Research Paper on Expressive Facial Animation Synthesis by Learning Speech Co-articulation and Expression Spaces

Wu Wei
Art College of Henan University, Zhengzhou, Henan, China

ABSTRACT: In the current project, the techniques and different algorithms of the researchers are mentioned in the paper. The introduction is mentioned along with the common method to identify the facial expression which is also discussed in the paper. The literature review contains the historical researches by the authors regarding the 3D visualization and facial expression resignation. It was purposed to find the three different facial expressions of a person by dividing the face of a person into six sub-regions. The range of the markers was selected to be between 90 and 102. According to the markers or sub-region on the face, the expressions were identified which were red for the anger, green for the happiness and blue for the sadness. According to the data obtained and the results, it is identified and concluded that the 3D representation of the facial expression is more accurate than the single line representation.

Keywords: facial expression; sub-regions; 3D representation; skin color; connecting region; singular matrix; visualization

1 INTRODUCTION

The concept of face detection and realistic facial motions is increasing for many different purposes like intelligence companies, in the field of science and most importantly identify criminal especially at different airports. The fact detection is an important topic at present. However, face detection in the field of computer and graphics is quite challenging. The involvement of humans in the digital world is becoming common like understanding the visual facial expression of a human. At present, there are different studies that tell us about the facial expression algorithms and methodology regarding detecting a facial expression. These methodologies and algorithm are not enough to get the accurate results and has become one of the important topics. However, to get accurate or nearly accurate results, it is high dimensional and complex method. The human face has very complex muscular structure and bone structure. The human face contains different complex muscles and expression by a human is normally made by the stretching and contracting of a facial muscle or by relaxing or stressing of facial muscles[1].

For the current project, video of a face was taken that has hundreds of facial marks with the help of optical motion capture system. The video or a person for the facial expression is required to speak a designed corpus along with the facial expression. To have accurate results and outcome regarding the detection of face and emotions it is important to divide the face region into different sectors. In the current research, we have divided the facial expression of a person into 6 different sectors of facial regions. Six facial markers represent a specific region on the face. With the help of 3D rendering techniques, the final facial expression can be made. In the project, different methodologies and techniques are mentioned to detect the face or facial expression[2].

In a simple way, the face detector can be made with the help of simple algorithms and techniques for example; for the purpose of face detection from an image, seven different faces in an image were taken. To detect the faces on an image, spectrum of color techniques and neural networks and support vectors were used to support vector machines were used. Moreover, maximum rejection model were also used to detect some faces in an image. It is challenging to design an algorithm that helps to detect a face expression of facial structure from an image[3].
1.1 Block diagram for a common face detector

To detect the face from an image, it is important to follow the block diagram in a sequence so that the accurate outcome can be received with minimum or fewer errors. The resection-based method to detect the facial structure was chosen so that it will reject the non-facial region like non-skin colors with the help of color segmentations. According to the block diagram, the first image is selected that have some faces. The rejection process will reject the non-facial region in the image based on the skin color schemes. After the rejection of the non-facial schemes, it is important to follow the morphological processing. Morphological processing is kind of a filter that removes the clutter from an image.

The next step is to find the connected region analysis. This technique helps to find out the holes or number of holes. Based on the number of holes, template matching is used that helps to detect the number of faces or no face in the image.

1.2 Segmentation of skin color

The main purpose of the segmentation of skin color is to reject the non-facial region from the image. According to the face color across the world, it is identified that the human color varies based on luminance value and are closely related to the chrominance value. For the detection of the color of the face of the human, it is important to use the HSV that is hue, saturation value techniques. These techniques with the help of luminance information of the human face decouple the chrominance information of the human face. However, only focus on the hue and saturation component will help. H and S that is hue and Saturation color components. According to the graph, it can be identified that the H and S color components of the human face are nicely clustered. The segmentation process can be performed based on following steps:

1. The input image or selected image should be sub-sampled at 2:1 to have improved computational efficiency.
2. The resulted image should be converted to the HSV color schemes.
3. After the color schemes, the pixels, that fall outside the facial expression of an image should be rejected. In other words, the rejected pixels must be in black color.

1.3 Morphological process

The process of morphological is used to filter the image from the clutter. The morphological process helps to get or obtained the image almost equal to the input image without any noise and distortion. The algorithm of morphological process includes:

![Figure 3. Morphological process.](image-url)
In detail, the morphological process usually works on the image with an intensity that means the image is usually converted into gray and white color schemes. The division of regions into small regions helps the morphological process to clean or reject the nonfacial region from the image. As the threshold is usually low, it creates holes or clean spaces on the face region. In other words, holes are created. However, single connecting region helps to obtain the features on the face on the image[4].

1.4 Analysis of connecting region

The resulted images obtained by the morphological process contains some holes unclear regions like the matching of clothes with the skin color etc. so it is important to perform the connected region analysis. Geometry of connected region can be explained as follows:

The short region on the image would be the region that has a small height; the narrow region would be a region of the image that has a small width. The narrow and tall region on the image would be the region that has large height, but small width and the side and short region on the image would be the region that has small height but the large width. Algorithm or flow chart for the mentioned regions can be defined as:

![Flow chart for the connected region](image)

Another technique that is becoming popular to find the facial expressions along with the designed corpus required the use of Principal Component Analysis. This technique helps to create the 3D facial expressions of a person. In these techniques, the different emotions or emotions on the face of a person are usually described by the different color of schemes. The recorded facial expression of a person is then visualized with the 3D trajectory. The running visualization of the facial expressions is defined in Figure 5 below:

![Figure 5. 3D visualization of facial expression.](image)

In the above-mentioned figure, the region of the specific part of the face is mentioned with the color schemes like the jaw region, eye region, forehead region of the face of a person, etc.

In the current paper, consist of different techniques used to find the facial expression of a person. The paper is based on different sections the first section of the paper is consist of the introduction and the simple facial expression techniques that are normally used and the second section of the paper includes the visualization of the facial expression with the help of videos. The facial expressions are identified bases on the different color schemes assigned to the specific part of the face of a person. Moreover, the paper includes information about the temporal, spatial data regarding visualization. The paper also includes detail about the rejection of region based on data and the number of application that are used for the purpose of visualization etc.

2 LITERATURE REVIEW

In the past, there are many types of research, and the data is available regarding the facial expression and common techniques that are commonly used by the researchers. However, as it is mentioned above, it is a quite complex technique to design the algorithm. In the past different facial animation and modeling techniques were introduced and developed by different researchers at that time. However, in the recent years, the concept of data drove facial expression has been increased[5]. The data drove facial expression to use the pre-collected data of the facial motion. For the purpose of novel animation and the purpose of editing application like statistical models etc., the concept of facial coding systems is used in the past to find out the
human expressions. These expressions are usually identified with the help of the basic facial units of actions\[6\].

As it is discussed above that, the facial expression visualization requires hundreds of points or markers at the specific region of part of the face for example for the jaw side, forehead side of the face, eyes side of the face, etc. The 3D data or facial markers can also be regarded as the time-varying data or multivariate data. To explore the high definition data sets, different techniques can because for example; the scatter plot matrices, glyphs, pixel-oriented methods and parallel coordinated methods\[7\]. All these methods are used to find the high definition data sets. In the above-mentioned techniques, the dimensions on the screen are usually positioned either one or two arrangements. On the other hand, multi-resolution techniques are normally used for the purpose of grouping the obtained data in the form of hierarchies and to visualize them at the required level\[8\].

Another major reason of the clutter is the high dimensionality. At present there are different approaches to remove or reduce the dimensions, for example, the principal component analysis, self-organizing Maps and the multidimensional scaling, etc. these techniques are commonly used to reduce the dimensions and to visualize the data. Another technique that is visual hierarchy technique used for the dimension reduction helps to create the lower dimension spaces based on the original database\[9\]. For the reduction of clutter from the image the technique, generate the low dimensional spaces. These lower dimensional spaces also reduce the original data information so in that case, this technique is not so good as compared to the above three mentioned techniques\[10\].

For the volumetric data, the most common technique is the visualizing time-varying data. Recently different researchers have identified the different method to visualize the data instead of using or relying on the animation\[11\].

2.1 Comparison of the single and 3D rendering results

On comparing the results that are single line representation of the markers or the points and the 3D representation of 3D tube rendering it is identified that the 3D rendering results are more accurate and effectively represents the emotion of a person. The visualization of the emotions is clearer because of the different colors that represent the different emotions. Red represents the anger, green represents the happiness, and blue color represents the sadness. Moreover, on comparing the data, it is identified that the 3D tube renderings effectively represents the trajectory trends as compared to the single line.

2.2 Application of the system

The above system is applicable for the visual exploration of persons face expressions. With the help of the above-mentioned system, a person can easily employ the 3D system on the face of another person to get the 3D visualization. The system is easy to use as it helps to rotate the image in any direction and to get the 3D visualization\[12\]. The experiments show that how different region or sub-regions on a face can be used to detect the face expression of emotion of a person while talking. However, it is important to note that the facial expression of the same or almost same emotions may cause a problem or may become difficult to recognize by the system.

3 METHODOLOGY

Algorithm and common method was used to expressive facial animation synthesis by learning speech co-articulation and expression spaces. For that purpose, the electronics visuals were also generated to examine the focus and the direction of the facial expression. This method is quite useful to detect the accurate visual expression and the framing. Moreover the 3D visualization is also important and it is used for the facial expression resignation. Moreover, three different kinds if the facial expressions of a person were used that is subject to the division of the face of a person in a more appropriate manner. Total of the six sub-regions were included and the range of the markers was also put in to the selection that is between 90 and 102. Moreover, it is to be noted that the markers and the submission of the sub-region were also used as a face; the expressions were also used to identify which categorized in to the different regions. These regions were red for the anger as well as the green for the happiness. Moreover, the blue for the sadness was also used. Moreover, it is said that the data was obtained and the way results are obtained and they were also identified and concluded in that realm of the 3D representation that is subject to the of the facial expression. It is more accurate as well as than the single line representation in that regard.

4 TESTS AND RESULTS

4.1 Data acquisition of the facial motion

To the current project, the facial expression of the person is taken with the help of the VICON motion capture system. O the face of the selected person facial motion, 102 markers is marked on the face. After the allocation of the markets, the person is then asked to speak the designed corpus along with the emotions of several times. Reach time it is required by a person to speak the designed corpus with the different emo-
It is required by the person who reads the designed corpus with the three basic emotions each time that is, happiness, anger and with sadness. It is required by the person who reads the designed corpus with the three basic emotions each time that is, happiness, anger and with sadness. Figure 6. Facial motion with the markers that represents the specific region on the face. The data of person including sad, happy or anger was taken in 120 frames per second. Due to the large movement and the requirement of the removal of unnecessary markers on the face, the markers are used in the range of 90 to 102. On selecting the 90 markers range the facial motion was accurately captured. After getting the required data, it is important to get the normalized data of the facial motion. After getting the data, the markers that represent the specific region and expression of a person are then translated so that the specific ordinates on the face can be obtained. For the reference picture or image, the closed mouth face expression was chosen. The 90 markers are then packed into the 90 by 3 matrix that is b. Each motion of the person is then created and stored in the matrix. The calculation was made based on SVD Singular value Decomposition. The expression can be written as:

\[ b^T a_i = UDV^T \]

Now the rotation matrix can be defined as;

\[ Z = VU^T \]

The current project is based on the reduction of a region based on the dimension and 3D Visualization. The first step is to divide the region into six basic regions or sub-regions. With the use of principle, component analysis the singular vector was generated. The singular matrix was then generated which was then further reduced. With the effective use of the visualization techniques, the different motion expressed by a person is then characterized in different colors. The red color is assigned to represent anger, the green color assigned to express the feeling of happiness and the blue color is then assigned to express the feeling of sadness. The face, jaw side of the face, mouth region on the face, etc. some of the points or markers are not positioned at equal level off in symmetry. To have symmetry, it is important to expresses the markers of the specific region of the face in the form of vector. The expression for the above-mentioned technique can be expressed as:

\[ EigMX = (U_1U_2U_3) \]

\[ C = Eig \ M X^T. (A_i - \mu) \]

In the above expression the variable \( \mu \) represents the mean vector that is \( a(A_i) \). EigMX is the Eigen matrix that represents the reduce vector coefficient. In the project, the reduced vector coefficient is used in the sub-region of the face instead of the overall face of a person.

### 4.3 3D Visual

As the facial motions are now obtained based on six different facial sub-regions, the plots of the markers are then plotted in the 3D this is called 3D visualization. As mentioned above the red region represents the anger, the green region represents the happiness, and the blue region represents the sadness. The dots are plotted along with the assigned colors to get the visual of the image.

Figure 7. Representation of the markers or dots with the single line.

To get the more accurate visualization of the markers of dots that represents the face expression, the Phong shading were used. The cylinder shading used in the project has the constant radius. However, the cylinder is placed in the single line to get the more 3D view of the image. The image obtained by placing the cylinder on the each line was in the 3D space. The concept is known as 3D tube rendering.
4.4 Comparison of the single and 3D rendering results

On comparing the results that are single line representation of the markers or the points and the 3D representation of 3D tube rendering it is identified that the 3D rendering results are more accurate and effectively represents the emotion of a person. The visualization of the emotions is clearer because of the different colors that represent the different emotions. Red represents the anger, green represents the happiness, and blue color represents the sadness. Moreover, on comparing the data, it is identified that the 3D tube renderings effectively represents the trajectory trends as compared to the single line.

4.5 Application of the system

The above system is applicable for the visual exploration of persons face expressions. With the help of the above-mentioned system, a person can easily employ the 3D system on the face of another person to get the 3D visualization. The system is easy to use as it helps to rotate the image in any direction and to get the 3D visualization. The experiments show that how different region or sub-regions on a face can be used to detect the face expression of emotion of a person while talking. However, it is important to note that the facial expression of the same or almost same emotions may cause a problem or may become difficult to recognize by the system.

5 DISCUSSION AND CONCLUSION

In the visualization system, there is different emotion category, as there is the reorganization of the visual inspection, through the experiment and through focusing on the motion frames, the analysis was done that the facial motion sequences have several of different types of pattern. There are focuses on the hundreds of thousands of the facial motion sequences. Consequently, the thinking is done on the machine learning methods and the recognition of the alternative emotions concerning the facial motion are defined. Through doing the intuitive visual inspection or doing, focusing on the facial expressions data, there is the learning on the facial motion sequences that different emotion can be there in the personality of a person[18].

For the current project, the techniques and different algorithms of the researchers are mentioned in the paper. The common method to identify the facial expression is also discussed in the paper. It was purposed to find the three different facial expressions of a person by dividing the face of a person into six sub-regions. The range of the markers was selected to be between 90 and 102. According to the markers or sub-region on the face, the expressions were identified which were red for the anger, green for the happiness and blue for the sadness. According to the data, it is identified that the 3D representation of the facial expression is more accurate than the single line representation.

REFERENCE


