EAR BIOMETRIC TECHNIQUES: A COMPARATIVE APPROACH

[1]Tejal Nanaware, [2]Siddharth Nandargi, [3]Kavya Parag, [4]NehaPote, [5]ChandanSingh Rawat

[1]Student, [2] Student, [3] Student, [4] Student, [5]Associate Professor
[1]Electronics and Telecommunications,
[1]Vivekanand Education Society' Institute of technology, Mumbai, India

Abstract : Biometric technology has always been about innovation and new technology. It was the world of passwords and identity proofs that ruled the world of security before biometrics happened. Ear biometric is invariant from childhood to early old age as compared to other biometrics. Ear biometric is a non-invasive type of recognition. This paper presents a delineated comparison between two techniques which are used for recognition based on various characteristics of the ears. Use of SURF features along with K NearestNeighbours classification technique has been incorporated along with the training of Convolution Neural Network model where the features extracted in the first few layers are classified in the last layer. The experimental results show the relationship between computational time taken and varying number of features for different scenarios with the aid of KNN classifier, also the proposed approach of recognition by training CNN model gives an accuracy of 86% on AWE database.

***IndexTerms* – Ear biometrics, convolutional neural network, SURF, K-NN classifiers**

I. INTRODUCTION

Human Identification is an emerging field in research and technology. As more systems involving human identification using biometrics are developing, the existent methods are exhausted. Ear being a unique part of the body is recently used for biometrics technology. Ear features, shape, posture, appearance and structure do not age much with time. Ear data is rigid and remains the same with minor changes between eight and eighty years. Ear recognition has an advantage over facial recognition in the identification of identical twins due to indiscriminate characteristics. Outer part of human ear called Auricle does not change with age. The background also plays an important role in image segmentation. Facial recognition is a failure in case of twins, also facial characteristics change with a person's growth. It is considered better than facial recognition because the background is predictable in the case of ears. Occlusions, view point and pose variations can affect the recognition status. Challenges to the use of the ear to identify people, including hair on the ear that obscures a large part of them, headscarf worn by Muslim women to cover their hair therefore cover their ears and the level of illumination. Apart from the previous reasons, all other conclude in deciding ear as a suitable choice for human biometrics. In the recent times, deep convolutional neural networks have proved an excellent approach to solve complex problems. A lot of research regarding the application of CNN on complex problems resulting in notable success has been carried out in the recent years. In this scope of paper, VGG Net has been used which contains the definition of each layer with pre-trained set of weights. VGG neural network is built using convolution layers and can find the object name of image. SURF features have been used because of their robustness and fast computation, KNN is used to classify the test image into one of the predetermined classes.

Medical study has shown that change of ear be clear during the period from 4 months to 8 years old and over 70 years old. There are many advantages of using the ear as a source of data for human identification such as:

1- Ear does not change considerably during human life.
2- Ears have both reliable and robust features which are extractable from a distance.
3- Ear prints could be printed at a scene of crime.

The rest of the paper is organized as follows, section II discusses the related work, section III describes the proposed approach and its phases, section IV presents the results and discussions, and finally conclusion and future work are provided in section V.

II. RELATED WORK

The goal of this section is to provide an overview of two popularly used methods in ear biometrics. A comprehensive study of different domains of image processing and neural networks has been compared.

Depending on the type of feature extraction technique used, 2D ear recognition approaches can be divided into geometric, holistic, local and hybrid methods. Machine learning techniques are used in different applications such as ones used in [1] and [2]. On the other hand, many approaches and several researches have been proposed to extract unique features from human ear for people identification.

Mark Burge and Wilhelm Burger presented one of the most cited ear biometric methods in the literature [1]. They located the ear by using deformable contours on a Gaussian pyramid representation of the image gradient. Then they constructed a graph model from the edges and curves within the ear, and invoked a graph-based matching algorithm for authentication. They do not report any performance measures on the proposed system.

Yuizono et al. treated the ear image recognition problem as regular search optimization problem, where they applied a Genetic Algorithm (GA) to minimize the mean square error between the probe and gallery images. They assembled a database of 660 images corresponding to 110 persons. They demonstrated an accuracy of 99-100%. Like other biometric traits, research in ear recognition is directed by the databases that are available for algorithm evaluation and performance analysis. Therefore, we first discuss the various databases that have been assembled by multiple research groups for assessing the potential of ear biometrics. Pyramid and sequential similarity computation to speed up the detection of the ear from 2D images.

Morento et al. were the first to describe a fully automated system for ear recognition. They used multiple features and combined the results of several neural classifiers. Their feature vector included outer ear points, ear shape, and wrinkles, as well as macro-features extracted by a compression network. To test that system, two sets of images were acquired. The first set consisted of 168 images pertaining to 28 subjects with 6 photos per subject. The second set was composed of 20 images, corresponding to 20 different individuals.

To the best of our knowledge, no self-generated models are mentioned in CNN literature. The lack of research in the field is due to the lack of a large-scale ear dataset.

METHODOLOGY

In the comparative study, the first method used is SURF. The proposed approach consists of four main phases: Pre- processing, Ear detection, Feature extraction and Classification as shown in figure (1) and will be described at the rest of this section.

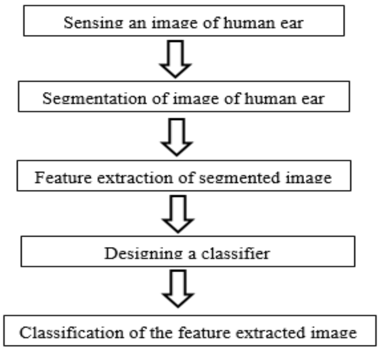
Sensing an image of human ear
Segmentation of image of human ear
Feature extraction of segmented image
Designing a classifier
Classification of the feature extracted image

Fig. 1 Model for ear recognition approach

Phase 1- Pre-processing:In this phase, we convert the image of the ear to gray scale level according to equation (1) as follows:

Gray Image = 0.2989R + 0.5870G + 0.1140B (1)

Then we applied the median filter to enhance image and remove noise

Phase 2- Feature Extraction: This step uses SURF technique for feature extraction which provides representation of an image in terms of a set of salient feature points. Equal number of feature values must be present in the feature vector for classification purpose, in order to obtain equal number of SURF feature 10 strongest SURF features have been taken from image. For this we need to consider those points which shows very high amount of intensity variation. Hence SURF algorithm makes use of Hessian matrix.

$$H(P,\sigma) = \left[L_{xx}(P,\sigma)\, L_{xy}(P,\sigma)\, L_{yx}(P,\sigma)\, L_{yy}(P,\sigma)\right]$$

where Lxx(P,σ), Lxy(P,σ), Lyx(P,σ) and Lyy (P,σ) are the convolution of the Gaussian second order derivatives, and with the image I at point P respectively. To speed up the computation, second order Gaussian derivatives in Hessian matrix are approximated using box filters.

The SURF detector algorithm can thus be summarized by the following steps:

1. Form the scale-space response by convolving the source image using DoH filters with different σ.

2. Search for local maxima across neighboring pixels and adjacent scales within different octaves.

3. Interpolate the location of each local maxima found

4. For each point of interest, return x, y, σ, the DoH magnitude, and the Laplacian sign.

Phase 3- Designing a classifier – Depending on the designing of a classifier the accuracy of the pattern recognition system varies. In classification process, the classifier classifies the given input image into one of the predetermined classes or into a brand new class based on the designing of the classifier. When an input image arrives at the classifier for classification purpose its features are extracted and compared with the features of the images which are present in the database. If most of the features of the input image are matching with the features of any one of the image present in the database, then the input image is said to belong to the class of that image which is present in the database.

In this method, the SURF extracted features are classified using the k-nearest neighbours algorithm (k-NN). K-nearest neighbours is a method used for classification and regression. A peculiarity of the k-NN algorithm is that it is sensitive to the local structure of the data. The training phase of the k-NN algorithm consists of storing the feature vectors and class labels of the training samples. In the testing phase, the features of the image to be classified are extracted and are plotted on to the feature space. The distance between features of testing image is compared with features of each and every class, and then the test image is assigned to the class which is near to the test image. Amongst various methods Euclidean distance is usually used to calculate distance between features of test image and features of training images in feature space.

Convolutional Neural Networks- In the next method, Convolutional Neural Network (CNN's) is applied for ear recognition. CNN is an end to end system that extracts all types of features in the images [13]. The model is trained in various layers in order to obtain low-level and high-level features in the way a traditional feature extraction algorithm does. The parameters obtained through forward and back propagation results in a learned feature being more generative than the traditional features [11].

In this paper, the Smaller VGGNet architecture is used for training the model. The VGGNet network was introduced by Simonyan and Zisserman in their 2014 paper, *Very Deep Convolutional Networks for Large Scale Image Recognition* [10]. The reason for using a compact version is the quality of the images. As the ear images aren't of high dimensionality, we use the Smaller VGGNet model. Being a compact version, it uses 8 layers instead of 16 layers [8]. The two important phases of the CNN approach are: Phase 1- Selection of a dataset: The available datasets on the internet contain a small number of samples. The requirement for CNN is a minimum of thousands of images. Fine-tuning can be achieved but it isn't enough to train a deep CNN model from scratch. The more the number of images or the availability of datasets, the better trained is the model. In the following case, we have used the AWE dataset. The AWE dataset contains 10 images of 100 people, thus summing up to 1000 images.

Phase 2- CNN Architecture: The deep learning environment is set up. The initial layers of the VGGNet architecture can be characterized by the 3x3 kernel convolutional layers stacked on top of one another for increasing the depth and better feature extraction. ReLU activation function adds non-linearities in the network for a more robust model. Max pooling is used to reduce the image size. A soft-max classifier is used in the last layer of the model for classification using probability values [12].

The four parameters concerning images are width, height, depth - number of channels and classes of the image. Images under consideration are resized to 128×128×3 for hardware limitations. Firstly, we perform a single convolution operation and ReLU activation to convert image with 3 channels into a feature map of 32 channels.

Dropout is used in the network architecture as it helps by reducing the overfitting of the network. Dropout works by randomly disconnecting nodes from the current layer to next layer. This process of random disconnects during training batches helps naturally introduce redundancy into the model. No one single node in the layer is responsible for predicting a certain class, object, edge, or corner.

For the next part of the block the layers added are (CONV => ReLU) × 2 layers followed by max pooling. This multiple stacking of convolution and ReLU helps in obtaining a rich set of features. We increase the filter size, while reducing the max pooling filter dimensions. A dropout layer is again used in this part of the block. Dropout of 25% is used to reduce overfitting. The same block is repeated before the FC layer for higher abstraction.

Before using the final classification, a fully connected layer with ReLU activation functions is used. The fully connected layerhas ReLU activation and batch normalization. Dropout is performed for a final time with 50% of the nodes during training. The reason for a higher percentage of dropouts is the fully connected layer. Lastly, we run the model with a softmax classifier that will return the predicted probabilities for each class label.

IV. RESULTS AND DISCUSSION

The comparative experiment was carried out on a personal computer having specifications as 8gb ram and ram speed of 2133MHz. All other programs were closed on the hardware. The simulation was performed with no multitasking and in the presence of a few background applications that are inevitable to shut down. The software simulation was performed on MATLAB and Google Collaboratory. Google Collaboratory is an online platform that allows deploying live codes using Jupyter Notebook, an open source web application and running them on TPU's (Tensor processing Units) which is a hardware accelerator. The code is implemented using Keras library along with TensorFlow backend.

Ear pattern recognition has been done using SURF feature extraction algorithm and K Nearest Neighbors classification algorithm. This work has been carried out on 20 test images by changing the number of images present in the database and also by taking number of SURF features taken. This has influenced operation time of this algorithm which is mentioned in following Tables 4.1-4.4.

Table 4.1 Computation time for one image

SURF features taken	Computation time (in seconds)
5	1.10
10	1.30
20	1.43

Table 4.2 Computation time for five images

SURF features taken	Computation time (in seconds)
5	1.94
10	2.08
20	2.22

Table 4.3 Computation time for fifteen images

SURF features taken	Computation time (in seconds)
5	2.47
10	2.63
20	3.01

Table 4.4 Computation time for twenty images

SURF features taken	Computation time (in seconds)
5	3.07
10	3.17
20	3.20

From Table 4.1 to 4.4 it can be observed that the run time of the algorithm will increase as the features taken and the number of images taken will increase. From Table 4.3 and 4.4 it can be observed that time required for computing 5 features for 20 images is not twice as great as time required to compute 5 features for 10 images, hence relationship between delay and number of images is non- linear.

Time required to calculate 10 SURF features for 10 images is 2.63 seconds while the time required to calculate 20 SURF features for 20 images is 3.20 seconds, hence time required for calculating first 10 features for first 10 images is 2.63 seconds but the time required to calculate next 10 SURF features of next 10 images is mere 0.57 seconds. Hence, while dealing with larger database, it is wise to feed entire database as the input to the program rather than splitting the database. Accuracy obtained is 100% for the 20 images input.

The recognition results of the Smaller VGGNet for the 100-class dataset are from 82% to 88%. The average accuracy obtained is 86%. The computational time required for classification of a single image is on an average 1.25 seconds and is the same for testing a batch up to 64 images. Hence when a set of 20 images were classified, the computational time was 1.25 seconds.

Table 4.5 Accuracy and time for different image size

Image Size	Accuracy	Time Required
96×96	80%	0.8 seconds
128×128	86%	1.25 seconds

From Table 4.5, it can be observed that changing the image size, helps achieve better accuracy while simultaneously increasing the computational time. For an image size of 256×256, greater accuracy can be achieved with a slight increase in the computational time.

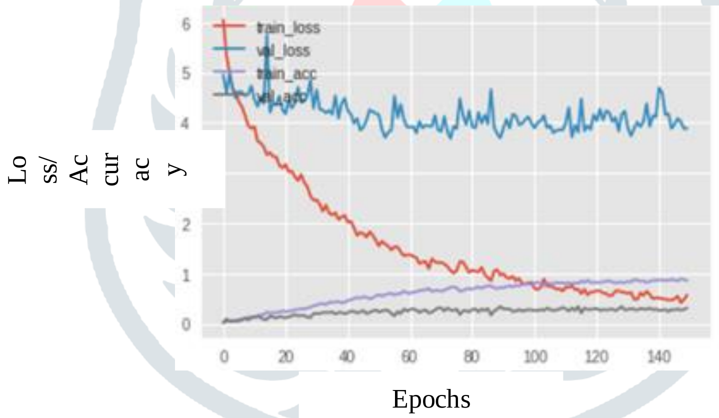

Fig.2 Training Loss and Accuracy

Fig.2 shows the loss and accuracy of the training model with respect to epochs. As the network is trained the validation and training accuracy increases and the validation and testing loss decreases exponentially.

```
[INFO] 011: 98.64%
```

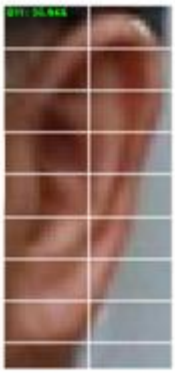

Fig.3 Classified ear image

Fig.3 shows a test image from the ear database. The percentage score indicates the confidence of the model that the ear belongs to subject number 011. Jupyter Notebook uses matplotlib, hence uses grids to plot the classified ear result.

The SURF method was tested on 20 test images, if this was to be carried out on a larger dataset of 100 images, the training time would increase drastically. As in the case of CNN's, the training time required for all the AWE dataset images is relatively shorter.

V. CONCLUSION

In this paper, two methods for ear recognition and matching have been discussed. Their comparative study delineates between the two methods on basis of performance metrics like accuracy and computation time. SURF uses blobs in order to detect intensities and extract features while through the training of a larger dataset, the Smaller VGGNet can extract more effective features and obtain better classification results than the traditional methods like SURF.

The contemporary method of Deep Convolutional Neural Networks has proven to be much more effective. Hence, it can be concluded that Neural Networks can be used for high efficiency and accuracy in real time applications.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[5] J Zhang, X Nie, Y Hu, "A method for land surveying sampling optimization strategy", *Geoinformatics 2010 18th International Conference on IEEE*, pp. 1-5, 2010.

[6] Hinton, G.E. and Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. Science, 313 (5786):504–507, 2006.

[7] Duchi, John, Hazan, Elad, and Singer, Yoram. Adaptive subgradient methods for online learning and stochastic optimization. The Journal of Machine Learning Research, 12:2121–2159, 2011.

[8] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv:1409.1556 [cs.CV], 2015.

[9] L. Tian and Z. Mu, "Ear recognition based on deep convolutional network," 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, 2016, pp. 437-441. doi: 10.1109/CISP-BMEI.2016.7852751

[10] FevziyeIremEyiokur, DogucanYaman, Hazım Kemal Ekenel. Domain Adaptation for Ear Recognition Using Deep Convolutional Neural Networks, arXiv:1803.07801v1 [cs.CV], 2018.

[11] Y. LeCun, F.J. Huang, and L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 2, pages II–97. IEEE, 2004.