On Distinctiveness and Symmetry in Ear Biometrics

Manisha Aeri,
Asst. Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand India 248002,

Abstract
In this procedure, a gender categorization and authentication system have been implemented using ear biometrics. A lot of real-world applications need for the ability to recognise people and categorise them as male or female. The fundamental reference point for identifying a person has been thought to be the earhole. In order to determine the relative distances (Euclidean distance) amongst the ear identifying spots (ear features) and the ear hole, the features are taken. The outer lobe edge, the outer and inner curves of the helix, along with edges are taken into consideration. For the categorization, we used a dataset with both male and female participants. The major goal of this technique is to identify a person and determine their gender using a photograph of their ears.Moreover, to increase the process' accuracy and decrease the ear region's feature mismatching. 206

DOI Number: 10.48047/nq.2021.19.6.NQ21088 NeuroQuantology2021;19(6): 206-211

1. INTRODUCTION

Ears are a trusted biometric since they have long been thought to be distinctive to their owners. Ear identification has been relied upon holistic as well as model-based techniques, integrating current research incorporating deep learning. There are apparent benifits to employing an ear for biometric systems as it is resistant to expression, yet ears can be concealed by hair. Several techniques have been targeted at, and assessed on, standardised datasets where the ear position is regulated (containing both older and larger datasets, such as SC face, and earlier ones, such as XM2VTS). (Which involves subsequent and bigger datasets like SC face as well as earlier ones like XM2VTS). This is a vital stage in every biometric, as the vital element in biometrics is distinctiveness to a person.

We must be able toto translate ear biometrics into practical applications show and comprehend the capacity on unconstrained ear pictures, where the ear is not always in a plane orthogonal to the direction of the camera. As of yet, no. publicly available whether or if the ears are bilaterally symmetrical, and if so, which areas should be targeted for recognition. This piece covers studies aiming to study these difficulties, as either a part of the advancement of ear biometrics for applications across the laboratory. The current Covid-19 epidemic and the rising use of masks might impede the advancement of facial recognition technology.To do that, we must loosen restrictions about the installation of a security/surveillance systems so they may handle photographs in which the ear is turned or not presented normally to the camera. Since that only one ear is frequently visible, it is also wise to look for symmetry. HoweverIt is usually recommended to save two photographs per subject it is feasible that a match would be necessary when just one ear can be viewed as an alternative scenarios.

Although there have been several research on the use of ear pictures for person identification [1–6], there have only been a few on the extraction of soft biometric features from ear images, such as age and gender. To the extent possible, this research is the first effort on age categorization using ear pictures. The use of ear pictures for gender categorization has, however, been the subject of a few earlier studies [7], [8], [9], and [10]. In [7], the ear-hole is employed to serve as the measurement's starting point. The Euclidean distances among ear hole and seven characteristics of ear, which are detected from masked ear pictures, are computed. They was using an internal database, which includes 342 samples, for the studies.In [8], support vector machines with a histogram intersection kernel are used to classify profile face and ear pictures individually. They did score level fusion relied upon Bayesian analysis to increase the accuracy. For the studies, 2D photos from UND biometrics dataset collection F [11] were utilised. Fusion results in an accuracy rate of 97.65%, compared to 95.43% for face-only performance and 91.78% for ear-only accuracy. In [9], features were retrieved using Gabor filters, and classification was carried out using the features on the basis of dictionary learning. A test specimen is represented as a linear concatenation of the training data in the test phase using a dictionary that has been constructed from training sample data.

[10] Performs gender categorization on both 2D and 3D ear pictures. Automatic 3D ear detection and alignment is performed. Collections F and J2 of the UND were used for the tests [11]. SVM was used to extract and categorise features from the histogram of the indexed shapes. The system performed on average at 92.94%. We give a thorough investigation of how to identify age and gender from ear pictures in this research. For ear representation, we have investigated the use of both geometric elements and features according to appearance. Eight landmarks identified on the ear serve as the foundation for geometric elements. We estimated 14 alternative distances between these locations and two area computations to extract the characteristics from them.

Four distinct classifiers—logistic regression, random forests, support vector machines, and neural networks—have been used to categorise these collected characteristics. The appearance-based approaches are based on wellknown deep convolutional neural network models, particularly, AlexNet [12], VGG-16 [13], GoogLeNet [14], alongsideSqueezeNet [15]. They have undergone two iterations of refinement: once on a large-scale ear dataset to offer domain adaptability, and once using a modest target ear dataset. Geometric feature-based techniques have underperformed appearance-based algorithms in the experiments. Our gender categorization accuracy of 94% is higher than that of the earlier research' results. 52% accuracy in determining age has been attained.The following is a summary of the paper's contributions: In order to categorise age as well as gender from ear pictures, geometric and appearance-based criteria has been investigated. We have taken eight spots on the ear that serve as landmarks and deduced 16 geometric characteristics from them. For appearance-based techniques, we have leveraged a large-scale ear dataset [16], It was created from close-up and profile face images from the Multi-PIE face collection [17]. In this manner, we successfully applied as well as benefitted from the popular CNN designs to the relevant issue. Compared to the prior study, we have produced gender categorization results that are superior.

2. LITERATURE SURVEY

This research illustrates that the stochastic structure of very complicated activities may be captured by modelling the combined restrictions on the number of codewords in the video. We employ the sum-product network (SPN) to represent different BoW designs. SPN is a generalised directed acyclic network with terminal, sum, and product nodes as its three different types of nodes. The EM technique is used to learn BoW distribution parameters and SPN connection. The video's most probable explanation (MPE) for recognising activityas well aslocalisation is produced via SPN inference.A fresh volleyball dataset is assembled and evaluated. A sum-product network (SPN) of bags-of-words (BoWs) makes up the activity model. Fast and scalable recognition is made possible by SPN inference, which yields the most likely explanation (MPE) and features nodes that are linearly complicated in terms of number. The findings show that we outperform competing approaches and that we are rather insensitive to particular selections of the SPN height, breadth, number of points in the counting grid, along with sizes of BoWs over a set period of time [1].

In order to create a single detection process, we want to combine the categorization and localisationcomponents. We provide a rapid method that makes use of top-down activity information to quickly locate the area of a video that increases a classifier's score. We create a 3D graph where nodes represent local video subregions and connectedness is based on spatial and temporal closeness. Each node has a learning weight that indicates how much the action class of interest is supported by the node's appearance and motion. Due to the branch-and-cut solver, the detection solution is similar to an exhaustive sliding window search but takes orders of magnitude less time.Author demonstrates how to build more flexible graph structures that allow for "non-cubic" detection zones and even temporal hops that would otherwise confuse the classifier. Last but not least, we present a brand-new high-level descriptor that can be searched by sub-graphs and represents human positions with objects alongside their relative temporal ordering. The suggested technique is a quick approach that can assess a larger space of candidates than formerly practicable, which frequently results in more precise detection [2].

In order to portray complex human activities, this work aims to understand what activity pieces and their spatiotemporal relationships should be represented. It also explores their relevance for facilitating effective inference in realistic films. The learning process has two objectives: to become familiar with the activity model's structure as well as the PDFs connected to its nodes and edges. Then, using this model, additional movies are parsed in order to localise important activity portions that are present at various sizes. Our inference is making arbitrary permutations of nodes in spatiotemporal networks invariant, to allow for any mistakes in extracting video tubes.This study introduces a novel volumetric-based method to activity detection as well as video parsing that automatically learns the hierarchy, temporal organization, and spatial relationships of complex human activities. Within afavourable complexity-vs-accuracy trade-off, it exceeds the latest developments in standard Olympic and UT human-interaction datasets [3].

The use of CNNs to identify human activity in videos is discussed in this study. It suggests doing 3D convolution in the CNN's convolutional layers to collect distinguishing characteristics in both the spatial and temporal dimensions. Convolution and subsampling are carried out individually in each channel of a 3D CNN architecture that creates several channels of data from nearby video frames. The results of the studies demonstrate that the constructed 3D CNN model performs competitively on the KTH data and outperforms other baseline approaches on the TRECVID data. Prior research demonstrates that when such a model is pre-trained using unsupervised techniques, the number of labelled examples may be greatly decreased.In this study, the constructed 3D CNN model was trained using a supervised technique, which calls for a substantial amount of labelled data. In the future, we'll investigate 3D CNN models' unsupervised training [4].

This research provides a unique learning framework for action detection in video that concurrently executes temporal segmentation and event recognition in time series. Using labelled data and a multi-class SVM that maximises the separation margin between classes, the discriminative recognition model is trained. Dynamic programming is used to efficiently perform simultaneous segmentation and recognition once the model has been learnt, maximising the SVM score of the victorious class simultaneously repressing members of the lower classes. Experimental findings on the honeybee, Weizmann, as well as Hollywood datasets highlight the advantages of our strategy over cutting-edge techniques. Experimental evidence supporting common datasets demonstrated the competitiveness of our strategy over cutting-edge techniques.Researchers replicated the experiment using the previous five classes of activities as the null class in order to assess how well the suggested strategy performed in the presence of those classes. In comparison to MaxScoreSeg's accuracy of 77.9%, our technique produced an average accuracy of 93.3% [5].

3. PROPOSED SYSTEM

This study introduces a hybrid thresholding by fuzzy c-means (THFCM) algorithm-based strategy for segmenting images. The suggested method aims to identify a discerner cluster capable of determining an automated threshold. The frequency values (y-values) on the smoothed histogram are applied to the conventional FCM clustering method to create the algorithm. When choosing a discerner cluster for the grey level image, look for the cluster with the greatest peak, which in the image histogram shows the highest frequency. By the suggested method, experimental results with typical test photos have been attained. Due to the necessity for effective segmentation implementation, segmentation also has its limits. Being one of the most crucial steps in the design of any algorithm or application, segmentation must be done carefully in order to avoid oversegmentation. The straight analytical calculation cannot handle the complexity of the solution space for these two factors. These techniques could rely on certain domain-specific information that might not be relevant for other domains.

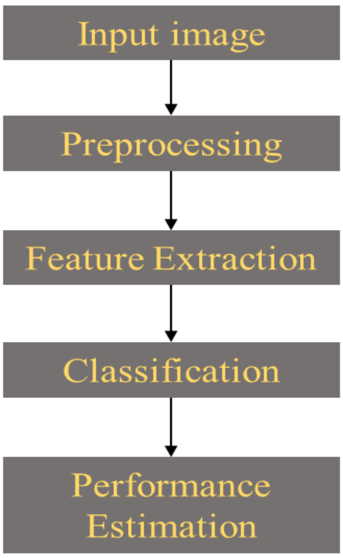

Fig 1: System Architecture

Several earlier studies have demonstrated how human ears may be utilised for identification, gender categorization, and lineage verification. They have also looked into whether a person's ears are symmetrical, although the results have been less than ideal. In order to identify the ear portions it is the source of recognition, our article expands the analysis of gender categorization on ear pictures along with examines how the human ears are symmetrical on both sides. To benefit from structure and performance, we employ deep learning techniques and model-based analysis, respectively. By imagining the ear as a flat surface connected to the head, we may check how ear images rotate when subjected to an affine transformation. We explore the subject of whether a person can be identified from their second ear after being shown a picture of one ear. Such a symmetry-based approach might ease restrictions on ear biometrics applications. Python and MATLAB were coupled to enhance the process's performance. The process's feature stability is improved by categorization in addition to features.

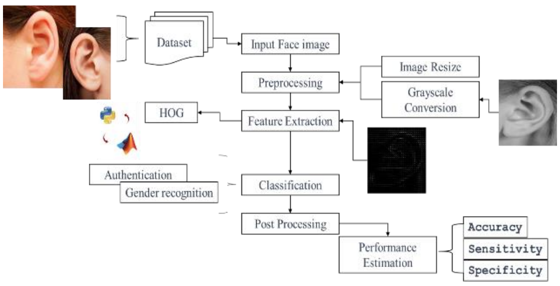

Fig 2: Flow Diagram

The next section provides an explanation of the many phases that are involved in putting the suggested technique into practise:

1. Input image

A rectangular array of numbers makes up a picture (pixels). Each pixel is a measurement of a different aspect of a scene over a limited region. There are several ways to measure the characteristic, but often we either measure the average brightness (one number) or the brightest areas of the image after applying red, green, and blue filters (three values). An eight bit integer is often used to represent the values, offering a brightness range of 256 levels. When we talk about an image's resolution, the terms "pixel count" and "brightness values" are used.

2. Preprocessing

The range of pixel intensity values can be changed in image processing by using the normalisation approach. Applications include images with poor contrast caused by glare. Histogram stretching or contrast stretching are other names for normalisation. In more generic data processing fields like digital signal processing, the term "dynamic range expansion" is employed. The act of moving an image or other type of signal into a more comfortable or normal for the senses range is referred to as "normalisation." Dynamic range expansion serves this function in a variety of applications.

3. Feature Extraction

A feature descriptor called HOG, or it is common practise to use the histogram of oriented gradients to extract characteristics from image data. It is commonly employed for object identification in computer vision tasks. Let's examine a few key features of HOG that set it apart from other feature descriptors: The HOG description emphasises an object's structure or form. You might be wondering how this differs from the edge characteristics we extract for photos at this point. We only determine When using edge features, determine if a pixel is an edge. The edge direction can be provided via HOG as well.To do this, the gradient and orientation are separated (or magnitude either direction, depending on your preference) of the edges.These orientations are also computed in "localised" sections. This implies that the entire image is divided into smaller sections, and the gradients and orientation are determined for each region. In the next parts, we'll go into considerably more depth about this. In the end, the HOG would create a distinct Histogram for each of these zones. The phrase "Histogram of Oriented Gradients" describes the histograms that are produced utilising the gradients and orientations of the pixel values.

4. Classification

An individual quantifiable quality or characteristic of a phenomena is referred to as a feature in machine learning and pattern recognition. Effective pattern recognition, classification, and regression algorithms must carefully select informative, discriminating, and independent characteristics. Although structural characteristics like strings and graphs are employed in syntactic pattern recognition, features are often numerical in nature. The term "feature" has a connection to the term "explanatory variable," which is used in statistical methods like linear regression.

5. Performance Estimation

Statistical measurements of the effectiveness of a binary classification test, commonly known as a classification function in statistics, include sensitivity and specificity: Sensitivity is the percentage of positives that are accurately classified as such (also known as the true positive rate, the recall, or likelihood of detection in some industries) (i.e., the proportion of patients who are diagnosed with their illness accurately.). Specificity measures the percentage of negatives that are accurately classified as such (also known as the true negative rate) (i.e., the percentage of people who are in good health who are appropriately diagnosed as not having the illness.)

4. RESLUTS

In order to recognise ears, classify genders, and recognise ears with bilateral symmetry, this study introduces model-based and deep learning techniques. In order to identify the ear portions from which recognition is derived, the research expands the analysis of gender categorization on ear pictures and investigates the bilateral symmetry of human ears. Accuracy rates for ear recognition, gender categorization, and ear recognition with bilateral symmetry using the model-based technique are 95.1%, 82.9%, and 66.9%, respectively. Accuracy rates for the deep learning approach are 92.9%, 90.9%, and 93.1%, respectively. The outcomes show symmetry to be 100% effective.

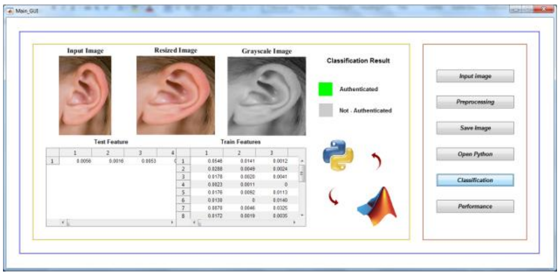

Fig 3: Classification

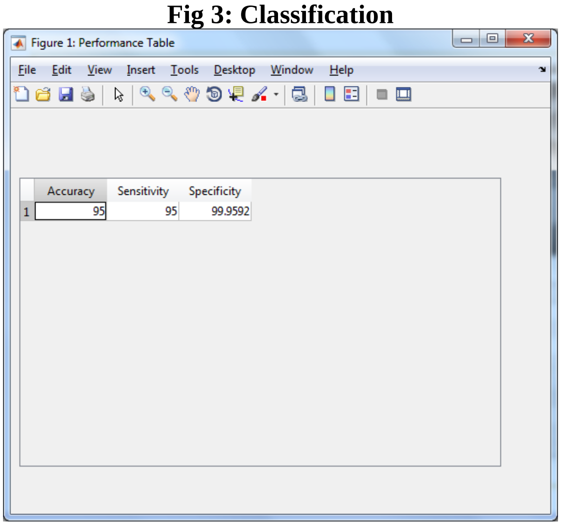

Fig 4: Performance Metrics

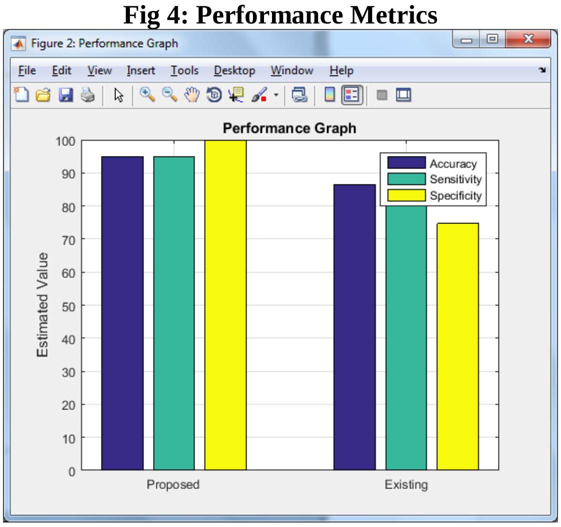

Fig 5: Performance Analysis

5. CONCLUSION

For ear recognition with bilateral symmetry, gender categorization, and deep learning, this research proposes model-based and deep learning approaches.The accuracy rates for ear recognition, gender categorization, and ear recognition with bilateral symmetry using the model-based technique are 95.1%, 82.9%, and 66.9%, respectively. The performances for ear identification with bilateral symmetry, gender classification, and deep learning are 92.9%, 90.9%, and 93.1%, respectively. Additionally, deep learning is used to recognise pairs of ears. Our findings provide 100% performance, confirming symmetry. Being able to recognise ears with bilateral symmetry using deep learning is a first for us.

6. FUTURE ENHANCEMENT

The feature tracking algorithm is the Kanade Lucas Tomasi algorithm. It is a really well-liked one. The KLT algorithm was first developed by Lucas and Kanade, and Tomasi and Kanade later expanded on their work. This approach is used to find feature points that are dispersed but yet have enough roughness to allow for accurate tracking of the needed locations. Here, the Kanade-Lucas-Tomasi (KLT) algorithm is utilised to track human faces constantly throughout a frame of video. They do this by identifying the parameters that provide a decrease in the dissimilarity measurements among feature locations connected to the original translational model.

REFERENCE

[1] A. Huang, S.-Y. Kwan, W.-Y. Chang, M.-Y. Liu, M.-H. Chi, and G.-S. Chen, "A robust hair segmentation and removal approach for clinical images of skin lesions," in Proc. 35th IEEE Int. Conf. (EMBC), Jul. 2013, pp. 3315–3318.
[2] H. C. Shih, "An unsupervised hair segmentation and counting system in microscopy images," IEEE Sensors J., vol. 15, no. 6, pp. 3565–3572, Jun. 2015.
[3] Y. Wang, Z. Zhou, E. K. Teoh, and B. Su, "Human hair segmentation and length detection for human appearance model," in Proc. 22nd IEEE Int. Conf. (ICIP), Aug. 2014, pp. 450–454.
[4] D. Wang et al., "A novel two-tier Bayesian based method for hair segmentation," in Proc. 16th IEEE Int. Conf. (ICIP), Nov. 2009, pp. 2401–2404.
[5] D. Wang, X. Chai, H. Zhang, H. Chang, W. Zeng, and S. Shan, "A novel coarse-to-fine hair segmentation method," in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), Mar. 2011, pp. 233–238.
[6] D. Wang, S. Shan, H. Zhang, W. Zeng, and X. Chen, "Data-driven hair segmentation with isomorphic manifold inference," Image Vis. Comput., vol. 32, no. 10, pp. 739–750, Oct. 2014.
[7] K.-C. Lee, D. Anguelov, B. Sumengen, and S. B. Gokturk, "Markov random field models for hair

and face segmentation," in Proc. 8th IEEE Int. Conf. (FG), Sep. 2008, pp. 1–6.

[8] N. Wang, H. Ai, and F. Tang, "What are good parts for hair shape modeling?" in Proc. CVPR, Jun. 2012, pp. 662–669.

[9] N. Wang, H. Ai, and S. Lao, "A compositional exemplar-based model for hair segmentation," in Proc. 10th Asian Conf. Comput. Vis., Nov. 2010, pp. 171–184. [10] P. Aarabi, "Automatic segmentation of hair in images," in Proc. IEEE Int. Symp. Multimedia, Dec. 2015, pp. 69–72.

[11] P. Julian, C. Dehais, F. Lauze, V. Charvillat, A. Bartoli, and A. Choukroun, "Automatic hair detection in the wild," in Proc. 20th Int. Conf. Pattern Recognit. (ICPR), Aug. 2010, pp. 4617–4620.

[12] C. Rousset and P. Y. Coulon, "Frequential and color analysis for hair mask segmentation," in Proc. 15th IEEE Int. Conf. (ICIP), Oct. 2008, pp. 2276–2279.

[13] Y. Yacoob and L. S. Davis, "Detection and analysis of hair," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 7, pp. 1164–1169, Jul. 2006.

[14] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut': Interactive foreground extraction using iterated graph cuts," ACM Trans. Graph., vol. 23, no. 3, pp. 309–314, Aug. 2004.

[15] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in Proc. ECCV, May 2006, pp. 1–15.

[16] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg, "Retrieving similar styles to parse clothing," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 5, pp. 1028–1040, May 2015.