Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 7354–7358

Fourier Series Model for Facial Feature Point Land-Marking

H. Arabian*. N. Ding*. J. Geoff Chase. K. Moeller***

**Institute of Technical Medicine (ITeM), Furtwangen University, VS-Schwenningen 78054, Germany (Tel: 7720-307-4390; e-mail: H.Arabian@hs-furtwangen.de).*
***Center for Bioengineering, Department of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand}*

Abstract: The field of digital health apps, combined with intelligent learning systems, is new and expanding to incorporate a wide range of possibilities in different domains. An application in the field of digital therapy is for the incorporation of emotion recognition systems as a tool for therapeutic interventions. Adopting an individually tailored virtual world combined with a novel reward system in a gaming scenario, complemented with the technical affinity of most autism spectrum disorder (ASD) children makes a suitable atmosphere for therapeutic intervention. In this paper the use of image processing techniques coupled with Fourier models is used to generate point land-mark annotations on facial features in an image. The OULU-CASIA database was used for the analysis process. The images were first pre-processed based on previous work to reduce background noise and focus on the face. Afterwards a de-correlation stretch was executed to separate different features. A series of morphological, region detections and boundary traces followed. Fourier series models were used to transition the rough segmented pixel data into a smooth geometric representation. Twenty evenly distributed land-mark points are then selected from a fine mesh. Results showed that the geometric representation adhered to the segmented pixel data with a mean of 81.88% Dice similarity. The positive outlook highlighted the effectiveness of such a technique in automating the land-mark annotation process, which is tedious and time consuming. This method leads to explainable machine learning feature representations, which lead to more robust emotion recognition models.

Keywords: Autism Spectrum Disorder, Digital Health, Emotion Recognition, Fourier Series, Geometric Feature Representation, Image Land-Marking, Therapeutic Application.

1. INTRODUCTION

Digital health has become one of the key topics in the field of research. The pandemic has restructured some healthcare services to be able to provide more remote access to patients for a better quality of care with lower exposure risk. This is accomplished by eliminating burden of travel and at the same time reduces the strain on an overwhelmed healthcare system (Dorsey et al., 2013; Saanvi, 2020; Campbell, 2020; H. Arabian et al., 2022a). The field of digital health apps, combined with intelligent learning systems, is new and expanding to incorporate a wide range of possibilities in different domains (H. Arabian et al., 2022a).

Significant interest and growth has been seen in the field of emotion recognition over the past few years. The use of emotion recognition for therapeutic interventions i.e. emotion training of children with autism spectrum disorder (ASD) is currently being studied. ASD is a developmental brain disorder that impairs social interaction, communication, and behaviours (Disabilities, 2001). Estimates reveal 1 out of 59 individuals are affected by ASD (Rylaarsdam and Guemez-Gamboa, 2019). These therapeutic intervention systems can help individuals by coping better to different social interactions (Golan et al., 2010; Yuan and Ip, 2018; Herag Arabian et al., 2022b).

Children with ASD are familiar to a given routine and a divergence from this normalcy can cause psychological and emotional challenges to the child, and increased stress levels for the caregiver (Lugo-Marín et al., 2021). Typical therapy is defined by comprehensive individual and small group interaction sessions with trained facilitators and instructors. Adopting an individually tailored virtual world combined with a novel reward system in a gaming scenario, and technical affinity of most ASD children makes a suitable atmosphere for therapeutic intervention. This reduces the economic impact as well as streamlines clinician's workload.

Some studies (Leo et al., 2015; Ravindran et al., 2019; Voss et al., 2019) revealed that such intervention systems provide positive outcomes that support ASD patients in strengthening their social skills. These systems also provide vital feedback to clinicians in evaluating and selecting the proper therapeutic intervention level to move forward.

Facial expressions are perceived to project 55% of an emotion recognition process (Mehrabian, 2017). In order to classify visual data into a corresponding emotion, machine learning techniques are required. Machine learning requires relevant and distinct descriptive features in order to make a proper decision. There are two methods for feature extraction, the first is image based and the second is geometric based. In image based feature extraction the image

10.1016/j.ifacol.2023.10.350

is passed through feature extractors i.e., convolution layers in deep learning neural networks or via other traditional means i.e., histogram of oriented gradient.

In the geometric based approach, the features rely on geometric relations between particular points in the image. In this approach the images are usually annotated with land-mark points at certain key areas that represent a particular image feature i.e. the mouth. The relations between these points are used as features for the machine learning process. Manual land-mark annotation is a tedious and time consuming process and error prone, as different people perceive important feature points differently even when following a common standard.

In this paper the use of image processing techniques coupled with Fourier series models is used to generate point land-mark annotations. The region of interest is focused on the mouth area. In this study the OULU-CASIA (Zhao et al., 2011) emotion database is used to test the performance of the annotation model.

The aim of this study is to establish an accurate automatic land-mark point annotation algorithm for use in feature generation in the development of an efficient and robust emotion recognition model.

2. SYSTEM DESCRIPTION

2.1 Methodology

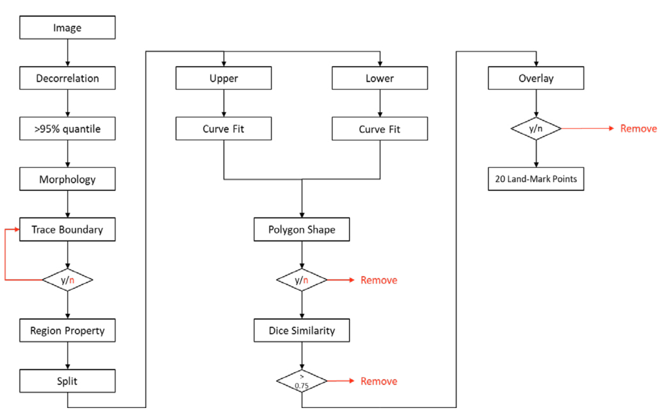

Fig. 1. Flow chart of the algorithm workflow. Starting from the original image input until the 20-point land-mark goal. y/n represent success and failure of the applied operation. Red colour represents the path of failed operation. Remove syntax is when the image did not pass though the relevant operation stage and subsequently was neglected and removed from further analysis.

The images of the OULU-CASIA database were first pre-processed based on previous work in (Arabian et al., 2021), in order to reduce background noise and focus on the face. Afterwards the images were enhanced with a de-correlation stretch to change the colour spectrum. A series of morphological, region detections and boundary traces followed.

The outcome from the image processing yielded a rough sketch of the mouth region. This rough estimation is then used as a basis to fit two Fourier series models that smooth the region of interest. A polygon shape is then created from the two curves. The points are then overlaid with a fine squared pattern mesh. Twenty points, distributed evenly, are then selected across the boundary as the point based land-marks.

A Dice (Dice, 1945) similarity metric is performed to analyse the adherence of the geometric reconstruction of the mouth region to the pixel image estimate. A threshold of 0.75 was set such that any value below this threshold was excluded from further analysis. Figure 1 represents the system workflow.

The mouth was selected to test the proposed method, as the area does not include obstructions unlike that of the eyes and eyebrows were artefacts i.e., eyeglasses can cause problems during the estimation process. The point land-marking of the eyes and eyebrows will be addressed in future works.

2.2 Image Enhancement

In order to distinguish the mouth more specifically the lips, from the surrounding skin tone, an image enhancement targeting the RGB colour spectrum was conducted. A linear pixel-wise de-correlation was performed to enhance the image colour space. A tolerance of 0.01 was set in order to achieve a separation between the facial features. The output was then raised to the power of 3 in order to brighten the skin tone to maximum while keeping the desired region of interest at a lower and separable colour code.

After obtaining an enhanced image, a threshold was chosen to convert the image into a binary. The threshold was set to the 95% quantile of the image, as each image's colour distribution differs. Next, the image is complemented such that the white regions become black and vice-versa. A series of morphological operations follow in order to improve the shape and eliminate any anomalies at the borders.

A region analysis is then performed to identify the shape with the largest area. This shape represents the mouth region. A boundary trace is performed to extract key starting points from the given shape. The shape is then split into top and bottom sections of the mouth, i.e. upper and lower lip. At this stage the shape of the mouth is rough and requires a smoothing operation.

2.3 Land-Mark Point Extraction

Two first term Fourier models are fitted with the pixel information. The fitting was performed based on the non-linear least squares method. A smooth representation of the upper and lower lips is obtained. After which new points are interpolated for a smoother and more detailed outcome feature representation. A fine square mesh grid is then overlaid with the upper and lower mouth curves. A limit of 20 points in total is assigned across the periphery of the polygon created from the combination of the 2 curves.

$$y = a_0 + \sum_{k=1}^{n} a_k \cos(kwx) + b_k \sin(kwx) \quad (1)$$

Equation (1) represents the Fourier series model used. Where a_0 is a constant, x is the input i.e. the pixel information of the top or lower lip boundaries, y is the output of the fitted curve, n is the number of harmonics and w the fundamental frequency of the signal. In this study n was set to 1.

The nearest grid point that is inside the polygon shape is considered, this is performed in order to have similar distance between two subsequent points which will be identified as the land-mark points. Figure 2 represents the fine square meshed grid with the polygon overlaid. The red points indicate the points that are inclusive while the blue dots are the grid points that are excluded.

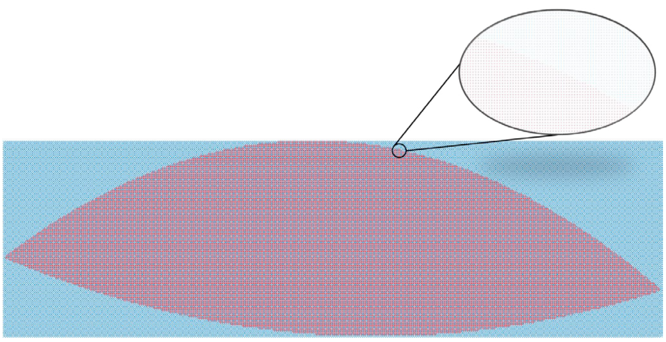

Fig. 2. Example of the grid point overlay. Red dots represent the grid points that belong within the polygon. Blue dots are the fine squared mesh grid points.

2.4 Database Description

The database chosen for the analysis process is the OULU-CASIA database. It is composed of image sequences from 80 participants expressing the six basic emotions of anger, disgust, fear, happiness, sadness and surprise (Zhao et al., 2011). Each sequence begins with a neutral expression and ends with a strong expression of the particular emotion. Image sequences of original colour, visible light with strong illumination were selected for this study.

For this approach the neutral expression was taken into consideration. The first two frames from each sequence were considered as neutral expressions and the last three frames of each sequence were set to be the corresponding emotion expression. The selected dataset consisted of 2,400 images in total.

3. RESULTS

In Table 1 the image distribution into each class, prior to and following the image processing algorithm, at different stages of the OULU-CASIA dataset is shown. The image pre-processing stage resulted in a loss of around 11.80% of the original image data, due to the inability of the region detection algorithm to correctly segment the areas of interest. During the image enhancement stage a reduction of around 19.93% was noticed with a further decrease of 0.18% in the final stage of point land-marking.

Table 1. Image distribution before and after image processing

Class	Original	Pre-Processed	Enhanced	Land-Marked
Anger	240	192	146	146
Disgust	240	182	152	152
Fear	240	233	178	178
Happiness	240	237	128	128
Neutral	960	911	807	805
Sadness	240	193	151	150
Surprise	240	169	133	133
Total	2400	2117	1695	1692

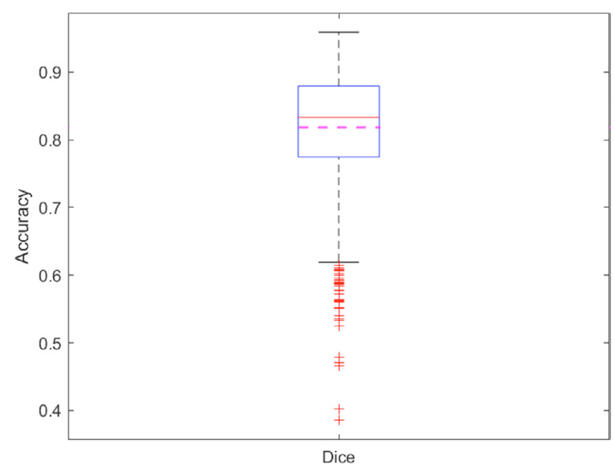

Fig. 3. Boxplot of the Dice similarity metrics of the different images in the selected dataset. The red solid line indicates the median and the dashed magenta line represents the mean. The red plus points are the outliers.

Fig. 4. Sample results of the 20-point land-mark system. The green plus signs represent the land-marks.

In figure 3 the Dice similarity metrics are represented as a boxplot. The statistical analysis revealed a median of 83.31% and a mean of 81.88% with a standard deviation of 8.14. Figure 4 represents the outcomes from the land-mark algorithm on different samples from the dataset.

4. DISCUSSIONS

The results from the image processing and enhancement revealed that the method was able to mark the key points of interest. However, a loss of around 30% of the original data was reported, which is quite high. The Dice coefficients alone revealed that 17.10% of the samples fall below the defined threshold. This set of rejected images represent more than half the data lost.

This loss is attributed mainly to the image enhancement process. An iterative tolerance and parameter tuning by an optimization technique would lead to a lower rejection rate and thereby retain more data which is vital for machine learning algorithms during the classification training process. This will be addressed in future works.

The mesh results from Fig. 2 showed that the geometric representation is quite detailed. The more data points available, the more flexibility is attained when it comes to selecting the number of land-mark points. This fine grid also helps in the feature extraction process e.g. the profile and direction of the curvature.

As observed from Fig. 4, the outcomes from the system are mixed. The left column shows the samples that outlined the shape of the mouth quite well. This indicates that the Fourier model is capable of representing the pixel data quite effectively. When looking at the right column samples of Fig. 4, it is observed that the landmarks fall short of a complete outline. This is related to the colour enhancement stage of the algorithm.

In the last sample of Fig. 4, right column bottom row, the land-mark points move towards a divergent pattern. This divergence at the boundaries is due to the relative distance chosen for the point interpolation. This anomaly occurs because the Fourier model is a harmonic series, and the algorithm was not designed to consider only one cycle.

The model simulation for the landmark point estimation recorded a run time of 2.12 ± 0.36 seconds. This duration is quite high, therefore the use in real-time simulations is not possible at this time. However, with further code optimizations the use in real-time could be achievable.

The proposed model is quite robust in the point land-mark quantity selection. In this study, 20 points were chosen as a starting point to verify the approach. More points can be selected for more feature generation without impacting the performance of the proposed approach. However, a balance must be determined to set the number of points that provide meaningful information without adding excess information that is redundant for the classification process.

Some limitations are taken into consideration during this study. The use of only a one term Fourier model, as well as the limitation thresholds of 95% quantile and 75% Dice measure.

5. CONCLUSION

In this study the design of an accurate automatic land-mark point annotation algorithm for use in feature generation was assessed. The outcomes revealed, that the Fourier models were quite effective in representing the pixel data in the geometric space with an 81.88% mean accuracy. However, some anomalies were detected in the 20-point land-mark system. The system showed a large potential in the automation of the annotation process that will reduce errors and conserve resources.

Further research is needed to enhance the pre-processing and image enhancement stages of the algorithm. An optimization model is being considered for this process for the fine tuning of the parameters to reduce the amount of discarded images.

6. ACKNOWLEDGMENT

Partial support by grants from the project KOMPASS funded by the Ministerium für Wissenschaft, Forschung und Kunst (MWK) of Baden-Wuerttemberg and the German Federal Ministry of Research and Education (BMBF) under project No. 13FH5I06IA – PersonaMed are gratefully acknowledged.

7. AUTHOR STATEMENT

Conflict of interest: Authors state no conflict of interest.

REFERENCES

Arabian, H., Abdulbaki Alshirbaji, T., Jalal, N.A., Ding, N., Laufer, B. and Moeller, K. 2022a. Identifying User Adherence to Digital Health Apps In: *IFMBE Conference Proceedings. SpringerNature*.

Arabian, H., Wagner-Hartl, V., Geoffrey Chase, J. and Möller, K. 2021. Facial Emotion Recognition Focused on Descriptive Region Segmentation In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*., pp.3415–3418.

Arabian, Herag, Wagner-Hartl, V. and Moeller, K. 2022b. Image Pre-processing Effects on Attention Modules in Facial Emotion Recognition In: *IFMBE Conference Proceedings. SpringerNature*.

Campbell, S. 2020. From Face-to-Face to FaceTime. IEEE Pulse. 11(5), pp.7–11.

Dice, L.R. 1945. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 26(3), pp.297–302.

Disabilities, C. on C.W. 2001. The Pediatrician's Role in the Diagnosis and Management of Autistic Spectrum Disorder in Children. *Pediatrics*. 107(5), pp.1221–1226.

Dorsey, E.R., Venkataraman, V., Grana, M.J., Bull, M.T., George, B.P., Boyd, C.M., Beck, C.A., Rajan, B., Seidmann, A. and Biglan, K.M. 2013. Randomized, controlled trial of “virtual housecalls” for Parkinson disease. *JAMA neurology*. 70(5), pp.565–570.

Golan, O., Ashwin, E., Granader, Y., McClintock, S., Day, K., Leggett, V. and Baron-Cohen, S. 2010. Enhancing Emotion Recognition in Children with Autism Spectrum Conditions: An Intervention Using Animated Vehicles with Real Emotional Faces. *Journal of Autism and Developmental Disorders*. 40(3), pp.269–279.

Lugo-Marín, J., Gisbert-Gustemps, L., Setien-Ramos, I., Español-Martín, G., Ibañez-Jimenez, P., Forner-Puntonet, M., Arteaga-Henríquez, G., Soriano-Día, A., Duque-Yemail, J.D. and Ramos-Quiroga, J.A. 2021. COVID-19 pandemic effects in people with Autism Spectrum Disorder and their caregivers: Evaluation of social distancing and lockdown impact on mental health and general status. *Research in Autism Spectrum Disorders*. 83, p.101757.

Mehrabian, A. 2017. Communication without words In: C. D. Mortensen, ed. *Communication Theory*. Routledge.

Rylaarsdam, L. and Guemez-Gamboa, A. 2019. Genetic Causes and Modifiers of Autism Spectrum Disorder. *Frontiers in Cellular Neuroscience*. 13, p.385.

Saanvi, A. 2020. IoMT (Internet of Medical Things): Reducing Cost While Improving Patient Care – EMBS. IEEE Pulse, a magazine of the IEEE Engineering in Medicine and Biology Society. (Best of 2020 Compendium).

Yuan, S.N.V. and Ip, H.H.S. 2018. Using virtual reality to train emotional and social skills in children with autism spectrum disorder. *London Journal of Primary Care*. 10(4), pp.110–112.

Zhao, G., Huang, X., Taini, M., Li, S.Z. and Pietikäinen, M. 2011. Facial expression recognition from near-infrared videos. *Image and Vision Computing*. 29(9), pp.607–619.