Generalized Perceptual Modeling: Virtual Human Face Modeling and Expression Recognition Considering Emotion Understanding

Ziyang Weng*, Ziyu Zhang, Yinger Liang and Biyi Dai
School of Information Management, Wuhan University, Wuhan, China
weng_ziyang@whu.edu.cn, zhang_ziyu@whu.edu.cn, Vincy020125@163.com, taipeki@163.com
*corresponding author

***Abstract*—Human facial emotion recognition (FER) plays an important role in human-computer interaction applications. Given the widespread use of convolutional neural networks (CNNs) in automatic video and image classification systems, higher-level features can be automatically learned from hierarchical neural networks with large data. However, learning CNNs requires a large amount of training data for adequate generalization, while scale-invariant feature transform (SIFT) does not require large training samples to generate useful features. In this paper, we propose a representation of a generalized perceptual model for recognizing facial expressions from a single image frame, a model with a set of common interpretable features. The method combines SIFT and deep learning features extracted from CNNs models at different levels, and then employs these combined features and classifies the expressions using a support vector machine (SVM). The performance of the method has been validated in experiments testing expression recognition for virtual human face modeling, demonstrating the great potential of combining shallow features with deep learning features. The experiments show that features predicting aesthetic preferences can emerge in layers of deep convolutional neural networks trained only for emotion understanding recognition. It is found that human preferences for facial observation can be effectively translated into computationally inferable, systematic integration of emotionally latent behavioral organization features.**

Keywords- facial emotion recognition, convolutional neural networks, generalized perceptual models, deep learning features, support vector machines, virtual human

I. INTRODUCTION

Human facial emotion recognition (FER) is a challenging task in the field of intelligent development of virtual human micro-expressions. The subjective representations and interactive feelings of facial emotions can be interpreted as the basis for building trust in the process of human civilization. Finding the same or similar perceptual understanding laws in them and recapitulating them to technical prototypes that can be reused in the field of intelligent research requires identifying the concurrent logic of subjective emotions, underlying psychology, neural control, Superficial muscle tissue, eye movement trajectory constraints, and other compound response behaviors in the process of human emotion expression. Developing and designing algorithmic models that can interpret emotions from the face is an important part of the human-computer interaction environment to predict emotions, deduce behaviors, and achieve sympathy[1].

One human ability is to acquire information where it is shared or easily accessible through the detection of gaze. It is an open question whether one can predict human visual preferences in response to aesthetic desires from the basic constituents of visual images. Here, we develop and test a computational framework to investigate how aesthetic values are formed. We show that considering a mixture of image accuracy levels and high-level knowledge links can explain human preferences for visual preferences. Subjective value ratings can be predicted both within individuals and between individuals, using regression models with a common set of interpretable features. Previous research has shown that features predicting aesthetic preferences not only emerge hierarchically in deep convolutional neural networks trained for object recognition, but that human preferences for emotion understanding during pairwise viewing can be explained, at least in part, as a systematic integration of underlying visual features of the face.

To realize the analysis for human expression recognition from facial images, mathematical, computer, and ergonomic researchers look forward to different approaches to try to replicate high-precision facial models and explore the computability of expression patterns, with approaches gradually radiating from traditional research in the fields of image processing, computer vision, pattern recognition, and artificial intelligence to the fields of psychology, semiotics, performance, biology, and anthropology. In particular, with the advent of neural network (CNNs) convolutional computation, it is as if training and interpretation can be achieved in a very efficient manner as sensor input signals (e.g., skin temperature, iris changes, and brain wave curves) are augmented with breakthroughs in high-precision model acquisition techniques. Interdisciplinary differences and ineffective correlations between data have severely hampered the reuse of such research in the study of virtual human emotional output, and the frequent occurrence of the " Uncanny Valley " effect in virtual human simulations is the result of developers' insistence on ignoring the subjectivity, sensitivity, and growth characteristics of human perceptual experience.

Is it possible to effectively predict the preference of individual emotion understanding from the basic components of generalized perception such as visual physiological and psychological laws? In particular, global facial perception is associated with the traversal feedback of the total emotional experience of the subject, and micro-emotional perception is

2693-9371/22/$31.00 ©2022 IEEE
DOI 10.1109/QRS-C57518.2022.00067

associated with the biased evaluation of the subjective emotion of the subject at the moment are both urgent antecedent issues to be addressed. Here, this study develops and tests a computational framework to investigate how values of emotion understanding are formed. Experiments show that base-level portrait-based modeling and high-level feature blending can explain human preferences for virtual human micro-expression recognition. Using a regression model with a common set of interpretable features also shows that predicted emotion understanding features can emerge in layers of deep convolutional neural networks trained for object recognition only. Moreover, human preference for avatar expressions can be at least partially explained by the integration of emotion understanding that is determined by line of sight as context and driven primarily by lip angle changes.

II. MOTIVATION

In the last decade, human facial emotion recognition (FER) has aroused and become a hot research field. Currently, there are various applications of automatic FER, such as data-driven animation, interactive games, social robots, and many other human-computer interaction systems [2]. Psychologists have developed different systems to describe and quantify facial behavior. The Facial Action Coding System (FACS) developed by Ekman and Friesen is the most popular system among them. FACS was created to classify human facial actions based on facial expressions and more systematically classify criteria [3] for emotions, body, and expressions. Recent studies tend to pose bias to study image changes due to light, age, and occlusion. Current automatic FER algorithms usually include three main steps: face acquisition, feature extraction and classification [4]. The most important step in the FER systems proposed by most researchers is feature extraction, which aims to represent facial images as feature vectors. Extracting features from the input data will significantly affect the accuracy of the final classification. For face representation, most existing works use various manual features, including Gabor wavelet coefficients, local binary pattern histograms (LBP), directional gradient histograms, and scale-invariant feature transform (SIFT) descriptors, or a combination of these features to obtain the representation [5]. Then, various machine learning algorithms are adapted to perform the classification task.

Facial expression recognition (FER) plays an important role in human-computer interaction applications. Given the wide use of convolutional neural networks (CNNs) in automatic video and image classification systems, more advanced features can be automatically learned from hierarchical neural networks with large data. However, learning CNNs requires a large amount of training data for adequate generalization, whereas generalized perceptual transform (SIFT) does not require large training samples to generate inefficient data and can be optimally evaluated to constrain quality feature generation.

The facial expression hypothesis is motivated by the possibility that facial regions may have developed from the original retinal image and by the selectivity of facial features acquired through natural visual experience. The face has a typical feature configuration, usually perceived in a typical vertical orientation and often fixed at a specific location. From an objective perspective, highly accurate and precise facial models do not achieve an effective interactive feedback chain supporting emotional expression, recognition, and response. After all, humans have not evolved over thousands of years to acquire information from facial perception based on high-quality acquisition of skin texture, skin tone and similar details. Rather, they are sensitive to the implicit linkage of adjacent regions observed from almost invisible areas of change. For example, the occipital facial area (OFA) and the fusiform facial area (FFA) are considered to be brain regions dedicated to facial perception [6]. However, the field of medical neurology is still unable to effectively identify their intrinsic functional and organizational status. Here, the intrinsic functional association of emotion perception with "facial expressions" is a particular hypothesis. The visual areas of the face would contain the texture of facial features on the cortical surface and would have physical contiguity with adjacent areas, and common behavior of emotion synergistic transmission.

To test the facial expression hypothesis, we presented subjects with images of isolated facial features during an experiment on the perception of ancient Chinese figure painting faces. (This class of evaluative responses is best explained by the low-level image properties of the stimuli.) When each facial feature was attended to, the aggregated distance between feature points reflected the physical distance between facial features. The facial examination would be an example of reflecting the topology in the expression of emotion, not as part of the organism itself (retina is the retina, the body is the soma), but as an external object with particular perceptual significance.

III. BACKGROUND

Simulated virtual human face construction estimates a given 3D image model based on the skull mapping and the positional interrelationship between the facial features in the captured image, and this technique has been widely used in film and television, game development and ancient human face restoration [7]. Among them, facial expression capture and expression translation is the most difficult research area to break through. In particular, the measurement and statistics of facial soft tissue thickness is the basis of 3D face restoration, and the skull feature point localization and soft tissue coupling connection will directly affect the effect of facial micro-expression realization. The definition of skull feature points is not yet a unified standard, and the definitions of feature points given by scholars in different fields meet certain rules: feature points include two parts, one part is located on the median sagittal plane of the head, and the position of such feature points is basically fixed; the other part is symmetrically distributed on both sides of the median sagittal plane, and there is no unified standard for such feature points, and most of them are defined by researchers themselves.

The core expression area concept of bilateral brow arch-upper and lower eyelids-core facial muscle groups-nasal flanks-lip corners region proposed by the authors in the previous research work was deepened in the present study [8]. The main superficial muscle layer and sympathetic nerves are concentrated in this area, and the main facial information transmission of eye and lip attraction, guidance and speech output, respectively, occur in this area. The same pattern can be found in the validity evaluation of different facial images through eye-movement experiments.

In recent years, with the development of medical image acquisition equipment, ultrasound and radiographs have been applied to the measurement of facial soft tissues in live samples, expanding the possibility of enhancing the performance of superficial muscle groups [9]. Due to the constraints of the working principle of the acquisition equipment, the above methods can only measure the soft tissue thickness values of the feature points in a single pass and cannot obtain a 3D model of the skull and face, then it is necessary to rely on domain experts to achieve alignment based on the bone point features with the facial image position labels and adjust the matching model in combination with the skull model. The method can not only calculate the soft tissue thickness of skull feature points and calculate the soft tissue thickness distribution of the face, but can also separate the hidden deformation layer from the 3D model of the reconstructed cranium and face, which is more suitable for representing micro-expression changes than other methods.

In this paper, we achieve soft tissue thickness as well as soft tissue thickness at facial dense points through matching skull feature points, facial position mapping, and labeling at head and facial feature points to generate a facial soft tissue distribution map, and achieve skull appearance restoration according to the facial soft tissue thickness of sample age, gender and physique. This stage mainly includes three issues: the definition of skull feature points and the measurement method of facial soft tissue; the analysis of soft tissue thickness distribution at the feature points and giving the face physique classification method; the realization of 3D face recovery of a given cranium according to the mean value of soft tissue thickness corresponding to age, gender and posture of the facial model volunteers.

Most of the skull feature points defined in this paper are located at the turnings and protrusions of the skull surface, such as the zygomatic arch points and the outer edge points of the eye orbit. Based on the developed expert knowledge in anthropology and forensic medicine, and combined with the experience of traditional manual facial restoration, 78 skull feature points were defined, among which 12 feature points were located in the median sagittal plane and the remaining 66 feature points were symmetrically distributed on both sides (Figure 1). The skull feature points are further divided into two categories: five senses feature points and facial feature points, five senses feature points are located in the five senses and the neighboring positions, and facial feature points are located in other positions on the face except for the five senses feature points, five senses feature points affect the basic form of the facial features, and facial feature points affect the basic form of the face, and the names and categories of the feature points are shown in the figure.

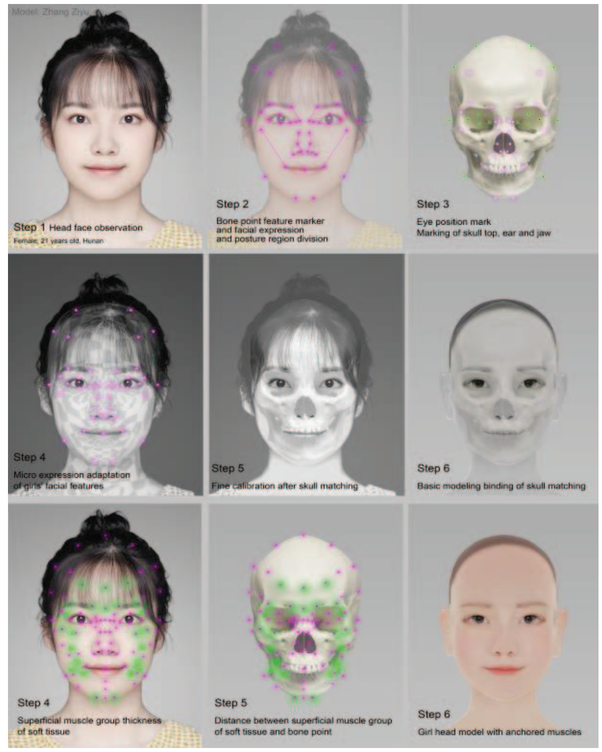

Figure 1. A modeling method based on skull bone point location features and soft tissue structure label

IV. METHODS

In this paper, we propose a new generalized perceptual model with hybrid feature representation for achieving optimization in perceiving micro-expression data embedded in a single avatar model, which combines SIFT and deep learning features extracted from CNNs models at different levels, and then employs these combined features and uses support vector machines (SVMs) to classify expressions. As shown in Figure 2, to enhance the efficacy of perceptual evaluation, the location features of superficial muscle layers and neurosensory networks are extracted as constraints on the expression output association regions.

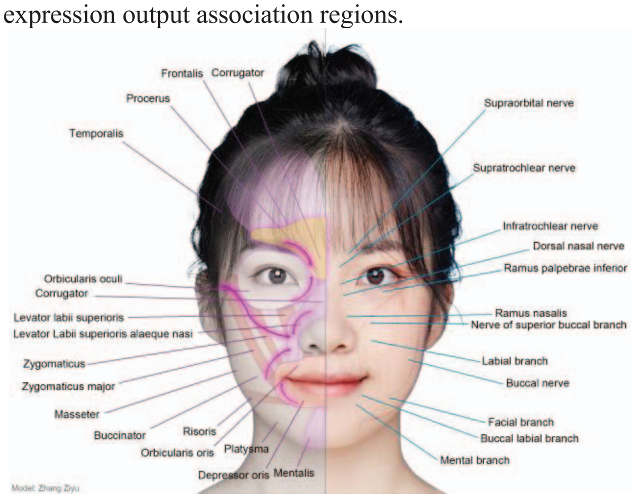

Figure 2. A Method of Distinguishing Features Based on Superficial Muscle Layer and Neural Sensor Network

A. Methods

To evaluate the generalization ability of our method, we built a high-precision acquired virtual human face model and conducted experiments by location marking of the occipital face area (OFA) and the fusiform face area (FFA), and constructed a computational architecture for emotional perception evaluation.

Operation procedures of generalized perceptual modeling:

1) Firstly, calibrate the skull feature points, then determine the face feature points according to the direction of soft tissue motion constraint, and predict the soft tissue thickness (including muscle layer, fat layer, and sympathetic nerve distribution).
2) Predict the direction of soft tissue motion constraint during the prognosis and directly determines the soft tissue thickness value, along the normal direction of the face or skull feature point; top-down labeling along the face or skull feature point; determine the corresponding skull and face expression muscle direction features separately according to the anatomical relationship, and their connecting lines are used as the labeling direction. Due to the complex geometry of the individual's head and the data noise of inconsistent left and right faces in the modeling process, resulting in the normal line direction at the same position of different skulls is not necessarily the same, so the soft tissue motion constraint direction defined in this paper is labeled along a straight line. The soft tissue thickness and its evaluation are shown in (Figure 2), the light-colored points indicate the defined skull feature points, the dark-colored points indicate the calculated face counterparts and the line segments indicate the soft tissue thickness, and the sampling effect is ideal because the volunteer's face is tight and smooth.
3) Based on the common specifications of virtual human modeling, realize the model pulling and texture matching of 3D head images, complete the wiring of facial structure and the modeling embedding of fine eyes in the eye area; focus on completing the cutting and morphological repair of the upper and lower lips of the mouth, and build the basic model for the subsequent eyesight change and the realization of speech actions of the mouth.

B. Experimental design for emotion understanding

Facial features were extracted using two sets of high-resolution images (ancient painting slices and photo slices). The faces in the images were first aligned using Matlab by computing the midpoints of the eyes and mouth in each image from a set of eye movement trajectories, then finding rigid spatial transformations between these points, and applying the transformations to the images. From the aligned facial images, the following 12 facial features were sampled using non-overlapping windows of equal size: left and right eyes, the distance between eyes, nose, mouth, left and right hairlines, left and right ears, left and right jawlines and chin. The vertical position of the ear sampling window needed to be manually adjusted to match individual changes in ear position, but all other features were sampled using the same window for each face. Subjects Fifteen healthy volunteers (13 female, 2 female, age 21) with normal or corrected-to-normal vision participated in this material and study.

Three different spatial layouts were used to represent emotion understanding: trajectory data, thermal data, and implicit target. Throughout the experiments, 15 different samples with the same facial features were evaluated during a 15-second attention period. Each experiment consisted of two blocks for each facial feature, and every fifth block was a baseline block presented separately with a fixed crossover. Since the comprehension metrics were created using Matlab, their timing was controlled using the presentation (neurobehavioral system).

1) *Eye-tracking comparison:*

Based on the extraction of valid data from the observation trajectories of volunteers, point clusters with distribution characteristics were formed, and the trajectory map data of A-1 and A-2 orthophoric sight change control groups and B-1 and B-2 side-looking sight changes were used for the control experiments, which were calculated using the Perceptual Hash Algorithm.

Perceptual Hash Algorithm (PHA) [10] is a hash algorithm to calculate the similarity of images. The algorithm calculates the similarity by reducing the size, simplifying the color, calculating and reducing the DCT (Discrete Cosine Transform), calculating the average value, calculating the hash value, comparing the Hamming distance and other steps, independent of the height and width, brightness and color of the picture. The algorithm can better remove noise (such as inscriptions and small tears in ancient paintings) and judge the overall shape of the graphics from the perspective of subjective visual perception with high accuracy. When using the perceptual hash algorithm to calculate the picture similarity, it is generally considered that if the Hamming distance $d(x, y) < 10$, the two are similar pictures.

Table 1. Results of picture similarity calculation

Sample1	Sample2	Similarity
A-1(frontal direct)	A-2(frontal direct)	0.9140625
B-1(profile advert)	B-2(profile advert)	0.8984375

The results of the calculations are shown in Table 1.It can be seen that the distribution area of the observer's eyes is concentrated between the eyebrows, the bridge of the nose, the cheekbones, the inner side of the facial muscle, the corner of the lips; the top of the forehead, the cheeks and the lower jaw are not in the gaze area; among them, the eyebrows and eyes, nose and lips constitute the core area.

2) *Dense lines comparison:*

Based on the distribution point clustering, linking trajectory routes with the highest density area, we run a control experiment with A-1, A-2 frontal direct change as the control group and B-1, B-2 side view change in curve features, and use cosine similarity and SIFT similarity to calculate.

Cosine similarity [11] measures the similarity between two vectors by measuring the cosine of the angle between them and is applicable to the comparison of vectors in any dimension. The similarity between images can be measured by transforming the set of image features in a sample into vectors in a high-dimensional space and calculating the cosine of the angle of the inner product space of these vectors representing each image feature. The cosine value between two vectors can be found by the Euclidean dot product formula.

$$\mathbf{a} \cdot \mathbf{b} = \| \mathbf{a} \| \| \mathbf{b} \| \cos\theta \quad (1)$$

Given two attribute vectors A and B, the remaining chord similarity θ is given by the dot product and the vector length, as shown in Equation (2). where A_i, B_i represent each component of vectors A and B, respectively.

$$\cos(\theta) = \frac{A \cdot B}{\| A \| \| B \|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad (2)$$

SIFT (Scale Invariant Feature Transform) [12] algorithm is an efficient region detection algorithm that can be used in object recognition, robot perception and navigation, image stitching, 3D modeling, action comparison, etc. Its main idea is to find the extreme points in the scale space, then filter the extreme points to find the stable feature points, and finally extract the local features of the image around each stable feature point to form local descriptors. Results of calculation are in Table 2.

It can be seen that the observer's sight trajectory occurred in a large difference, A-1, A-2 frontal direct change control group appeared the features of deviating from the inter-brow and pupil, with the deviation of the sight, the behavior of the sight clearly appeared to move in anticipation, fix focus with the bridge of the nose, and jump to the corner of the mouth near; in B-1, B-2 profile advert sight change control group, it appeared to change from exploring features from the edge of the whole face to expectation of anticipating the accompanied expression that the mouth will appear, the focus of the sight moved to the corner of the mouth and have a stay, and this observation of the sight tends to take a cross correspondence between the left and right up and down representations.

Table 2. Results of cosine similarity and SIFT similarity calculation

Sample types	Sample1	Sample2	Cosine similarity	SIFT matching
Trajectory points	A-1	A-2	0.777	0.181
Trajectory points	B-1	B-2	0.781	0.3773
Feature Lines	A-1	A-2	0.744	0.7778
Feature Lines	B-1	B-2	0.851	0.65

3) *Pattern discovery：*

Based on the discovery of implicit targets, this experiment obtained multiple expression slices of the model's eyebrow-eye region and nose-lip region, and found the asymmetry and lateral deviation trends in the model's expression changes by using the facial features deformation promoted by typical expressions, marking the displacement characteristics with auxiliary lines, and marking the superficial muscle group deformation characteristics with thermal marks, among which the changes in the eyebrow arch and the decree area are the strongest; the lip could not be generalized due to its special structure, but it also showed the same directional representation as the glabella height change.

V. EVALUATION

In this experiment, we tested the facial expression hypothesis by presenting images of isolated facial features to subjects in an eye-movement experiment. Although this experimental simulation did not include perspective changes, most of the facial visual observation data in the experimental slices were modeled. Response results obtained using general linear model (GLM) analysis to estimate facial emotion understanding exceeded study expectations and could also explain the hypothesis that facial emotion regions are intrinsically organized for emotion expression supply.

Starting from an oversimplification, assume that the face always appears in front of the observer at the same distance and at the same retinal position (e.g., fixed in the center).

1) If the retinal position cannot be directly opposite to the observation slice, the whole will be quickly perceived and the selectivity corresponding to the facial parts (including eyes, nose, and mouth) can be obtained by adjusting the vision to cover the whole head.
2) The spatial organization of the parts will be similar to the spatial layout of the face, with the nose represented in some morphological area between the areas representing the eyes and the mouth. In natural experience, people can view faces from different distances, and faces are not always fixed in the center. To test whether the facial expression assumption is reasonable when we consider more natural viewing conditions, we used a simple simulation (Figure 3). For example, the spatial distribution of facial emotional expression features over the retina and the similar distribution of facial features to each other. Despite the differences in viewing conditions, the peaks of the retinal feature exposure maps still formed a face map.

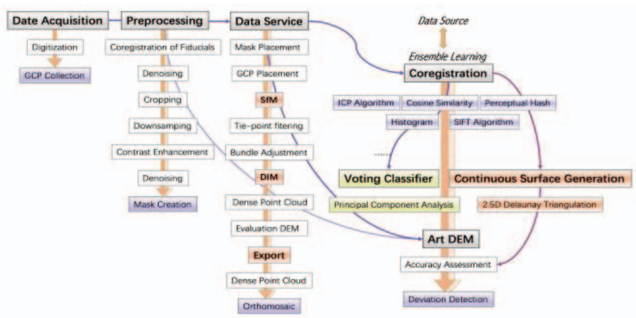

Figure 3. Modeling based emotional perception evaluation computing architecture

3) For each facial feature, we estimated the spatial distribution of retinal exposure when viewing conditions were randomly drawn from the actual distribution of sight distance and gaze points. This suggests that accepting the original retinal image of a size roughly corresponding to the face, if it is selective for its most easily understood features, may develop into an understanding of the overall emotion of the face even under very variable viewing conditions, and the brain completes the semantic complement of sparse data to automatically feed back a deep understanding of facial expressions.
4) FFA and OFA respond to faces even when they are presented at the edges (Hasson et al., 2003) [13]. Even though selectivity for certain types of natural shapes evolved from retinal proto-images, the resulting shape detectors may be quite tolerant to retinal locations. We found that during the observation of facial emotions, observers respond to each feature detection for its preferred feature, e.g., eye, lip; in contrast organs with stable features for precise retinal locations are tolerated to a certain extent and are essentially ignored in the visual trajectory.
5) Based on the above findings, the optimization steps of the virtual head portrait model established in the course of this experiment are reflected in: 1strengthening the eye modeling accuracy (eye movement pattern and eye material enhancement); strengthening the normality and alienation of the nose in breathing, where optimization will bring subconscious psychological identity to the observer; strengthening the backstage of the superficial facial muscle group deformation association to constrain the overall facial expression with micro-asymmetry; independently addressing the nose-lip region, for its strongly deformed performance, in addition to the conventional position changes, it is necessary to emphasize the lip deformation accompanying the mouth movements, especially the collection of atypical large movements such as swallowing, spitting, swallowing, gnashing, chewing, sucking and sucking during the mimicry process, which can be refined and summarized as the constraint data for serving micro-expression presentation.

VI. Discussion

To characterize each facial feature, representational similarity analysis in generalized perceptual models calculated differences between response patterns evoked by emotion understanding and compared them with model predictions of representational distance (Kriegeskorte, Mur, & Bandettini, 2008; Nili et al., 2014) [14].

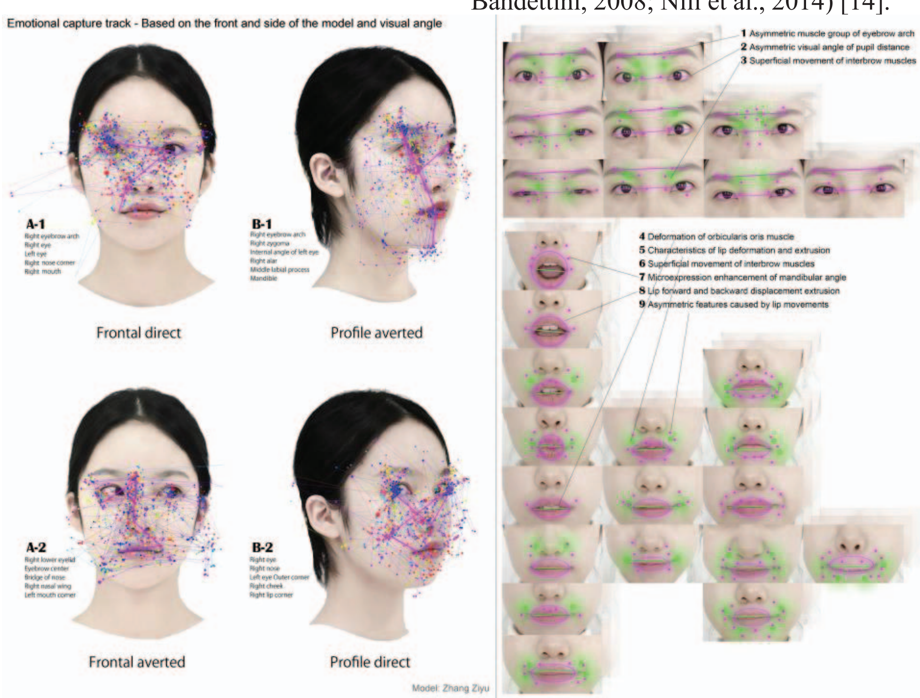

Figure 4. Micro-expression Tendency and Emotion Analysis of Object's Face Staring Direction Based on Eye Movement Trajectory

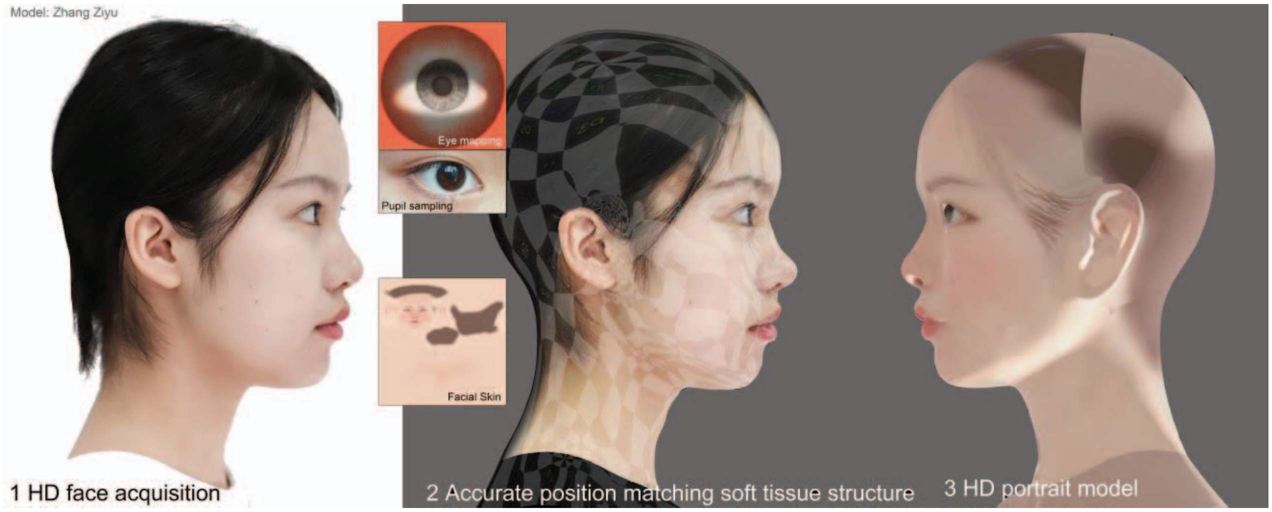

Figure 5. Video generation of high definition head models considering soft tissue structures

Some researchers have compared the difference in facial feature representation with the representation between facial features using three predictions: 1) the Gabor Wavelet Pyramid (GWP) model, 2) the physical distance reference between facial features, and 3) the physical distance between facial features when symmetric facial features are represented at the same location (symmetric reference). The GWP model captures the low-level image similarity between facial feature stimuli and was adopted by Kay, Naselaris, Prenger, and Gallant (2008) [15]. The values of the physical distance reference matrix are the distances between facial feature sampling windows. The distance matrix captures the spatial relationships between elements in the face. Symmetric reference is measured using the logic that the distance between symmetric features (eyes, ears, hairline, jawline) is zero and the distance between two symmetric features and the other features is the same.

And at the end of this perceptual evaluation of micro-expressions in the subject's face modeling, we tested whether face feature representations reflect emotions, i.e., understand underlying emotional states, i.e., whether subtle face feature representations can be explained by the physical distance between facial features. Within the model, we use the wiring data to accurately estimate the individual location of each facial feature point. We then considered local spherical neighborhoods around each location in an OFA-like Region of Interest (ROI) and assigned preferred features for locations with common deformations in the neighborhood. This process was repeated until all features had a cluster of positions or all positions above the threshold had been assigned to features. The size of the neighborhood (radius of the sphere) and the T-value threshold (the assumption of a replicable distance matrix between the locations of the no facial features) were optimized by evaluating the replicability of the exact distance matrix during the two measurements. The clusters of feature preferences are searched in 3D space (voxel coordinates) and assigned to the locations to control the deformation during the constrained expression animation.

Experimental results show that generalized perceptual models can achieve better classification rates when compared with current state-of-the-art CNNs approaches considering facial emotion understanding, which demonstrates the great potential of combining shallow features with deep learning features.

VII. Related work

The evaluation and application of generalized perception models have evolved from considering the emotion understanding to the regularity of soft tissue thickness distribution of the face in virtual human face modeling (the soft tissue thickness of the forehead is the thinnest and the difference in soft tissue thickness between different samples is not significant, the soft tissue of the cheeks is the thickest and the difference in soft tissue thickness between different samples is significant), and then to the facial restoration method based on generalized perception evaluation which essentially uses computer technology to simulate the traditional manual restoration method, the restoration effect is still constrained by the development of arithmetic power of 3D modeling technology. The more feature points are defined and widely distributed in the generalized calculation, the closer the simulation validity is to the actual face realism, and the more accurate the face restoration results are. The law of data changing in the above modeling process is applied to the reference face model to realize the non-rigid deformation of the reference face model, and the deformed face model is the restored face. The soft tissue thickness at the skull feature points of the restoration result is consistent with the average soft tissue thickness of the selected craniofacial classification library, and the facial soft tissue distribution of the restoration result is similar to the facial soft tissue distribution of the reference face, which also has application scenes in the field of police and security.

VIII. CONCLUSION

The experiments focus on processing single portrait facial emotion data to achieve predictions by analyzing multiple consecutive complex structural features. In deep learning-based facial micro-expression feature estimation, the deep network is transformed into a series of heat maps representing probabilities, where the most probable points are key coordinate points. Each key coordinate point represents the node part of the associated structural point trajectory. The use of adversarial networks can enhance the ability of deep neural networks to grasp micro-expression features in images.

In this paper, we combine a generative countermeasure network, a sparse detection network, and a Digital Elevation Model (DEM) spatial dataset, and transform the labeled information into an activation function to perform nonlinear processing of the matrix data. Meanwhile, a cosine dynamic learning rate decay strategy is used to control the gradient update to reduce overfitting. In this paper, a self-focused generative adversarial network is used instead of the common generative adversarial network to extract key location nodes.

ACKNOWLEDGMENT

This paper was completed under the support of Major Project of the National Social Science Foundation of China (NO.21&ZD326) "Virtual Reality Media Narrative".

REFERENCES

[1] Iigaya, K., Yi, S., Wahle, I.A. "Aesthetic preference for art can be predicted from a mixture of low-and high-level visual features." Nature human behaviour 5.6 (2021): 743-755.

[2] Felipe Zago Canal, Tobias Rossi Müller, Jhennifer Cristine Matias, Gustavo Gino Scotton, Antonio Reis de Sa Junior, Eliane Pozzebon, Antonio Carlos Sobieranski. "A survey on facial emotion recognition techniques: A state-of-the-art literature review." Information Sciences 582 (2022): 593-617.

[3] Ekman, Paul, and Wallace V. Friesen. "Facial action coding system." Environmental Psychology & Nonverbal Behavior (1978).

[4] Yahui Nan, Jianguo Ju, Qingyi Hua, Haoming Zhang, Bo Wang. "A-MobileNet: An approach of facial expression recognition." Alexandria Engineering Journal 61.6 (2022): 4435-4444.

[5] Jun Young Park, Tae An Kang, Yong Ho Moon, Il Kyu Eom. "Copy-move forgery detection using scale invariant feature and reduced local binary pattern histogram." Symmetry 12.4 (2020): 492.

[6] Charlotta Marina Eick, Gyula Kovács, Sophie-Marie Rostalski, Lisa Röhrig, Géza Gergely Ambrus. "The occipital face area is causally involved in identity-related visual-semantic associations." Brain Structure and Function 225.5 (2020): 1483-1493.

[7] Preim, Bernhard, and Patrick Saalfeld. "A survey of virtual human anatomy education systems." Computers & Graphics 71 (2018): 132-153.

[8] Ziyang Weng, Shuhao Wang , and Weixin Yan. "Understanding Rembrandt: Directed Knowledge Improves Robustness and Evolution of Facial Phenotype Modeling" . In 2022 9th International Conference on Dependable Systems and Their Applications (DSA), in press.

[9] Meikle, Brandon, and Carl N. Stephan. "B - mode Ultrasound Measurement of Facial Soft Tissue Thickness for Craniofacial Identification: A Standardized Approach." Journal of Forensic Sciences 65.3 (2020): 939-947.

[10] Zhu, Wenlie, and Jing Chang. "Research and implementation of perceptual hash algorithm for image retrieval." AIP Conference Proceedings. Vol. 2073. No. 1. AIP Publishing LLC, 2019.

[11] Soumyajit Saha, Manosij Ghosh, Soulib Ghosh, Shibaprasad Sen, Pawan Kumar Singh, Zong Woo Geem , Ram Sarkar. "Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm." Applied Sciences 10.8 (2020): 2816.

[12] Gupta, Surbhi, Kutub Thakur, and Munish Kumar. "2D-human face recognition using SIFT and SURF descriptors of face's feature regions." The Visual Computer 37.3 (2021): 447-456.

[13] Avidan, Galia, and Marlene Behrmann. "Spatial integration in normal face processing and its breakdown in congenital prosopagnosia." Ann. Rev. Vis. Sci 7 (2021).

[14] Hu, Xin, Fei Wang, and Dan Zhang. "Similar brains blend emotion in similar ways: Neural representations of individual difference in emotion profiles." Neuroimage 247 (2022): 118819.

[15] Henriksson, Linda, Marieke Mur, and Nikolaus Kriegeskorte. "Faciotopy—A face-feature map with face-like topology in the human occipital face area." Cortex 72 (2015): 156-167.