



A NOVEL EAR DETECTION TECHNIQUE IN NOISY ENVIRONMENT USING MEAN STRUCTURE SIMILARITY INDEX MEASURE

V. Ratna Kumari
Department of ECE

PVP Siddhartha Inst. of Technology, Vijayawada, A.P., India.

Dr. P. Rajesh Kumar
Department of ECE

Hyderabad Institute of Tech. & Management, Hyderabad, T.G., India

Dr. B. Leela Kumari
Department of ECE

JNT University Kakinada, Kakinada, A.P., India

Abstract - In this paper, an efficient automated ear detection technique that uses skin colour and template-based methods is presented. There are two steps involved in the proposed ear detection technique. The first step in this technique is to remove non-skin pixels from the input image, which is called skin segmentation, and the second step is to detect the ear structure using Structure Similarity Index Measure (SSIM). Furthermore, the proposed technique can also be used in noisy environments to detect different shaped ears. This is made possible by creating an ear template using Principal Component Analysis (PCA) and Mean Structure Similarity Index Measure (MSSIM) for template matching and ear localization. An evaluation of the proposed technique is performed using the UND-E Database, which consists of 464 side face images of 114 subjects. It has achieved 97% accuracy in detecting the ear even when the ear image has additive Gaussian, impulsive salt-pepper and multiplicative speckle noise with variance $1/m$, with template size of $m \times n$. In particular, the proposed technique is well suited for detecting the ear image of a person moving as it can also detect the ear in motion blurred images.

Keywords- Skin-segmentation, ear detection, PCA, Gaussian Noise, Salt-Pepper Noise, Speckle Noise and Motion Blur Image, MSSIM.

I. INTRODUCTION

In recent studies, researchers have found that the external ear can be used as an attribute for recognizing and identifying humans [1]. Two steps are involved in human identification, the first is ear extraction from a side profile

and the second is ear recognition. Extraction and detection of the ear affects recognition accuracy. To achieve a reasonable recognition rate, the ear part has to be extracted properly from the side profile face image. By providing some initial input, a variety of algorithms have been implemented to automatically extract the ear. Automatic ear detection is essential to identify humans in real time, but it is challenging since the appearance of the ear can vary depending on the viewing angle and lighting conditions. Surya Prakash et al. devised a technique for automatic ear detection that uses normalized cross correlation for ear localization. For the IITK database, ear detection rate is reported as 94% [2].

A snake-based method of background removal and ear localization was proposed by Deepak et al. to achieve ear detection within 4 seconds, but the accuracy was less than 85% [3]. PP Sarangi et al. proposed a modified Hausdorff distance method for ear localization that achieved an accuracy of 94.54 % [4]. Emersic et al. and Iyyakutti et al. applied Convolution networks to ear detection in an unconstrained environment and achieved good detection accuracy [5, 6].

Resmi et al. proposed yet another method based on banana wavelets and circular Hough transforms, comparing its detection accuracy with the existing four algorithms. All these five methods were applied on GTAV database and RR databases. The accuracy of ear detection is between 84 to 94 % using manual and automatic verification methods, including the proposed method [7].

All the above methods were applied on high quality images and achieved good detection rate. However, these methods failed in some instances when applied to images of poor quality or noisy images.



In this paper, a new method for automatic ear detection is presented that works with noisy images. This paper is organized in five sections. First, various works in the field of ear detection from the literature are presented and the need for ear detection in a noisy environment is discussed. The second section discusses how the skin segmentation reduces the amount of search space required for ear detection. In the third section, the creation of a template using PCA is explained to handle different ear shapes. The fourth section discusses the proposed new algorithm and its implementation. In the last section, the results obtained with the proposed technique for ear detection are summarized.

II. ADOPTED TECHNIQUES OF SKIN TONE AND STRUCTURE SIMILARITY INDEX

Due to the differences in skin tones between people of different regions, it is challenging to develop a uniform method for segmentation and human skin detection. It is clear from the literature that there are a variety of color spaces used to detect skin colors [8]. RGB color space, however, is not preferred due to its non uniform characteristics and mixing of chrominance and luminance information [9]. YCbCr is an effective color space for the separation of color image pixels with uneven illumination and has proven to be effective for complex color images as well. The difference in the appearance of an un segmented image and a segmented image of a same object is shown in figure1.

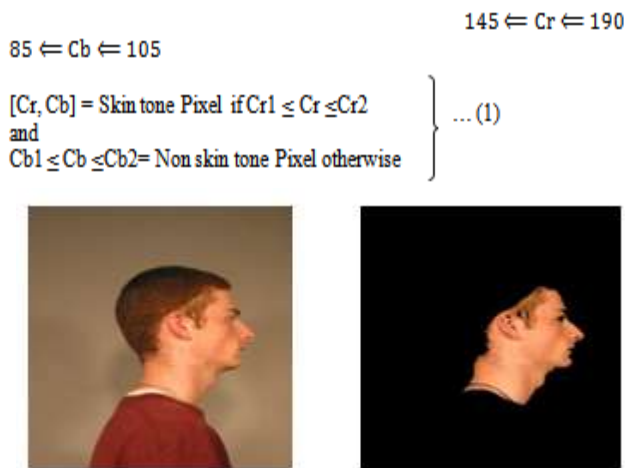


Fig.1. a) Input image b) Skin segmented image

There are different methods for measuring the similarity between two images. In contrast, Structural Similarity Index (SSIM) has shown superior performance and lower computational complexity than Human Visual System-based methods.

III. TEMPLATE CREATION

A template represents data and is very important in template-based approaches. Generally, ears can be classified into four types based on their shape: elliptical, circular, triangular, and rectangular. A template for extracting ear portions from side profiles of face images is generated through Principal Component Analysis (PCA) based on each of the four categories of ear images that are present in the database.

IV. EAR LOCALIZATION USING SSIM

The structural similarity index measure (SSIM) is a method for assessing the quality of digital images. SSIM is used for measuring the similarity between two images. The SSIM index is a full reference metric, which means the image quality measured by using an uncompressed or distortion-free image as a reference. Since SSIM is used to determine the similarity of structural features in this application, it is preferred for finding the ear structure within an input image. Localization based on template matching is used to locate the ear portion from a side profile image of a face. The template is moved on to each and every pixel in the image and the similarity index is measured at every point. A high similarity index at a particular point indicates the best match between the template and the region in the input image. An ear portion is extracted from its surroundings once the best match has been found. The process of structural similarity measure of the image is shown in figure 2 in the form of a block diagram.

Mean SSIM is computed as follows:

Luminance comparison is computed with

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \dots (2)$$

Contrast comparison is given by

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \dots (3)$$

Structure comparison is given by

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \dots (4)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \dots (5)$$

$$SSIM(x, \dots (6) \quad c(x,y)]^2 . [s(x,y)]^2$$

$$SSIM(x,y) = \frac{(z\mu_x\mu_y + C_1)(z\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \dots (7)$$

The distortion of images is sometimes related to localized image statistics and sometimes to space variations, but more importantly, the features of the image are highly spatial and non-stationary. To account for image distortion, mean SSIM (MSSIM) is used to locate the ear.

A normalized circular symmetric Gaussian weighting function $w = \{w_i | 1, 2, \dots, N\}$ of size 11×11 with a standard deviation of 1.5 is used for computing local statistics such as μ_x , σ_x and σ_{xy} . The estimates of local statistics are modified accordingly.

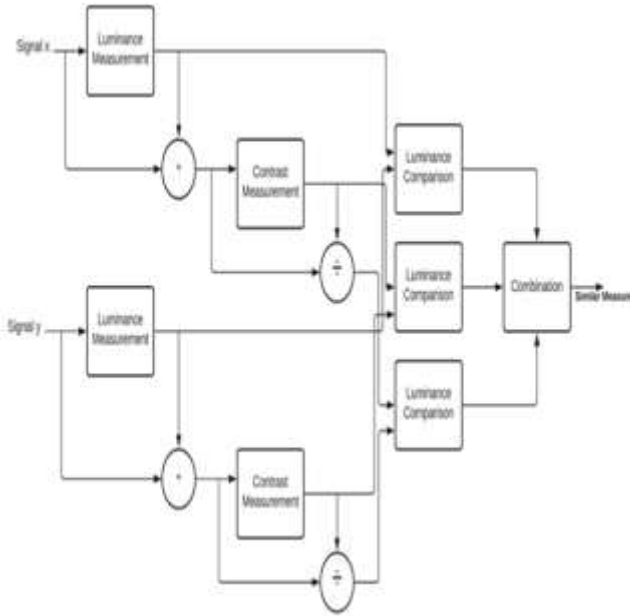


Fig.2. Block diagram showing the proposed algorithm

$$\mu_x = \sum_{i=1}^N w_i x_i \quad \dots (8)$$

$$\sigma_x = \left(\sum_{i=1}^N w_i (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad \dots (9)$$

$$\sigma_{xy} = \sum_{i=1}^N w_i (x_i - \mu_x) (y_i - \mu_y) \quad \dots (10)$$

This study considers K1 and K2 as the SSIM measure parameters: K1=0.01 and K2=0.03, respectively. Even though these values are arbitrary, their variations have no effect on the performance of SSIM since this algorithm is most insensitive to variations [10].

The overall image matching is computed by Mean SSIM (MSSIM) index as

$$MSSIM(X,Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j) \quad \dots (11)$$

Where X and Y in the above equation are template and input images, M represents the number of local windows and x_j and y_j represents the j-th local window image contents.

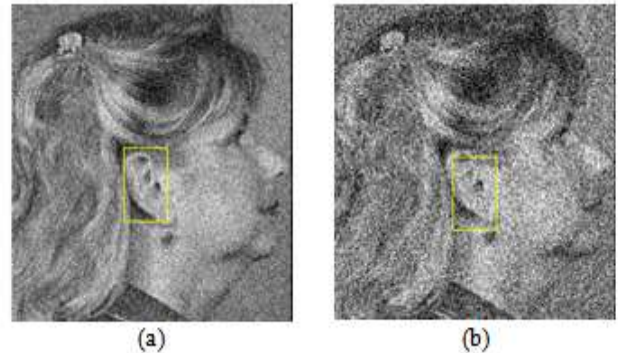
V. RESULTS AND CONCLUSION

The experiment was conducted on the UND -E ear database, which contains 464 side profile images of 114 subjects with a resolution of 1200×1600 . Skin segmentation is the first step in the process of ear detection to isolate the skin portion. Figure 3 shows the extracted portion of the ear from the given input side profile of the face.



Fig.3. Images with detected ear

Figure 4 presents the ear images detected from Gaussian noise with zero mean and with different variances. Increasing the variance observed reduction in the detected image quality.



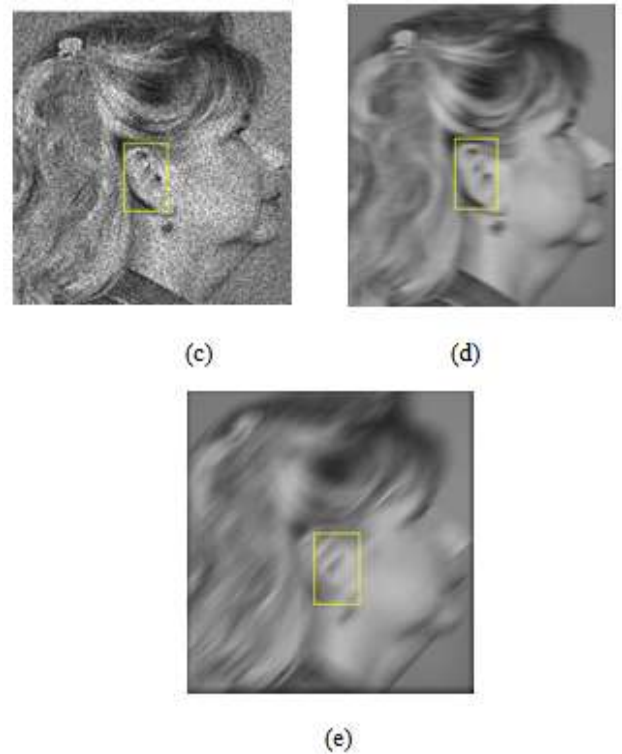
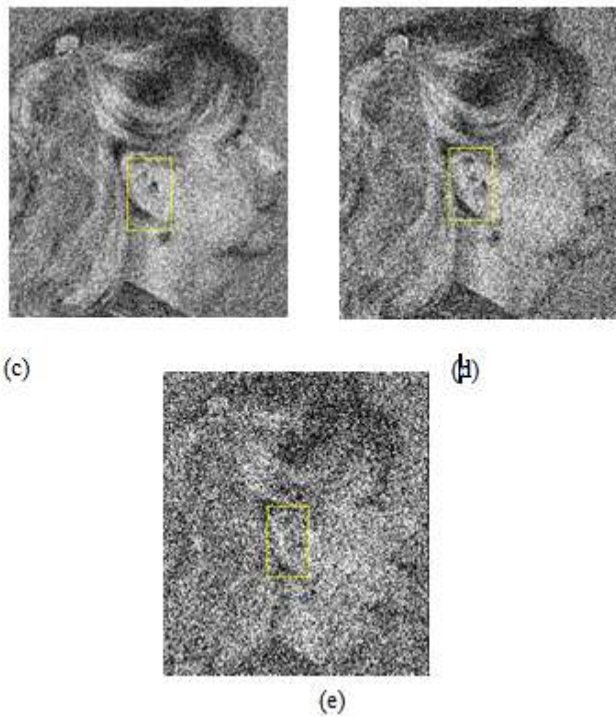


Fig.4. a to e. Detected ear images in presence of Gaussian noise with zero mean and variances of 0.01, 0.03, 0.05, 0.08, 0.2

Figure 5 presents the ear images detected from other noises such as Poisson's noise, Salt and Pepper noise and Speckle noises. Ear detected from motion blurred images are shown in figures d and e.

The search space for ear detection is reduced by more than 50% as a result of skin segmentation. In this method, the ear is detected using Mean SSIM, resulting in a detection accuracy of 97% even for the images with noisy environment.

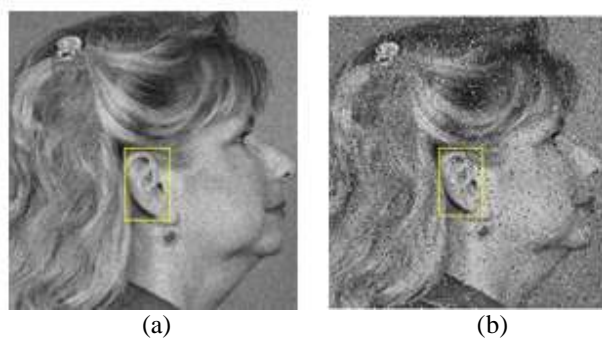


Fig.5 (a) Poisson's noise (b) Salt & pepper noise (c) Speckle noise (d, e) Different motion blurred images

VI. REFERENCES

- [1]. R. Purkait, "External ear: An analysis of its uniqueness," *Egyptian Journal of Forensic Sciences*, vol. 6, no. 2, pp. 99–107, 2016.
- [2]. S. Prakash, U. Jayaraman, and P. Gupta, "A skin-color and template based technique for automatic ear detection," *Proc. 7th Int. Conf. Adv. Pattern Recognition, ICAPR 2009*, no. February 2009, pp. 213–216, 2009.
- [3]. R. Deepak, A. V. Nayak, and K. Manikantan, "Ear detection using active contour model," *1st Int. Conf. Emerg. Trends Eng. Technol. Sci. ICETETS 2016 - Proc.*, 2016.
- [4]. B. Raman, S. Kumar, P. P. Roy, and D. Sen, "Ear Detection using Active Contour Model," *Adv. Intell. Syst. Comput.*, vol. 459 AISC, pp. V–VI, 2017.
- [5]. Ž. Emeršič, L. L. Gabriel, V. Štruc, and P. Peer, "Convolutional encoder-decoder networks for pixel-wise ear detection and segmentation," *IET Biometrics*, vol. 7, no. 3, pp. 175–184, 2018.
- [6]. I. I. Ganapathi, S. Prakash, I. R. Dave, and S. Bakshi, "Unconstrained ear detection using ensemble-based convolutional neural network model," *Concurr. Comput.*, vol. 32, no. 1, pp. 1–14, 2020.
- [7]. G. Raju, "An Empirical Study And Evaluation On Automatic Ear Detection," *International Journal of Computing* vol. 19, no. 4, pp. 575–582, 2020.



- [8]. R. Hsu and S. Member, "Face Detection in Color Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, 2002.
- [9]. K. Basha, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. Mary, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 41–48, 2015.
- [10]. Z. Wang, A. C. Bovik, H. R. Sheikh, S. Member, E. P. Simoncelli, and S. Member, "Image Quality Assessment: From Error Visibility to Structural Similarity," vol. 13, no. 4, pp. 1–14, 2004.