Investigation of an Innovative Approach for Identifying Human Face-Profile Using Explainable Artificial Intelligence

1st Nikhil Sarathy
FEIT, SEDE
University of Technology Sydney
Ultimo, NSW, Australia
Nikhil.Sarathy@student.uts.edu.au

2nd Mohammad Alsawwaf
FEIT, SEDE
University of Technology Sydney
Ultimo, NSW, Australia
College of CS & IT
Imam Abdulrahman Bin Faisal University
Dammam, Saudi Arabia
Mohammad.Alsawwaf@uts.edu.au
0000-0003-0359-9808

3rd Zenon Chaczko
FEIT, SEDE
University of Technology Sydney
Ultimo, NSW, Australia
Zenon.Chaczko@uts.edu.au
0000-0002-2816-7510

***Abstract*—Human identification is a well-researched topic that keeps evolving. Advancement in technology has made it easy to train models or use ones that have been already created to detect several features of the human face. When it comes to identifying a human face from the side, there are many opportunities to advance the biometric identification research further. This paper investigates the human face identification based on their side profile by extracting the facial features and diagnosing the feature sets with geometric ratio expressions. These geometric ratio expressions are computed into feature vectors. The last stage involves the use of weighted means to measure similarity. This research addresses the problem of using an eXplainable Artificial Intelligence (XAI) approach. Findings from this research, based on a small data-set, conclude that the used approach offers encouraging results. Further investigation could have a significant impact on how face profiles can be identified. Performance of the proposed system is validated using metrics such as Precision, False Acceptance Rate, False Rejection Rate and True Positive Rate. Multiple simulations indicate an Equal Error Rate of 0.89.**

***Index Terms*—Biometrics; Geometric Ratios; Pixel Segmentation; DNN; XAI; Identification; Feature Extraction; Human Profile Recognition; Feature Vector**

I. Introduction

Facial recognition, in general, is a base for many applications across multiple fields. These fields include but not limited to, surveillance, access control, entertainment, and identification [1]. Facial recognition as a research field includes multiple sub-fields. For example, feature extraction, land-marking, and expression detection [2]. Land-marking and extraction of facial features are widely used in the identification of humans because of their importance in creating the unique biometric signature for a given human [3]. However, many of the available approaches and even commercial products of identification are based on frontal face images [1], [2], [4]–[7].

Profile-based recognition, on the other hand, is becoming more recognized and researched because in many cases frontal images are harder to obtain and could even be impossible in some cases [5]. Just like the frontal recognition, profile-based one relies on the extraction of the features and their landmarks. Apparent features of individual profiles contain multiple features that could be of interest in achieving recognition. Ears, for example, are known to contain bio-metric signatures that are unique and precise. Moreover, other features like forehead, eyes, nose, lips, chin, and jawline are also visible inside images. They contain biometric signatures given that the subject in the image is not covering them with hair, scarf, cap, or any other obstacle (See Figure 1).

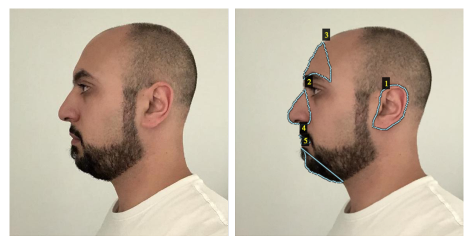

Fig. 1. Example profile image before and after segmentation

The process of face recognition [8] follows a generic flow that includes the following steps:

1) Face detection, to know that the image contains a recognizable face.
2) Extraction of the features needed to run the recognition algorithm.
3) Execution of the comparison or recognition algorithm and issuance of the results.

By locating the facial features on the face, it becomes possible to perform calculations and measures that can result in distinctive signatures between humans. the following adopted figure gives a visualization of the face recognition pipeline.

The aim of this paper is (i) to present an investigation of

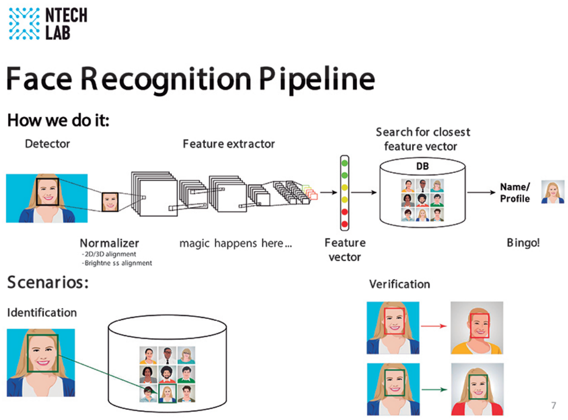

Fig. 2. NtechLab's Face recognition Pipeline [9]

an innovative approach to human face recognition based on profile images of the face using an explainable process, (ii) to explore the state of the art feature extraction methods using DNN with a strong focus on pixel segmentation approach, and (iii) to extract the unique geometric signatures of the facial profile features for similarity prediction. The adopted methodology attempts to overcome the known limitations with the explainability associated with the AI component used in similarity predictions [10]. The rest of this paper is divided as follows: Section 2 discusses related works. Then, Section 3 elaborates on the used approach. Section 4 explains the findings so far. Finally, section 5 provides a conclusion and future work.

II. Related Work

A. Holistic Nested Edge Detection

Holistic Nested Edge Detection (HED) is used for detecting edges and contours in images [11]. The ability to detect the feature's contours in a given image allows for a variety of opportunities to analyze and extract information for applications like object detection, tracking, medical imaging analysis, and motion detection. A Convolutional Neural Network (CNN) is used to train a model to detect the edges. This approach has much higher accuracy than the standard Canny Edge Detection approach [11]. Figure 3 shows HED prediction results of a sample input image. The feature contour points are then extracted using the predicted output.

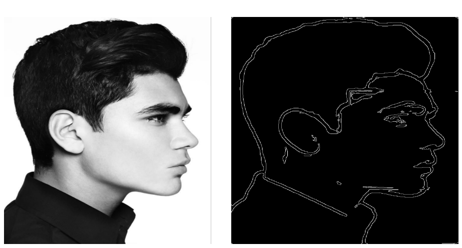

Fig. 3. Feature extraction phase: HED prediction results of a sample input image.

B. Segmentation

The image pixel segmentation process is an integral part of the feature extraction procedure. Segmentation, in this context, refers to identifying and selecting the different features on the face by annotating the image pixels around an entire feature. Thus, identifying the location of the feature which allows for the extraction of other valuable information. For example, finding a centroid and inter-features information like the geometrical positioning of the features on the face.

C. Geometric feature vectors

The use of feature vectors in face recognition has been previously explored since early works such as Jia and Nixon in [12]. The work presented in this paper adopts a novel approach to developing such vectors using geometric functions that considers a sequence of contour coordinates.

D. Inspiration for this work

The key hypothesis behind this approach is to measure the similarity of the side face profile is based on the following assumptions:

1) Human side face profile has sharp contours that are onus to the high degree of precision in feature detection, especially the ear.
2) The combination of various features of the side face can yield an array of geometric expressions - distance and angular ratios, and these can be unique to a person in combination with various facial positions.
3) These geometric measures can, therefore, attribute a signature to the face, which can be represented as a range of facial feature vectors. Consequently, the weighted similarity measure of these vectors can be used to predict the probability of a face match.

III. Methodology

The research work adopts a proof of concept experiment to develop a novel algorithm driven by the XAI approach. The algorithm has three basic functionalities as detailed further in below subsections - detecting features, extracting contour points and constructing geometric vectors for computing overall face similarity.

A. Experimental setup : Special considerations

It is important to highlight the assumptions that have been considered in this experiment.

- Input stream images are taken with good illumination where facial features are visible.
- Public dataset consisting of side faces of 420 subjects have been used for training and validation of the neural network-based prediction model. For the system performance measurement, five different subject's same face side were photographed.
- All the faces considered for the experiment have very less or no head rotation in both the vertical and horizontal axis.

- Ear has been considered as the key landmark, and all geometric computations involving the different face features use the ear centroid.
- Final validation images are segmented manually for the purpose of the experiment results, while on the other hand, a deep learning model is being trained to segment features automatically.

B. Facial feature detection: Deep Learning models

In this first step towards the facial feature extraction, two deep learning approaches are used for feature detection. The deep neural network (DNN) model, YOLO v2, has been adopted initially in which a custom model was trained using side profile faces (400 images) from publicly available datasets [13]. A loss convergence of around 3.0 was achieved. This resulted in reasonably good bounding box prediction results for features - ear, nose, forehead, eyes, mouth and chin.

The other alternative approach using the DNN is the pixel-based segmentation - Mask R-CNN model [14]. This custom model was trained using a sample set of side face images to primarily analyse the prediction output - the pixel boundary coordinates and the overall accuracy of the feature pattern points. However, the actual feature segmentation data was generated manually using image annotator tool [15] for reasons as highlighted in the assumption section. The dataset consists of around 50 images of side face for five different subjects, with a fixed resolution. After obtaining the detected features, the next step focuses on the contour points extraction for each feature. In the case of the bounding box results, the use of the pre-trained HED model along with the contour detection [16] yielded a set of contour points for each feature. The limitation of this approach is that clustered points often include outliers within the box which posed a challenge in constructing precise polygon representation of the feature as discussed in the further section.

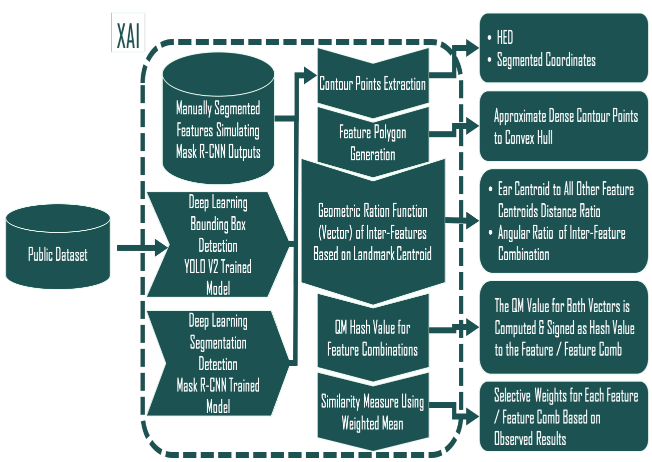

Fig. 4. Explainable Face-Profile Recognition Algorithm Flow : This process diagram shows the different steps in the methodology - facial feature extraction, geometric vector construction and similarity measure

C. Geometric Ratio Vectors: Inter feature properties

The feature detection follows the construction of the closed polygon shape from the contour points of the feature. The generated shape is further interpolated into a set of equidistant points. These points range from 30 to 50, depending on the feature and are used as the parameters of the vector function. Following this approach enables a point to point nodal distance and angular measurement between a combination of different features. For example, each point on the ear to all points on the nose. Points on the ear to points on the forehead and points on the forehead to points on the nose. These point-wise metrics are used to calculate a geometric signature expression for the entire feature - which is hypothesized as a unique geometric property for each feature / inter feature contour patterns [17]. In the experiments, two geometric functions to compute the feature vectors have been considered - 1. Distance ratio function, and 2. Angular function. Ratios and angular measures have been previously used in facial image recognition. In this approach, the application of functions on a set of a large number of points intensely exploits the unique inter-feature associations. For a brief example, this paper will highlight only a particular instance of a distance ratio function which is the ear-chin-nose association. The algorithm developed uses various inter-feature combinations iterated over every point of the contours. A feature set consisting of 15 and 9 have been used for distance ratio and angular functions, respectively. The following formulas (1) and (2) define the construction of a distance ratio feature vector.

$$f_{chin-nose_i} = \frac{d_{i-C_{ear}}}{d_{i-C_{nose}} * \cos\theta} \tag{1}$$

Where, $f_{chin-nose_i}$ denotes the point function of the specific feature combination while $d_{i-C_{ear}}$ and $d_{i-C_{nose}}$ denote the distances between a contour point and the respective feature centroid.

$$\mathcal{V}_{chin-nose} = \left\{ f_{chin_i-nose_i} \right\} \tag{2}$$

Where, $\mathcal{V}_{chin-nose}$ denotes the feature vector for the specific combination.

The vector representation of the geometric expressions computed from the sequence of coordinate points follow a specific order on the convex hull (polygon), ensuring that corresponding feature vectors of a different person are comparable. This concept of feature vector in this work is similar in context to an early experiment of [12], however differs significantly in its construction by the application of various geometric functions that associate features using relative centroids.

D. Feature Hash value and Similarity measures

From the multiple simulation experiments using the generated geometric feature vectors, interesting results have been observed. Initial attempts to measure the similarity of two similar feature vectors (ex: ear-nose distance ratio vector, ear - frontal features angular ratio vector) using cosine similarity, Euclidean and Minkowski distance resulted in inconsistent predictions - few false positives and false negatives. It can be inferred here that even when comparing feature vectors of same subject's different poses, sometimes there can exist significant difference in corresponding elements while the overall shape (area and line segment ratio) and positional (angle ratios) properties of the features equate to be similar. Therefore, to overcome these limitations, a more simplistic and explainable approach have been adopted - comparing the derived quadratic mean ($\mathcal{QM}$) of two similar vectors. These values prove to be distinctive and mostly consistent for the different feature combinations of a specific person. Hence, the $\mathcal{QM}$ of the vectors has been considered as the "hash" value to numerically represent the various features and their combinations. The expression of QM is given below as:

$$\mathcal{QM} = \sqrt{\frac{\sum_{i=1}^{n} \mathcal{R}_i^2}{n}} \tag{3}$$

The final step after attributing the features with the hash value ($\mathcal{QM}$) in formula (3) is the computation of the overall face similarity. The similarity score of two similar distance ratio vector for feature combinations or angle ratio vector for each feature is computed as the percentage difference in their hash value. Formula (4) denotes the calculation of the similarity using $\mathcal{QM}$.

$$\mathcal{S}_{similarity} = \frac{|\mathcal{QM}^1_{chin-nose} - \mathcal{QM}^2_{chin-nose}|}{max(\mathcal{QM}^1_{chin-nose}, \mathcal{QM}^2_{chin-nose})} \tag{4}$$

Table I below compares the quadratic mean ($\mathcal{QM}$) value of the distance ratio vector ($\mathcal{V}$) for various inter feature combinations of a specific subject against the prediction of different images- Subject 1. Each of the shown five combination's geometric computations are with respect to the ear's centroid.

TABLE I
QM (HASH VALUES) OF THE THE VARIOUS INTER-FEATURE COMBINATIONS OF SUBJECT-1

Feature	Picture ID of the Same Subject		
ID	*Pic1*	*Pic2*	*Pic3*
Eye-to-Nose	4.5175	5.1894	4.3468
Eye-to-Forehead	21.5947	19.1717	20.5560
Eye-to-Chin	2.2930	2.3617	2.2964
Nose-to-Nose	44.1705	45.1848	41.0454
Nose-to-Forehead	18.0604	17.0112	16.0835

Table II below compares the quadratic mean $\mathcal{QM}$ value of the angular ratio vector ($\mathcal{V}$) for the five side facial features of a specific subject against the prediction of different images - Subject 3. The angular feature vector is constructed using a combination of point to point as well as point to centroid angles in radians.

IV. FINDINGS AND DISCUSSIONS

A. Experiment Results

The results from the conducted experiments reveal interesting findings relating to the latent geometric properties of the side facial features. Furthermore, the typical challenges encountered during the feature contour extraction, and subsequent derivation of the inter feature geometric hash values from the experiments opened up new aspects for further research in this area. Standard performance metrics have been considered for validating the facial similarity predictions.

TABLE II
QM (HASH VALUES) OF THE FACIAL FEATURES OF SUBJECT-3 DERIVED FROM THE ANGULAR VECTOR

Feature	Picture ID of the Same Subject			
ID	*Pic1*	*Pic2*	*Pic3*	*Pic4*
Ear	0.5578	0.5617	0.5435	0.5549
Eyes	0.7768	0.7324	0.7327	0.7214
Nose	0.5736	0.5838	0.5775	0.5836
Forehead	0.5618	0.6398	0.6280	0.6215

Table III below indicates the performance metrics of the system captured during the multiple simulations. The threshold values represent the probabilistic measure of similarity between two given images of side faces. The change in the False acceptance rate (FAR), and False rejection rate (FRR) drops with the increase in the threshold whereas the Recall or True positive rate (TPR) starts decreasing. It can be inferred here that the average similarity score between two images of the same subject is around 0.96. Meanwhile the score of images of different subjects tend to range between 0.85 to 0.93. This implies that comparing side faces using geometric computations demands a high precision as there is only slight differences in the human side profile (considering the experimental setup)

TABLE III
PERFORMANCE METRICS OF THE XFPR ALGORITHM BASED ON MULTIPLE SIMULATIONS

Decision	Performance metric of the Algorithm			
Threshold	*Recall / TPR*	*Precision*	*FAR*	*FRR*
0.9610	0.91	0.70	0.14	0.10
0.9620	0.89	0.72	0.12	0.11
0.9630	**0.89**	**0.74**	**0.11**	**0.11**
0.9640	0.88	0.75	0.10	0.12

B. Pixel segmentation based DNN model for side face contours

The two approach that has been used to extract the detailed contour points from the side face provides a good insight of feature extraction. Although the HED based DNN model in conjunction with the YOLO-v2 deep learning model was effective in extracting detailed contour points, the limitations imposed by the outliers coordinates (from neighboring features) detected within the bounding box eventually impacts the accuracy of the similarity measures. The challenges can nevertheless be overcome to an extent by using more sophisticated DNN models such as YOLO v3, SSD (Single Shot Detector) etc. Furthermore, custom clustering methods specific to a feature pattern such as Ear can improve the precision of the feature vectors constructed using approximated polygons. Hence, It is believed this approach has a scope to be further improved and serve as a potential contribution in facial feature extraction techniques.

In the second approach using the manual pixel segmentation of features, extraction of precise set of contour points was achieved. This has inspired future work in the direction of using sophisticated DNN models such as Mask R-CNN, and experiments implementing this is currently in progress. The segmented pixels has a major advantage over the bounding box method since the ”region of interest” becomes more focused yielding far fewer outliers. This investigation propose the pixel segmentation based DNN model as one of the effective approaches for feature extraction with specific emphasis on full side face and other facial feature analysis. [18].

C. Distinctive geometric properties of the side face

Major findings from the experiments indicate the unique geometric values derived from the feature vectors for each feature such as ear,nose and forehead etc, and also other feature to feature combination. These geometric hash values have been observed to be largely consistent and within a distinctive range for each feature/feature combinations across multiple images of the same subject's side face. Therefore, this implies that the side face features of humans can be geometrically exploited to yield a set of distance and angular ratios unique to a person. Feature vectors can be then be constructed using these point functions to attribute hash values for the side face. Previous works that have applied the concept of geometric properties such as golden ratio have evidenced an effective approach for classifying human features. As demonstrated from the performance of the proposed algorithm (ERR rate of 0.89), this investigation can conclude that the methodology adopted is a useful contribution in the field of facial recognition research.

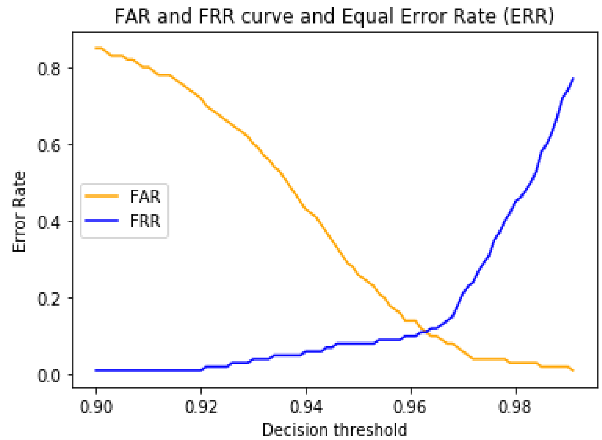

Fig. 5. This graph depicts the performance of the system under multiple simulations. The FAR and FRR curve meet at the error rate value of 0.11 and the EER is calculated as 0.89.

V. Conclusion

Analysis of the findings from the proposed novel methodology offer encouraging results. Considering the benchmark for performance in the area of facial profile recognition, the system is able to achieve a high level of prediction exceeding 0.89 accuracy. The proposed new algorithm can be improved further in the following areas:

- The automation of the pixel segmentation using deep learning models such as Mask R-CNN and SSD.
- Exploring more combinations of inter feature relationship
- Implementing the algorithm in the context of frontal face features.
- Future versions should include simulations using larger datasets.

References

[1] Yue Wu and Qiang Ji. Facial landmark detection: A literature survey. *International Journal of Computer Vision*, 127(2):115–142, 2019.

[2] Mrs Jyothi S Nayak, G Preeti, Manisha Vatsa, Manisha Reddy Kadiri, and S Samiksha. Facial expression recognition: A literature survey. *International Journal of Computer Trends and Technology (IJCTT)*, 48(1), 2017.

[3] Vikas Maheshkar, Suneeta Agarwal, Vinay Kr Srivastava, and Sushila Maheshkar. Face recognition using geometric measurements, directional edges and directional multiresolution information. *Procedia Technology*, 6:939–946, 2012.

[4] Celia Cintas, Mirsha Quinto-Sánchez, Victor Acuña, Carolina Paschetta, Soledad De Azevedo, Caio Cesar Silva de Cerqueira, Virginia Ramallo, Carla Gallo, Giovanni Poletti, Maria Catira Bortolini, et al. Automatic ear detection and feature extraction using geometric morphometrics and convolutional neural networks. *IET Biometrics*, 6(3):211–223, 2016.

[5] Boris Efraty, Emil Bilgazyev, Shishir Shah, and Ioannis A Kakadiaris. Profile-based 3d-aided face recognition. *Pattern recognition*, 45(1):43–53, 2012.

[6] Sima Soltanpour, Boubakeur Boufama, and QM Jonathan Wu. A survey of local feature methods for 3d face recognition. *Pattern Recognition*, 72:391–406, 2017.

[7] Yuhang Wu, Shishir K Shah, and Ioannis A Kakadiaris. Godp: Globally optimized dual pathway deep network architecture for facial landmark localization in-the-wild. *Image and Vision Computing*, 73:1–16, 2018.

[8] Mohammad Alsawwaf, Zenon Chaczko, and Marek Kulbacki. In your face: Person identification through ratios of distances between facial features. In *Asian Conference on Intelligent Information and Database Systems*, pages 527–536. Springer, 2020.

[9] Findface enterprise server sdk - facial recognition — kimaldi.com.

[10] Zenon Chaczko, Marek Kulbacki, Grzegorz Gudzbeler, Mohammad Alsawwaf, Ilya Thai-Chyzhykau, and Peter Wajs-Chaczko. Exploration of explainable ai in context of human-machine interface for the assistive driving system. In *Asian Conference on Intelligent Information and Database Systems*, pages 507–516. Springer, 2020.

[11] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

[12] Xiaoguang Jia and Mark S. Nixon. Extending the feature vector for automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12):1167–1176, 1995.

[13] Tiago F. Vieira, Andrea Bottino, Aldo Laurentini, and Matteo De Simone. Detecting siblings in image pairs. *The Visual Computer*, 30(12):1333–1345, Dec 2014.

[14] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[15] Abhishek Dutta and Andrew Zisserman. The vgg image annotator (via). *arXiv preprint arXiv:1904.10699*, 2019.

[16] Satoshi Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing*, 30(1):32–46, 1985.

[17] David Zhang, Qijun Zhao, and Fangmei Chen. Quantitative analysis of human facial beauty using geometric features. *Pattern Recognition*, 44(4):940–950, 2011.

[18] Dasari Shailaja and Phalguni Gupta. A simple geometric approach for ear recognition. In *9th International Conference on Information Technology (ICIT'06)*, pages 164–167. IEEE, 2006.