13] A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman and A. Blake, ”Real-time human pose recognition in parts from single depth images”, Machine Learning for Computer Vision book, vol. 411, pp 119-135, 2013.

[AH15] Alabbasi, H., F. Moldoveanu, and A. Moldoveanu, “Real Time Facial Emotion Recognition using Kinect V2 Sensor”, IOSR Journal of Computer Engineering (IOSR-JCE), Vol.17, Issue 3, Ver. II, pp 61-68, May- 2015.

[AK06] P. Aleksic, A. Katsaggelos, “Automatic facial expression recognition using facial animation parameters and multistream HMMs”. IEEE Transactions on Information Forensics and Security 1 (1): pp. 3-11, 2006.

[AK08] A. Koschan and M. John Wiley & Sons, Handbook “Digital Color Image Processing”, Inc., Publication, 2008.

[AK09] A. Krtshan,“ Evaluation of gabor filter parameter for image enhancement and segmentation“, M.Sc. thesis, the THAPAR university, July, 2009.

[AL] Al-aqrabawi, Mohammad and Dug, Fangfang, "Human Skin Detection Using Color Segmentation", internet projects, Web site:http://www2.pmf.fh-goettingen.de/~awolf2/Projekte/Face-Detection/Skin/color2/detection/Human/ Skin Detection.pdf.

[AL07] Alvarez, G. A., & Franconeri, S. L. “How many objects can you track? Evidence for a resource limited attentive tracking mechanism”, Journal of Vision, 7(13), 14 1-10, 2007.

[AL11] Albert Rizzo, Belinda Lange, and Evan A. Suma, “Virtual Reality and Interactive Digital Game Technology: New Tools to Address Obesity and

Diabetes", Journal of Diabetes Science and Technology, vol. 5, Issue 2, March 2011.

[AM07] A. Mota Ferreira, M. VelhoteCorreia, "Face Detection Based on Skin Color in Video Images with Dynamic Background", Springer-Verlag Berlin Heidelberg, LNCS 4633, pp. 254-262, 2007.

[AN] Ananya Mandal," What is the Hippocampus?", http://www.news-medical.net /health / Hippocampus-What-is-the-Hippocampus .aspx.

[AN07] An XHTML + MathML + SVG Profile. W3C, August 9, 2002. Retrieved on 17 March 2007.

[AN12] Anupama.H.S, N.K.Cauvery, and Lingaraju.G.M, "BRAIN COMPUTER INTERFACE AND ITS TYPES - A STUDY", International Journal of Advances in Engineering & Technology, 739 Vol. 3, Issue 2, pp. 739-745, May 2012.

[AN93] Andersson, R.L,"A real experiment in virtual environments: A virtual batting cage. Presence: Teleoperators and Virtual Environments", 2, 16-33,1993.

[AO90] A. Ortony, and T.J. Turner,"What's basic about basic emotions", Psychological Review, Vol 97(3), pp 315-331, Jul 1990.

[AR13] A. Reema. and S. Namrata., "Face Detection in Digital Images Using Color Space and Edge Detection Techniques" , international Journal of Advanced Research in Computer Science and Software Engineering, volume 3, Issue 6,June 2013.

[AS03] A. Singh1, D. Chauhan, A.ayank, "A Robust Skin Color Based Face Detection Algorithm", Taking Journal of Science and Engineering, Vol. 6, No. 4, pp. 227-234, 2003.

[AS14] A. Shingade and A. Ghotkar," Animation of 3D Human Model Using Markerless Motion Capture Applied To Sports", International Journal of Computer Graphics & Animation (IJCGA), vol.4, no.1, January 2014.

[AS92] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," Pattern Recognition, vol. 25, no. 1, pp. 65-77, 1992.

[AW11] A. Weiss, D. Hirshberg, and M. Black. Home 3d body scans from noisy image and range data. In Computer Vision (ICCV), 2011 IEEE International Conference on, pages 1951 –1958, Nov. 2011

[AY13] A.Youssef, S. F. Aly, A. Ibrahim, and A. Lynn, ”Auto-Optimized Multimodal Expression RecognitionFramework Using 3D Kinect Data for ASD TherapeuticAid”, International Journal of Modeling and Optimization, Vol. 3, No. 2, April 2013.

[AZ15] AZO SENSOR, Kal Kaur, “Motion Capture Sensor Systems “,2015, available online at: http://www.azosensors.com/Article.aspx?ArticleID=43#2.

[BA01] Bartlett, M. S., Donato, G., Movellan, J. R., Hager, J. C., Ekman, P. and Sejnowski, T. J., "Image representation for facial expression coding", supported by NIH Grant No. 1F32 MH12417-01, Web site:http://www.cnl.salk.edu/~marni/publications.html

[BC10] B C. Zhang and Z. Zhang, A survey of recent advances in face detection. Technical report, Microsoft Research, 2010.

[BE11] Belinda Lange, Chien-yen Chang, Evan Suma, Bradley Newman, Albert Rizzo, and Mark Bolas, “Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor”, In International Conference of the Engineering in Medicine and Biology Society, pages 1831–1834, Boston, Massachusetts, August 2011.

[BI] Billy Y.L., Li1 Ajmal S., Mian2 Wanquan Liu1and Aneesh Krishna1, “Using Kinect for Face Recognition Under Varying Poses, Expressions, Illumination and Disguise”, Curtin University, The University of Western Australia, Bentley, Western Australia Crawley, Western Australia.

[BL01] https://blaine901.wordpress.com/category/second-year-work/page/2/

[BL05] Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R. & Curio, G. (2005). The Berlin Brain-Computer Interface: Report from the Feedback Sessions. Fraunhofer FIRST, Tech Rep. 1.

[BO11] Bo, A. P. L., M. Hayashibe, and P. Poignet, “Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect. Boston, United States, 2011.

[BR10] Brijil Chambayil, Rajesh Singla, R. Jha, “Virtual Keyboard BCI using Eye blinks in EEG” IEEE 6th
International Conference on Wireless and Mobile Computing, Networking and Communications, pg 466- 470,2010.

[BR11] BRAIN INJURY INSTITUTE Lab. http://www.braininjuryinstitute.org/Brain-Injury-Types/Frontal-Lobe-Damage.html,2011.

[BR15] Bray, Tim (9 April 2003). "On Semantics and Markup — Taxonomy of Markup". www.tbray.org/ongoing. Retrieved 9 July 2015.

[CA] Carnegie Mellon University, the Robotic Institute , https://www.ri.cmu.edu /research_project _detail.html? project_id=465&menu_id=261

[CB04] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M. Lee, A. Kazemzadeh, S.B. Lee, U. Neumann, S. Narayanan Analysis of Emotion Recognition using Facial Expression, Speech and Multimodal Information, ICMI'04, PY, USA, 2004.

[CH12] Chang, C.-Y.et al., 2012. Towards Pervasive Physical Rehabilitation Using Microsoft Kinect. Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 159-162.

[CL04] C. L. Lisetti, D. E. Rumelhart ―Facial Expression Recognition Using a Neural Network‖ In Proceedings of the 11 th International FLAIRS Conference, pp. 328—332, 2004.

[CL12] Clark, R. A. et al., 2012. The validity of the Microsoft Kinect for assessment of postural control. Gait & Posture, 36(3), pp. 372-377.

[CL87] C. Lange, (1887). Ueber Gemuthsbewgungen. 3, 8.

[CO06] Coan,Touch and Emotion available online at http://alinenewton.com/neuroscience-of-touch-touch-and-the-brain/, 2006.

[CO15] COMPUTER GRAPHICS 2015, available online at http://graphics.berkeley.edu/people/jObrien-publications.html.

[CW27] C. Walter, "The James-Lange theory of emotions: A critical examination and an alternative theory", The American Journal of Psychology, pp 106–124, 1927.

[DA11] D. Alexiadis, P. Daras, P. Kelly, N. E. O'Connor, T. Boubekeur, and Maher Ben Moussa, "Evaluating a Dancer's Performance using Kinect- based Skeleton Tracking", the 19th ACM international conference on Multimedia, November 2011, Pages 659-662.

[DA12] Damasio A, Damasio H, Tranel D, "Persistence of feelings and sentience after bilateral damage of the insula", Cerebral Cortex.

2012.doi:10.1093/cercor/bhs077.

[DC01] D. Chai, S.Lam and A. Bouzerdoum, "Skin Color Detection For Face localization in human-machine communications", School of Engineering and Mathematics, Edith Cowan University, Perth, Australia, 2001.

[DC10] D. Coyle, T. Martin McGinnity, G. Prasad,“Improving the separability of multiple EEG features for a BCI by neural-time-seriesprediction-preprocessing“, Biomedical Signal Processing and Control, Vol. 5, No. 3, 2010, pp. 196-204.

[DD09] D. DUAN, W. XIA, X. ZHANG and L.YAN, “A Novel Method of Face Detection Based on Fusing YCbCr and HIS Color Space”, IEEE International Conference on Automation and Logistics Shenyang, China August 2009.

[DD13] D. Dandotiya, R. Gupta, S. Dhakad and Y. Tayal," A survey paper on Biometric based Face detection techniques", International Journal of Software and Web Sciences (IJSWS), ppt70, 2013.

[DG13] D. Ghimire, J, Lee, ”A Robust Face Detection Method Based on Skin Color and Edges”, J Inf Process Syst, Vol.9, No.1, March 2013, pp. 141-156.

[DK00] D. Keltner, P. Ekman, „Facial expression of emotion“, In Lewis, M., Haviland-Jones, J.M. (eds.) Handbook of Emotions, pp. 236–249. Guilford Press, New York (2000).

[DM13] D. McDuff, R. Kaliouby, D. Demirdjian, and W. Rosalind, ”Predicting online media effectiveness based on smile responses gathered over the internet. In IEEE International Conference on Automatic Face and Gesture Recognition, 2013.

[DN09] D. Noonan, P. Mountney, D. Elson, A Darzi and G.Zhong Yang, “A Stereoscopic Fibroscope for Camera Motion and 3D Depth Recovery During Minimally Invasive Surgery”. In proc ICRA 2009, pp. 4463-4468.

[DO] https://docs.unrealengine.com/latest/INT/Engine/.

[DO01] Douglas Chai, Son Lam Phung and Abdesselam Bouzerdoum, "Skin Color Detection For Face localization in human-machine communications", School of Engineering and Mathematics, Edith Cowan University, Perth, Australia,2001.

[DR11] D.Riek Laurel, P. Robinson, ”Using robots to help people habituate to visible disabilities”, 2011.

[DS12] D. Singh Raghuvanshi, D. Agrawal, “Human Face Detection by using Skin Color Segmentation, Face Features and Regions Properties”, International Journal of Computer Applications, Volume 38– No.9, January 2012.

[DV07] D. Vlasic, R. Adelsberger, G. Vannucci, J. Barnwell, M. Gross, W. Matusik, and J. Popovic. Practical motion captures in everyday surroundings. ACM Trans. Graph., 26(3), 2007.

[EA02] E. Acosta, L. Torres and A. Albiol“ An automatic face detection and recognition system for video indexing applications“ IEEE CONFERENCE, Acoustics, Speech, and Signal Processing (ICASSP), vol.4, may- 2002.

[EA13] Edward Alcamo,” Anatomy Coloring Workbook”, second edition, The Princeton Review. pp. 120–. ISBN 978-0-375-76342-7. Retrieved 10 January 2013.

[EH71] E. Hill, J. Faller, H. Hill. "New Experimental Test of Coulomb's Law: A Laboratory Upper Limit on the Photon Rest Mass". Physical Review Letters, Vol: 26, pp: 721-724. 1971.

[EM14] Emotion Markup Language (EmotionML) 1.0, W3C Recommendation 22 May 2014, http://www.w3.org/TR/2014/REC-emotionml-20140522/

[EO97] E. Osuna, R. Freund, and F. Girosi,”Training Support VectorMachines: An Application to Face Detection”, Proc. IEEE Conf.Computer Vision and Pattern Recognition, pp. 130-136, 1997.

[ER11] Erik Andreas Larsen, ”Classification of EEG Signals in a Brain- Computer Interface System”, Norwegian University of Science and Technology, Department of Computer and Information Science, Master of Science in Computer Science, June 2011.

[ER06] E. R. Bachmann, R. B. McGhee, X. Yun, and M. J. Zyda, “Inertial and magnetic posture tracking for inserting humans into networked virtual environments, In Proceedings of the ACM symposium on Virtual reality software and technology, pages 9–16, 2006

[EX00] E. Foxlin and M. Harrington. Weartrack,”A self-referenced head and hand tracker for wearable computers and portable VR”, In Proceedings of the 4th IEEE International Symposium on Wearable Computers, pages 155–162, 2000.

[EX14] "Extensible Markup Language (XML)". W3.org. Retrieved 2014-06-28.

[FA11] F. Alizadeh, S. Nalousi and C. Savari," Face Detection in Color Images using Color Features of Skin ", World Academy of Science, Engineering and Technology 52 2011.

[FB14] F. Burkhardt, C. Becker-Asano, E. Begoli, R. Cowie, G. Fobe, P. Gebhard, A. Kazemzadeh, I. Steiner, T. Llewellyn: Application of EmotionML. Proceedings of the 5th International Workshop on Emotion, Sentiment, Social Signals and Linked Open Data (ES3LOD), 2014.

[FC09] F. Cohn, T. Simon Kruez, I. Matthews, Y. Yang, M. Nguyen, M. Tejera Padilla, F.Zhou, and F.De la Torre, ”Detecting depression from facial actions and vocal prosody”, In Affective Computing and Intelligent Interaction, 2009.

[FG08] F. Galán, M. Nuttin, E. Lew, P. Ferrez, G. Vanacker, J. Philips, J. Millán, “A brain actuated wheelchair: Asynchronous and non-invasive brain-computer interfaces for continuous control of robots,” Clinical Neurophysiology, vol. 119, no. 9, pp. 2159-2169, 2008.

[FG12] F. Gasparini, R. Schettini, ”Skin segmentation using multiple thresholding”, DISCO, UniversitàdegliStudi di Milano-Bicocca, Via BicoccadegliArcimboldi 8, Milano, Italy, 2012.

[FI14] Final Report of the Emotion Markup Language Incubator Group, 22 May 2014, available online, http://www.w3.org/TR/2014/REC-emotionml-20140522/

[FL] L. Franco, Alessandro Treves ―A Neural Network Facial Expression Recognition System using Unsupervised‖ Local Processing, Cognitive Neuroscience Sector – SISSA.

[GE11] GERRY VASSAR,:HOW DOES ANGER HAPPEN IN THE BRAIN?”, lakeside connect,2011.

[GG] G. Gaurav, B. Samarth, Mayank, V. and Richa, S.,”On RGB-D Face Recognition using Kinect”, IIIT Delhi.

[GO] G-O-M, ATOS and PONTOS 3D Laser scanners, http://www.gom.com/DE/index.html.

[GR12] G R. Vineetha, C. Sreeji , and J.Lentin,“Face Expression Detection Using Microsoft Kinect with the Help of Artificial Neural Network”, Trends in Innovative Computing 2012 - Intelligent Systems Design.

[GS11] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, “A dynamic approach to the recognition of 3D facial expressions and their temporal models," in Proc. 9th IEEE International Conference on AutomaticFace and Gesture Recognition, March 2011.

[GS12] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. "Static and dynamic3D facial expression recognition: A comprehensive survey," Image and Vision Computing, Oct. 2012.

[GW91] G. W. Cottrell and J. Metcalfe, "EMPATH: Face, gender, and emotion recognition using holons," Advances in Neural Information Processing Systems, vol. 3, pp. 564-571, 1991.

[GX13] G. Xiaosi, R. Patrick Hof, j. Karl Friston, and Jin Fan,"Anterior Insular Cortex and Emotional Awareness", The Journal of Comparative Neurology | Research in Systems Neuroscience 521:3371–3388, 2013.

[HA10] Hans J. Markowitsch and Angelica Staniloiu, "Amygdala in action: Relaying biological and social significance to autobiographical memory", 2010.

[HA12] H. Ali Monash, "Motion Comparison using Microsoft Kinect", computer science project, MONASH UNIVERSITY, 2012.

[HA14] H. ALABBASI, F. Moldoveanu," Human face detection from images based on skin color", 18th International (ICSTCC), October 2014.

[HA15] Hesham ALABBASI, Alex Gradinaru, Florica Moldoveanu, and Alin Moldoveanu "Human Motion Tracking & Evaluation using Kinect V2 Sensor", presented to the 5th IEEE International Conference on E-Health and Bioengineering - EHB November 2015, Iasi, Romania (in press).

[HD] https://dev.windows.com/en-us/kinect/develop.

[HE15] Hesham. ALABBASI, Florica Moldoveanu, Alin Moldoveanu and Zaid Shhedi," Facial Emotion Expression Recognition With brain activities Using Kinect Sensor V2", International Research Journal of Engineering and Technology (IRJET), vol. 02, Issue 02, May-2015, pp 421- 428.

[HE16] Hesham. ALABBASI, Alex Gradinaru, Florica Moldoveanu, and Alin Moldoveanu "Virtual sports training system using Kinect V2 sensor", University POLITEHNICA of Bucharest Scientific Bulletin, C series: Electrical Engineering and Computer Science, 2016, accepted paper

[HH] https://dev.windows.com/en-us/kinect/hardware.

[HK09] H. Kanetaka, S. Yabukami, S. Hashi and K.Ichi Arai, "Wireless magnetic motion capture system for medical use”, Chapter Interface Oral Health Science, pp 329-331, 2009.

[HM] https://msdn.microsoft.com/en-us/library/microsoft.kinect.Jointtype .aspx.

[HO10] Hoffmann J, Brüggemann B, Kruger B. Automatic Calibration of a Motion Capture System based on Inertial Sensors for Tele-Manipulation. In proceedings of 7th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Funchal - Madeira, Portugal, June 2010.

[HP12] H. POURGHASSEM S.NASEHI AND, ”An Optimal EEG-based Emotion Recognition Algorithm Using Gabor Features” WSEAS TRANSACTIONS on SIGNAL PROCESSING, Issue 3, Volume 8, July 2012.

[HR15] H. Rimlawi Malone, “Hypothalamus Anatomy” **, MD** Resident Physician, Department of Neurosurgery, Columbia University College of Physicians and Surgeons, 2015.

[HT04] H. Takahashi, N. Yahata, M. Koeda, T. Matsuda, K. Asai, Y. Okubo, “Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study“, NeuroImage, Vol. 23, No. 4, 2004, pp. 967-974.

[HU05] Hugo D. Critchley, Pia Rotshtein, Yoko Nagai, D. John , Christopher J., Mathias,c and Raymond J. Dolan, “ Activity in the human brain predicting differential heart rate responses to emotional facial expressions”, NeuroImage 24,2005, PP. 751– 762.www.elsevier.com/locate/ynimg.

[HU10] Humintell’s scientifical, “the seven basic emotions: Do you know them?“, website: http://www.humintell.com/2010/06/the-seven-basic-emotions-do-you-know-them/#sthash.Xrb3V7ky.dpuf

[HY13] Hyunjin Yoon, Sang-Wook Park, Yong-Kwi Lee, and Jong-Hyun Jang, “ Emotion Recognition of Serious Game Players Using a Simple Brain Computer Interface”, 978-1-4799-0698-7, ICTC 2013.

[IB07] I. Baran and J. Popovi,”Automatic rigging and animation of 3d characters”, ACM Transactions on Graphics (TOG), SIGGRAPH 2007, vol.26, issue 3, July 2007

[IK05] I. Kim, J. Hyung Shim, and J. Yang, Stanford University,"Face detection", 2005.

[IK07] I. Kotsia, I. Pitas, “ Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines” in IEEE

Transactions on Image Processing 16 (1): pp.172- 187, 2007.

[ION10] Ion Marqu´es,"Face Recognition Algorithms", Proyecto Fin de Carrera, June 16, 2010.

[IW] Internet website, https://en.wikipedia.org/wiki/Gabor_filter.

[JA09] Jacek De˛biec and Joseph LeDoux, “The Amygdala and the Neural Pathways of Fear“, Spingers, 2009.

[JA14] Jacob Whitehill, Zewelanji Serpell, Yi-Ching Lin, Aysha Foster, and Javier R. Movellan,”The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions”, IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 5, NO. 1, JANUARY-MARCH 2014

[JA72] Jarwin, C. (1872). The Expression of the emotions in man and animals. London: John Murray.

[JA84] James, W. (1884). What is emotion? Mind, ix, 189.

[JA90] James, W. (1890). The Principles of Psychology.

[JC] J. Crowley, F. Berard, "Multi-model Tracking of faces for Video Communications".

[JC00] J. Cassell, N. Nudge wink, “ elements of face-to-face conversation for embodied conversational agents, volume 1, chapter 1, pages 1–27. MIT Press, 2000.

[JC11] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura. A virtual reality dances training system using motion capture technology. IEEE Trans. Learn. Technol., 4(2):187–195, 2011.

[JF98] J. F. Cohn, A. J. Zlochower, J. J. Lien, and T. Kanade, “Feature point tracking by optical flow discriminates subtle differences in facial expression,” in Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 396-401, 1998.

[JK14] J. Kalat and M. Shiota,(2007). Belmont, CA: Thomson Wadsworth from: http://en.wikiversity.org/wiki/Motivation_and_emotion/Book/2014/Facial_Action_Coding_System.

[JL15] J. Lee, M. Hong, and S. Ryu,“Sleep Monitoring System Using Kinect Sensor”, Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks , Article ID, April 2015.

[JO06] John M. Allman, Atiya Hakeem, Joseph M. Erwin, Esther Nimchinsky and

Patrick Hof, “The Anterior Cingulate Cortex, The Evolution of an Interface between Emotion and Cognition“, Article first published online: 25 JAN 2006 ,DOI: 10.1111/j.1749-6632.2001.tb03476.

[JO14] JointType Enumeration, 2014, https://msdn.microsoft.com/en-us/library/microsoft.kinect.Jointtype .aspx.

[JP03] J. Prinz, “Emotions Embodied”, the University of North Carolina at Chapel Hill, 2003.

[JQ12] J. Qiang-rong and L.Hua-lan, "Robust Human Face Detection in Complicated Color Images", IEEE journal, 2012.

[JR00] J. R. Wolpaw *et al.*, “Brain-computer interface technology: A review of the first international meeting,”*IEEE Trans. Rehab. Eng.*, vol. 8, pp.164–173, June 2000.

[JS05] J. Stam, “Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field,” Clinical Neurophysiology, vol. 116, 2005, pp. 2266-2301.

[JS11] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio. "Real-Time Human Pose Recognition in Parts from Single Depth Images". Computer Vision and Patter Recognition. 2011.

[JS13] J. Shotton. et al.,”Real-time human pose recognition in parts from single depth images”, Communications of the ACM, pp. 116-124, 2013.

[JU] Júlio Rocha do Amaral, MD & Jorge Martins de Oliveira, MD, PhD, “Limbic System: The Center of Emotions”, Neurobiology article on The Limbic System, The Healing Center On-Line , available at: http://www.healing-arts.org/n-r-limbic.htm.

[JZ11] J. Zhang, G. Sudre, X. Li, W. Wang, D. Weber, “Clustering Linear Discriminant Analysis for MEG-based Brain Computer Interfaces, IEEE Trans. on Neural Systems and Rehabilitation Engineering, No. 99, 2011, pp. 1-8.

[KH09] K. Hook, ” Affective loop experiences: designing for interactional embodiment”, Philosophical Transactions of the Royal Society B: Biological, Sciences, 364(1535), 2009.

[KH11] K.Hawari, J. Ma, R. Xiao “An Innovative Face Detection based on Skin Color Segmentation”, International Journal of Computer Applications (0975 – 8887) Volume 34– No.2, November 2011.

[KI14] Kinect hardware, 2014, https://dev.windows.com/en-us/kinect/hardware.

[KO04] Konstantinos Pehlivanis & Maria Papagianni & Athanasios Styliadis, "VIRTUAL REALITY & LOGISTICS", International Conference on Theory and Applications of Mathematics and Informatics – ICTAMI, Thessaloniki, Greece, 2004.

[KO11] Korpela, Jukka (2005-11-16). "Programs vs. markup". IT and communication. Tampere the University of Technology. Retrieved 2011-01-08.

[KY13] K.Yun Yeung, T. Ho Kwok, and C.L. Wang, "Improved Skeleton Tracking by Duplex Kinect: A Practical Approach for Real-Time Applications", J. Comput. Inf. Sci. Eng 13(4), Oct 2013.

[LA12] Lance, B.J.; Kerick, S.E.; Ries, A.J.; Oie, K.S.; McDowell, K. , "Brain–Computer Interface Technologies in the Coming Decades," Proceedings of the IEEE , vol.100, no. Special Centennial Issue, pp.1585-1599, May 13 2012. doi: 10.1109/JPROC.2012.2184830.

[LC11] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura, "A virtual reality dances training system using motion capture technology", IEEE Trans. Learn. Technol., 4(2):187–195, 2011.

[LD09] L. Dickholtz. "Magnetion Motion Capture Systems", MetaMotion. 2009. Available: http://www.metamotion.com/motion-capture/magnetic-motioncapture-1.htm.

[LI03] L.I. Aftanas, A.A. Varlamov, N.V. Reva, and S.V. Pavlov, „Disruption of early event-related theta synchronization of human EEG in alexithymics viewing affective pictures", Neuroscience Letters, Vol. 340, No. 1, 2003, pp. 57-60.

[LI05] List of use cases for an Emotion markup language http://www.w3.org/2005/Incubator/emotion/XGR-emotion/#AppendixUseCases.

[LI11] Lia Stannard,"What Parts of the Human Brain Correspond to Emotion or Love?", article, may 04,2011, available at: http://www.livestrong.com/article/77419-parts-human-brain-correspond-emotion/.

[LI13] Lin, Y.-H.et al., 2013. A Kinect-Based System for Golf Beginners Training.

Information Technology Convergence, pp. 121-129.

[LJ06] L. J. Olson, E. and S. Teller. Robust rangeonly beacon localization. Journal of Oceanic Engineering, 31(4):949–958, 2006.

[LL05] L. Leocani, M. Rovaris, F. Martinelli-Boneschi, P. Annovazzi, M. Filippi, B. Colombo, V. Martinelli, and G. Comi,“Movement preparation is affected by tissue damage inmultiple sclerosis: Evidence from EEG event related desynchronization“, Clinical Neurophysiology, Vol. 116, No. 7, 2005, pp. 1515-1519.

[LU12] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil,” Brain Computer Interfaces, a Review”, Sensors 2012, 12, 1211-1279; doi:10.3390/s120201211

[MA11] Mazen A Kheirbek and René Hen, " Dorsal vs Ventral Hippocampal Neurogenesis: Implications for Cognition and Mood”,Neuropsychopharmacology, 2011, pp. 373–374.

[MA60] M. Arnold. Emotion and personality. Columbia University Press, 1960.

[MA95] Mark S, et al., “Brain Activity During Transient Sadness and Happiness in Healthy Women“, psychiarity, 1995.

[MC14] M. Chauhan, M. Sakle, “Study & Analysis of Different Face Detection Techniques“, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), PP 1615-1618, 2014.

[ME] Merriam-Webster Dictionary language markup language.

[ME11] Merriam-Webster, I. (2011, August 03). Emotion - Definition and More from the Free Merriam-Webster Dictionary. Retrieved August 3, 2011, from Merriam-Webster: http://www.merriam-webster.com/dictionary/emotion.

[ME13] M. Eltaher, H. Wan and J. Lee,” Motion Detection Using Kinect Device for Controlling Robotic Arm”, 2013 ASEE Northeast Section Conference, Reviewed Paper, March 14-16, 2013.

[ME63] http://mewarnai.us/634655-gallery-of-muscles-of-head-and-neck

[MG13] M. Girard, F. Cohn, H. Mahoor, S. mohammad and Dean P. Rosenwald,” Social risk and depression: Evidence from manual and automatic facial expression analysis”, In IEEE International Conference on Automatic Face and Gesture Recognition, 2013.

[MH02] M.-H. Yang, D. Kriegman, and N. Ahuja, “Detecting faces in images: A survey,” IEEE Trans. On PAMI, vol.24, pp. 34–58, January 2002.

[MH12] M. Hofstetter. “Temporal Pattern-Based Active Marker Identification and Tracking Using a Dynamic Vision Sensor”. Master thesis. Institue of Neuroinformatics, ETH Zurich, 2012.

[MI] Michael Downes. "TEX and LATEX 2e".

[MI] Micro Expression Training, Online Micro Expression Training Using The Official Paul Ekman Tools, website: http://changingminds.org/explanations/emotions/facial_emotions.htm.

[ME14] Mepits, “BCI - Brain Computer Interface”, 11 August, 2014, available at: https://www.mepits.com/tutorial/173/Electronics-Devices/BCI---Brain-Computer-Interface.

[MI11] Michael S. Fanselow and Hong-wei Dong,” Are the Dorsal and Ventral Hippocampus Functionally Distinct Structures?”, Neuron. Author manuscript; available in PMC 2011 Jan 14.

[MI14] Microsoft Company, 2014, https://msdn.microsoft.com/en-us/library/microsoft.kinect.Jointtype .aspx.

[MJ13] M. Johanna Theresia Veltmaat, “Recognizing activities with the Kinect A logic-based approach for the support room”, Master thesis Artificial Intelligence, Radboud University Nijmegen, July 2013.

[ML98] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with Gabor wavelets,” in Proc. Third IEEE Conf. Face andGesture Recognition, pp. 200-205, Nara, Japan, Apr. 1998

[MM] M. Motion. Gypsy. http://www.motion-capturesystem.com.

[MM06] M. Marin, “Emotion and Stress”, Chapter 9 Psy.D. Adjunct Professor of Psychology, 2006.

[MM14] M. Modi, F. Macwan“Face Detection Approaches: A Survey“, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 4, April 2014.

[MO10] Morten l. kringelbach and kent c. Berridge, “the neurobiolo gy of pleasure and happiness“ GLYPH OUP, CHAPTER 2,OCTOBER-2010.

[MO12] MOTION CAPTURE, available online at http://motion--capture.blogspot.ro/2012/03/time-modulated-active-marker.html.

[MO13] Molly McAdams, ”What Are the Functions of Frontal Lobe of Brain?”, Aug

16, 2013, available at : http://www.livestrong.com/article/24256-functions-frontal-lobe-brain/.

[MP00] M. Pantie and L. J. M. Rothkrantz, “Automatic analysis of facial expressions: the state of the art,”IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424-1445, 2000.

[MP08] M. Pantic, “Facial Expression Recognition”, Imperial College London, London, UK; University of Twente, AE Enschede, The Netherlands, 2008.

[MP13] M. Puica,”Towards a Computational Model of Emotions for Enhanced Agent Performance”, Ph.D thesis, University Politehnica of Bucharest, Romania, 2013.

[MS] M. Schrooder, I. Wilson, W. Jarrold, D. Evans, C. Pelachaud, E. Zovato and K.Karpouzis, “What is most important for an Emotion Markup Language?”

[MS03] M. Stewart Bartlett, G. Littlewort, I. Fasel, R. Jvier Movellan, “Real Time Face Detection and Facial Expression Recognition: Development andApplications to Human Computer Interaction”, Computer Vision and Pattern Recognition Workshop, 2003 (CVPRW '03), DOI: 10.1109/CVPRW.2003.10057.

[MS11] M.SHARIF, A. KHALID, M. RAZA, and S. MOHSIN,“ Face Recognition using Gabor Filters“, Journal of Applied Computer Science & Mathematics, no. 11, pp 53-57,/2011.

[MS12] M. Schröder, P. Baggia, F. Burkhardt, C. Pelachaud, C. Peter, and E. Zovato: EmotionML - an upcoming standard for representing emotions and related states, ACII Affective Computing and Intelligent Interaction, 2012.

[MS14] M. Schröder, Paolo Baggia, Felix Burkhardt, Catherine Pelachaud, Christian Peter, Enrico Zovato: Emotion Mark up Language. In R.A. Calvo, S.K. D'Mello, J. Gratch and A. Kappas (Eds). Handbook of Affective Computing. Oxford University Press, 2014.

[MY04] Myers, D. G. (2004). Theories of Emotion. Psychology: Seventh Edition. New York, NY: Worth Publishers.Of the emotional state. Psychological Review, 69(5):379–399, 1962.

[NA03] N.A. Shapira et al, “lBrain Activation by Disgust-Inducing Pictures in Obsessive-Compulsive Disorder“, BIOL PSYCHIATRY ,2003.

[NO] Noraxon-Inc, http://www.noraxon.com/index.php3.

[OP] OptiTrack http://www.optitrack.com/products/prime-17w/.

[OP15] Optical Tracking Markers http://www.ps-tech.com/optical-trackers/accessories/optical-tracking-markers,2015.

[OQ] http://www.opaquemultimedia.com/kinect-4-unreal/#about-k4u.

[OR] Organic Motion.

[OR03] Ortony, Andrew (2003). On making believable emotional agents believable. In et al., R. Trappl, editor, Emotions in Humans and Artifacts, pp. 189.212. MIT Press.

[OR88] Ortony, Andrew, G.L. Clore, and A. Collins (1988). The cognitive structure of emotions. Cambridge University Press, Cambridge, MA.

[PA12] Panagiotis C. Petrantonakis and Leontios J. Hadjileontiadis, "Adaptive Emotional Information Retrieval from EEG Signals in the Time-Frequency Domain," IEEE Trans. Signal Processing, vol. 60, no. 5, pp. 2604-2616, 2012.

[PC10] P.C. Petrantonakis, L.J. Hadjileontiadis,"Adaptive extraction of emotion-related EEG segments using multidimensional directed information in the time-frequency domain, Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2010, pp. 1 – 4.

[PE13] Peter B. Williams, "Your Brain on Happiness: The Neuroscience of Joy", 2013.

[PH] PhaseSpace, I., http://www.phasespace.com/snm.html.

[PH05] Philippe R, et al., "The neural bases of amusement and sadness: A comparison of block contrast and subject-specific emotion intensity regression approaches", ELSEVIER, 2005.

[PI11] Pierre Beauchamp, and Jocelyn Faubert, "Visual Perception Training: Cutting Edge Psychophysics and 3D Technology Applied to Sports Science", High-Performance CIRCuit e-Journal, In Press, 2011.

[PK02] P. Ekman, Friesen, W. V., & Hager, J. C. (2002). Facial Action Coding System Investigator's Guide. Retrieved from: http://face-and- emotion.com/ dataface/ facs/guide / FACSIVTi.html.

[PK03] P. Ekman, Friesen, W. V, „Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions Paperback", September 1, 2003.

[PK06] P. Kakumanu, N. Bourbakis, "A Local- Global Graph Approach for Facial Expression Recognition", ICTAI, pp 685-692, 2006.

[PK81] P. Kleinginna and A. Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. Motivation and Emotion, 5(5):345– 379, 1981.

[PL13] P. Lemaire, L. Chen, M. Ardabilianand M. Daoudi, ” Fully Automatic 3D Facial Expression Recognition using Differential Mean Curvature Maps and Histograms of Oriented Gradients”, Workshop 3D Face Biometrics, IEEE Automatic Facial and Gesture Recognition, Apr 2013, Shanghai, China from:https://hal.archives-ouvertes.fr/hal-00823903/document.

[PN11] P. Nogueira, “Motion Capture Fundamentals A Critical and Comparative Analysis on Real-World Applications”, Faculdade de Engenharia da Universidade do Porto, Programa Doutoral em Engenharia Informática Instituto de Telecomunicações, 18 de Novembro de 2011.

[PR09] Professor Anthony A. Wright, Ph.D.,” Homeostasis and Higher Brain Functions” book, “Limbic System: Amygdala”, chapter 6, Department of Neurobiology and Anatomy, The UT Medical School at Houston, 2009.

[PR11] P. Redding, "Feeling, thought and orientation: William James and the idealist anti-Cartesian tradition" , Parrhesia 13, pp 41–51, 2011.

[PV04] P. Viola and, M. Jones,“ Robust Real-Time Face Detection“, International Journal of Computer Vision 57(2), 137–154, 2004.

[PW04] P. Wright, G. He, N. A. Shapira, W. K. Goodman and Y. Liu1,“Disgust and the insula: fMRI responses to pictures of mutilation and contamination“, Lippincott Williams & Wilkins, Vol 15 No 15, 25 October 2004.

[RA07] Ranganatha Sitaram, Andrea Caria, Ralf Veit, Tilman Gaber, Giuseppina Rota, Andrea Kuebler and Niels Birbaumer, "FMRI Brain–Computer Interface: A Tool for Neuroscientific Research and Treatment, 2007.

[RA12] R. Asiimwe and A. Anvar, “Automation of the Maritime UAV command, control, navigation operations, simulated in real-time using Kinect sensor: A feasibility study”, World Academy of Science, Engineering and technology, 2012.

[RB07] R. Poppe. Vision-based human motion analysis: An overview. Comput. Vis. Image Underst., 108(1-2):4– 18, 2007

[RG98] R. Gonzales, and R. Woods,, "Digital image processing", Addison-Wesley, 1998.

[RI13] Richard L. Wilson, “ Ethical Issues of Brain Computer Interfaces (BCI)”, IACAP Proceedings, 2013.

[RL15] R. Lun and W. Zhao,” A Survey of Applications and Human Motion Recognition with Microsoft Kinect”, International Journal of Pattern Recognition and Artificial Intelligence, vol. 29, Issue 05, August-2015.

[RL91] R. Lazarus. Emotion and Adaptation. Oxford University Press, 1991.

[RM03] R. Meir and G. R¨atsch,”An introduction to boosting and leveraging”, Springer-Verlag Berlin Heidelberg, pp 118–183, 2003.

[RO99] Ross Moore, Mathematics Department, Macquarie University, Sydney, 1999.

[RP11] R. Plutchik, "The Nature of Emotions", American Scientist. Retrieved 14 April 2011.

[RR12] R. Ramirez & Vamvakousis, Z, ”Detecting Emotion from EEG Signals Using the Emotive Epoc Device”, Brain Informatics Lecture Notes in Computer Science (pp. 175–184). Springer, 2012.

[PR14] Proudfoot M, et al. Pract Neurol, “ How to understand it: Magnetoencephalography”, Pract Neurol 2014;0:1–8. doi:10.1136/practneurol-2013-000768.

[RV10] R. Vijayanandh, G. Balakrishnan,”Human Face Detection Using Color Spaces and Region Property Measures”, 11th Int. Conf. Control, Automation, Robotics and Vision Singapore, 2010.

[RZ08] R. Zhi, Q. Ruan, "A Comparative Study on Region-Based Moments for Facial Expression Recognition,", in Congress, on Image and Signal Processing, Vol. 2, pp.600-604, 2008.

[SB11] S. Berretti, B. Ben Amor, M. Daoudi, and A. del Bimbo, “3D facialexpression recognition using SIFT descriptors of automaticallydetected keypoints”, The Visual Computer, vol. 27, no. 11, 2011.

[SB85] S. Baron-Cohen, M. Alan Leslie and Uta Frith, “Does the autistic child have a ”theory of mind”? Cognition, 21(1):37–46, 1985.

[SC08] Schalk, G; Miller, KJ; Anderson, NR; Wilson, JA; Smyth, MD; Ojemann, JG; Moran, DW; Wolpaw, JR; Leuthardt, EC (2008). "Two-dimensional movement control using electrocorticographic signals in humans". Journal of Neural Engineering.

[SC13] Schröder, M., Baggia, P., Burkhardt, F., Pelachaud, C., Peter, C. & Zovato, E., (2013), "Emotion Markup Language (EmotionML) 1.0 "

[SE03] Serruya MD, Donoghue JP. (2003) Chapter III: Design Principles of a Neuromotor Prosthetic Device in Neuroprosthetics: Theory and Practice, ed. Kenneth W. Horch, Gurpreet S. Dhillon. Imperial College Press.

[SE13] Sergio Ruiz, Korhan Buyukturkoglu, Mohit Rana, Niels Birbaumer and Ranganatha Sitaram, " Real-time fMRI brain computer interfaces: Self-regulation of single brain regions to networks", Biological Psychology article in press, Elsevier, 2013.

[SK08] S. Kannan,"Quantitative Analysis of Masticatory Performance in Vertebrates", in Mechanical and Aerospace Engineering, University at Buffalo, NY: Buffalo, 2008.

[SN10] Snell, Richard S., Clinical Neuroanatomy BOOK, "The Reticular Formation and the Limbic System", Chapter 9, 7th Edition, 2010.

[SN11] S. Nasehi, H. Pourghassem,"A New Feature Dimensionally Reduction Approach based on General Tensor Discriminant Analysis in EEG Signal Classification, IEEE International conference on Intelligent Computation and Bio-Medical Instrumentation (ICBMI2011), Wuhan, China, December 2011, pp. 188-191.

[SN12] S. Nasehi and H. Pourghassem, "Seizure Detection Algorithms Based on Analysis of EEG and ECG Signals: A Survey," Neurophysiology, Vol. 44, No. 2, 2012, pp. 174-186.

[SO] http://souvrsophie.blog.com/page/7/

[SO03] Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai, "Skin Segmentation using Color and Edge Information", IEEE, ISSPA 2003.

[SS62] S. Schachter and J. E. Singer, "Cognitive, social, and physiological determinants of emotional state", Psychological Review, 69, 379-399, 1962.

[ST88] Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. Journal of Personality and Social Psychology, 54(5), 768-777.

[SW12] Swapnil V Tathe and Andipan P Narote "Face Detection using Color Models" Proceedings of "Conference on Advances in Communication and Computing (NCACC 12)" Held at R.c.patel institute of Technology, Shirpur, Dist. Dhule, Maharashtra, India April 21, 2012.

[TA04] T. Ahonen, A. Hadid and M. Pietika¨inen,”Face recognition with local binary patterns”, In Proceeding of the European conference on computer vision (ECCV2004), pp 469–481, 2004.

[TA08] "Tags used in HTML". World Wide Web Consortium. November 3, 1992. Retrieved November 16, 2008.

[TA13] Taciana Saad Rached and Angelo Perkusich, “Emotion Recognition Based on Brain-Computer Interface Systems”, http://dx.doi.org/10.5772/56227, INTECH OPENSCIENCE, 2013.

[TB06] T. B. Moeslund, A. Hilton, and V. Kr¨uger. A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Underst., 104(2):90–126, 2006.

[TH08] THEODORE W. BERGER, JOHN K. CHAPIN, GREG A. GERHARDT, WALID V. OUSSOU, DAWN M. TAYLOR, and PATRICK A. TRESCO, “BRAIN-COMPUTER ,INTERFACES”, springer, 2008.

[TK02] T. Kanade, J. Cohn, and Y. Tian, “Comprehensive Database for Facial Expression Analysis,” in Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 45-63, 2000.

[TO97] Todorov, E., Shadmehr, R., & Bizzi, E.,” Augmented feedback presented in a virtual reality environment accelerates learning of a difficult motor task”, Journal of Motor Behavior, 29, 147-158, 1997.

[VG14] V. Gupta AND D. Sharma, ” A Study of Various Face Detection Methods”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 5, May 2014.

[VI] Vicon. Vicon motion capture system. www.vicon.com.

[VI10] Victoria J. Bourne, “How are emotions lateralised in the brain? Contrasting existing hypotheses using the chimeric faces test“, philpapers, Cognition and Emotion 24 (5): 903-911, 2010.

[VI15] VIRTUAL DICTIONARY available at : http://www.virtualworldlets .net/Resources/Dictionary.php? Term=Time%20modulated% 20active% 20 marker%20MoCap, 2015.

[W3] W3C Emotion Markup Language Incubator Group.

[W305] W3C Emotion Markup Language Incubator Group, website http://www.w3.org/2005/Incubator/emotion/.

[WA07] Wang, J, Yan, N, Liu, H, Liu, M, & Tai, C, “Brain-computer interfaces based on attention and complex mental tasks“, In Proceedings of the 1 st International Conference on Digital Human Modeling, ICDHM’07, Berlin, Heidelberg, 2007. Springer-Verlag., 467-473.

[WC27] W. Cannon,“Bodily Changes in Pain, Hunger, Fear, and Rage - an account of recent researches into the function of emotional excitement“, D. Appleton and Company, 1927.

[WE09] Wei-Yen Hsu, Yung-Nien Sun,“ EEG-based motor imagery analysis using weighted wavelet transform features“, Journal of Neuroscience Methods, Vol. 176, No. 2, 2009, pp. 310-318.

[WH15] http://whatis.techtarget.com/definition/virtual-reality,2015.

[WI] WiseGEEK,http://www.wisegeek.com/what-are-the-different-methods-for-tracking-human-motion.htm.

[WI03] Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G (October 2003). "Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust". Neuron 40 (3): 655–64.doi:10.1016/S0896-6273(03)00679-2. PMID 14642287.

[WJ84] W. James,“ What is an emotion? Mind“, (9):188–205, 1884.

[WO02] Wolpaw, J. R, Birbaumer, N, Mcfarland, D. J, Pfurtscheller, G, & Vaughan, T. M. Brain-computer interfaces for communication and control. Clinical Neurophysiology, March (2002). , 1, 767-791.

[WR09] W. Rosalind Picard,”Future affective technology for autism and emotion Communication”, Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 364(1535):3575–3584, Dec 2009.

[WR97] W. Rosalind Picard. Affective Computing. The MIT Press, 1997.

[XI14] Xiao-Wei Wang, Dan Nie, Bao-Liang Lu, Emotional State Classification from EEG Data Using Machine Learning Approach, Neurocomputing, vol. 129, pp. 94-106, 2014

[XS] Xsens. Xsens MVN. www.xsens.com.

[YA11] Yao-Jen Chang, Shu-Fang Chen, and Jun-Da Huang, “A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities", Research in Developmental Disabilities, 32(6):2566–2570, 2011.

[YT02] Y. Tian, T. Kanade, and J. F. Cohn, "Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity," in Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 229-234, 2002.

[ZH13] Zhang, L., Hsieh, J.-C. & Li, S., 2013. A Kinect Based Golf Swing Reorganization and Segmentation System. Pervasive Computing and the Networked World, pp. 843-847.