Edge Detection and Template Matching Approaches for Human Ear Detection

K. V. Joshi
G H Patel College of Engineering and Technology
vallabh vidyanagar, Gujarat, India

N. C. Chauhan
A D Patel Institute of Technology
New vallabh vidyanagar, Gujarat, India

ABSTRACT
Ear detection is a new class of relatively stable biometrics which is not affected by facial expressions, cosmetics, eye glasses and aging effects. Ear detection is the first step of an ear recognition system, to use ear biometrics for human identification. In this paper, we have presented two approaches to detect ear from 2D side face images. One is edge detection based method and the other is template matching method. For both the methods, the correctness of the detected ear is verified using support vector machine tool. For template matching method it is also verified by Euclidian distance. The purpose of the paper is also to compare the results of both the presented methods. The experimental results prove the effectiveness of these methods.

General Terms
Pattern recognition, pattern matching.

Keywords
Ear biometrics, ear detection, ear verification, edge detection, template matching.

1. INTRODUCTION
Ear is a viable new class of biometrics since ears have desirable properties such as universality, uniqueness and permanence. The ear has certain advantages over other biometrics. For example, ear is rich in features, it is a stable structure which does not change with the age. It does not change its shape with facial expressions. Furthermore, the ear is larger in size compared to fingerprints and can be easily captured although sometimes it can be hidden with hair and earrings. It has fixed background. For face recognition, when an image is a side face image, only the ear is unique feature from which a person can be identified. Although it has certain advantages over other biometrics, it has received little attention compared to other popular biometrics such as face, fingerprint etc. Human ear detection is the first task of a human ear recognition system and its performance significantly affects the overall quality of the system. Ear recognition is useful for person identification when an image of a side face is available.

The number of recent researches [1, 2, 3, 4] show that face recognition is possible and effective for side faces by detecting and recognizing components such as ears. Hence, in this paper we present and compare two methods of ear detection from 2D side face images. The rest of the paper is organized as follows. The proposed methods for ear detection and verification are described in section 2. The implementation and verification results are shown in section 3. The comparison of both the methods is shown the section 4 and Conclusion is discussed in the section 5.

2. EAR DETECTION METHODS

2.1 Edge Detection Based Method
A block diagram for this method is shown in the figure 1[5].

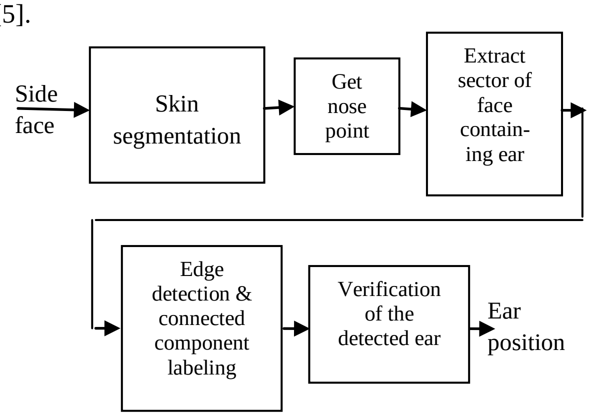

Fig 1 : Block diagram of the edge detection based method

As shown in block diagram of figure 1, first skin segmentation is performed from input image then nose tip is detected. After that sector of face containing ear is extracted. Then edge detection and connected component labeling is applied. The edges of ear have been found in extracted region. From the result of connected component labeling, where maximum connected edges are found a rectangle is drawn around it, that shows the detected ear.

2.1.1 Skin Segmentation

2.1.1.1 Color Space Selection
Here our input image is 2D side face color image. From side face image skin portion is separated out using the method suggested in [6, 7]. The goal is to remove the maximum number of non-face pixels from the images in order to narrow the focus to the remaining predominantly skin-colored regions. For this purpose we need to select appropriate color space from the wide variety of choices such as RGB, HSV, CMYK, YCbCr etc.

From these, RGB (red-green-blue), HSV (hue-saturation-value) and YCbCr are widely used [8]. In the RGB model, each of the three components may exhibit substantial variation under different lighting environments. The results of YCbCr and HSV are more robust to lighting variations because in both the color spaces, color classification is done using only pixel chrominance. It is expected that skin segmentation may become more robust to lighting variations if pixel luminance is discarded and it is also verified by results. Here HSV color space is preferred for color classification because of its similarities to the way human tends to perceive color. It decouples the chrominance information from the luminance information. Thus we can only focus on the hue and the saturation component.

2.1.1.2 Setting Threshold and Binary Image Creation

After choosing the suitable color space, the next step is to separate the skin colored region from the given input image. For this, the best technique is to apply threshold. To find the appropriate values for threshold, the many face images in HSV color space are examined and found out some specific ranges of the components for skin color.

When these threshold values are applied to the input image, the new binary image is formed in which the portions satisfying the conditions is made white and the remaining portion is made black. Figure 2 shows result of skin segmentation. This is a binary image created from a RGB input image. The thresholding is done on the basis of the HSV values.

2.1.2 Nose Tip Detection

Nose tip is detected by taking first white pixel from segmented skin image. The skin segmentation might not perfect and so any first non-black pixel in the binary image might not always represent a skin. So in order to obtain the nose point, the first non-black pixel is noted as a skin point only if it is surrounded by non-black pixels as well. In this way the first non black pixel in each column is noted.

(a) **(b)**

(a) Input Image
(b) Segmented skin
Fig 2 : Skin segmentation

Once a vector of the pixel locations is available, the minimum position is noted. Again, the surrounding pixels are examined in order to ensure that the identified point is a skin pixel. Figure 3 shows the identified nose point.

2.1.3 Extraction of Region of Face Containing Ear

Once the nose point is identified, a rectangular region is extracted from the face with dimensions 10cm * 20cm. The probability of finding the ear in this region is high. Once this region is extracted, the ear can be located by looking for strong edges in the region. Figure 4 shows the extracted region.

Fig 3 : Nose tip detection

Fig 4 : Region Extraction

2.1.4 Edge Detection and Connected Component Labeling

In this phase, edges are detected from input side face image. First of all color image is converted to gray scale image. For edge detection Sobel, Robert, Prewitt, Canny, Laplacian of Gaussian, Zero cross etc methods are available [9]. The Canny method finds edges by looking for local maxima of the gradient of intensity image. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges.

The experimental results using canny edge detector are shown in figure 5. The edges of ear have definitely been detected in extracted region which has higher probability to contain ear.

After edge detection, connected component labeling is applied, so extra unconnected small edges are removed. We have used an 8-connected neighborhood to label a pixel. The result of connected component labeling is shown in figure 6. From this result where maximum connected edges are found a rectangle is drawn around those edges. This rectangle shows detected ear. Figure 7 shows the final result of this method.

Fig 5 : Edge det*ection*

2.1.5 Ear Verification Using SVM Tool

A Support Vector Machine (SVM) [10, 11] is one such method that can perform pattern recognition; its use, though is not limited to this one application. While most classifiers work on designing rules that will place decision boundaries between data belonging to different classes, SVM goes a step ahead and designs "Support Vectors" such that the data belonging to different classes is now separated by a region rather than just a hyper plane. Thus, the distinction between classes is made more obvious, in an intuitive sense.

Fig 6 : Connected Component Labeling

Fig 7 : Final output

In this work, we validate the detected ear using SVM classification tool. The libsvm-2.91 [12] has been used in this verification task. Here we verify using SVM tool that weather our detection is correct or not. The choice of the appropriate Kernel parameter of SVM for a specific application is problem dependent and often a difficult task. Generally it is expected that if the data is known to be not linearly separable, a non-linear kernel performs better than the one based on a linear kernel. The model is created by features which are generated using detection results. The different kernel like Linear, Polynomial, Radial Basis Function, Sigmoid and accordingly parameters i.e. degree, gamma, cost co-efficient are adjusted manually to improve classification (verification) accuracy.

2.2 Template Matching Method

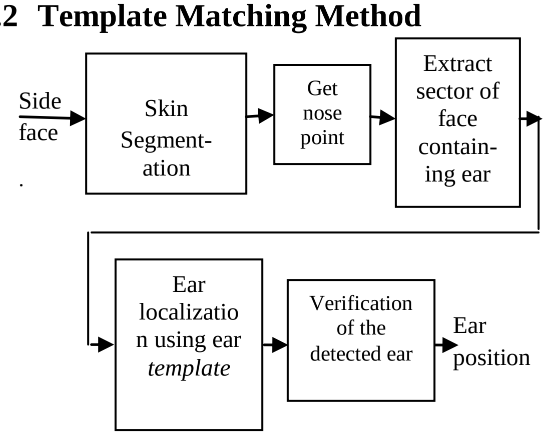

Fig 8: Block diagram of the template matching based method

A block diagram for this method is shown in figure 8.The template matching approach presented here is an enhancement of the approach presented in [4, 13]. Our presented ear detection technique involves three steps namely detection of probable area of ear, as described in the edge detection based method, ear localization and ear verification. In this method, instead of moving the template over the entire image we first detect the area having maximum probability of the ear. This step is added to the previously presented method [4], in order to reduce the time for ear detection. In our first step, skin segmentation is performed to eliminate all non-skin pixels from the image. Then nose point is identified and using distance estimation between nose tip and ear the probable area of ear is found. Second step employs an off-line created template to detect ears. Third step is about to verify the detected ears. In our presented approach, the detected ear is verified using a SVM machine learning tool in addition to Euclidean distance. Here the first step that is to find probable area of ear is described in the edge detection based method. The second and third steps are described as follows.

2.2.1 Ear Localization

The three steps involved in ear localization process [4] are discussed in following subsections.

2.2.1.1 Template Creation

For any template based approach, it is very much necessary to obtain a template which is a good representative of the data. In this technique, ear template is created by averaging the intensities of a set of ear images. Human ear shape can broadly be categorized into four classes: triangular, round, oval, and rectangular [4]. For creation of ear template, all types of the ear are considered to obtain a good representative template.

Intuitively, it seemed reasonable to us that the best template to use would be one derived by somehow averaging the some ear images of the dataset that would likely be in the testing images. We would like to find a good subset of the ears found in the training images that are clear, straight, and representative of typical lighting/environmental conditions. It is also important that these images be properly aligned and scaled with respect to one another. To this end, considerable time was spent for manually segmenting, selecting, and aligning ear images. At the end 17 ear images were chosen. These cropped images were first converted into grayscale and then the average was found which gives final template. The ear template T is formally defined as,

$$T(i,j) = \frac{1}{N}\sum_{k=1}^{N} Ek(i,j) \qquad (1)$$

where N is the number of ear images used for ear template creation and E_k is the Kth ear image. $E_k(i, j)$ and $T(i, j)$ represent the pixel values of the (i, j)th pixel of E_k and T respectively.

Thus, our final template for ear detection is a result of averaging together the 17 ear images. The actual template used in the matched filtering is of size 50×32 pixels. The template generated and used in the experimentation is shown in figure 9.

Fig 9 : Created template image

2.2.1.2 Resizing Template
It is observed that the size of the ear is proportional to the size of the side face image. This observation is used for resizing the ear template in the proposed technique. To handle the detection of ears of various sizes, ear template need to be resized to make it appropriate for the detection of ear in the image.

$$wi^{e} = \frac{wr^{e}}{wr^{f}}\, wi^{f} \tag{2}$$

By keeping the aspect ratio of the ear template same, it is resized to the width obtained in above Equation 2. Where wi^f and wr^f be the widths of the input face image and the reference face image respectively and wi^e and wr^e are the widths of input ear image and the reference ear image respectively.

2.2.1.3 Localization
To search an ear in the image *I*, ear template *T* is moved over the probable area of ear in the image and normalized cross correlation coefficient (NCC) [4] is computed at every pixel. NCC at point (*x, y*) is defined in Equation (3) as,

$$NCC(x,y) = \frac{\sum_{u,v}[I(u,v)-\bar{I}_{x,y}][T(u-x,v-y)-\bar{T}]}{\sqrt{\sum_{u,v}[I(u,v)-\bar{I}_{x,y}]\sum_{u,v}[T(u-x,v-y)-\bar{T}]}} \tag{3}$$

where sum is performed over *u,v* under the window containing T positioned at (*x,y*). $\bar{I}$ x,y and $\bar{T}$ are the average of brightness values of the portion of the target image under the template and template image respectively. Values of NCC lie between -1.0 and 1.0. Where it is found maximum, a rectangle is drawn around it to show detected ear. Value of NCC closer to 1 indicates a better match.

2.2.2 Ear verification
To determine whether a detected ear is a true ear or not two methods are used namely verification using Euclidian distance [4] and verification using SVM tool [10, 11, 12].

2.2.2.1 Ear Verification Using Euclidian Distance
Here to determine whether a detected ear is a true ear or not, shape based ear verification is performed. To measure the similarity, Euclidian distance between the two sets (one for template and another for detected ear) of mean is used, which is estimated as follows:

$$\text{Distance}= \sqrt{(|M^{T}| - |M^{E}|)^{2}} \tag{4}$$

where M^T and M^E are mean of ear template and detected ear respectively. Edge images of the ear template and the detected ear image are obtained using canny edge detector and the similarity distance between them is calculated using Equation (4). If the value of *distance* is less than a pre estimated threshold, detection is accepted otherwise it is rejected.

2.2.2.2 Ear Verification Using SVM Tool
Here libsvm-2.91 tool [12] is used for ear verification. The choice of the appropriate Kernel for a specific application is again problem dependent and often a difficult task. The different kernel like Linear, Polynomial, Radial Basis Function, Sigmoid and accordingly parameters i.e. degree, gamma, cost co-efficient are adjusted to improve the verification accuracy.

3. EXPERIMENTAL RESULTS

3.1 Data Acquisition
In this work CVL (CVL is library for image and data processing using graphics processing units (GPUs)) dataset [14] is used, which contains total 114 persons with 7 images of each. Resolution of each image is 640 x 480. All the Images are in JPEG format captured by Sony Digital Mavica under uniform illumination, and with projection screen in background. Age of most of the faces is between 18-25 years approximately. Although the method is tested on males of certain age groups, it can also be applied with persons of other age groups. An another dataset is produced by us, having images of 30 side faces with dynamic lighting condition with screen resolution of 2848 × 2144.

3.2 Ear Detection Results

3.2.1 Edge Detection Based Method
This method is tested on 130 side face images. In results of canny edge detector, there are some extra edges which are not of ear are also detected but the whole ear is detected with continuous edges. After edge detection threshold is applied for connected component labeling to remove unconnected small edges and a rectangular box is drawn around the place where maximum connected edges are found.

If the detected region contains part of ear, it is considered as a positive detection; otherwise it is a false detection. Figure 10 shows examples of positive detection. This method has failed in some cases, especially for the images where face is oriented with some angle because in that image nose tip detection is not correct. For some people the distance between nose tip and ear is differ than the generally estimated distance so there are chances for false detections or partial true detections. Figure 11 shows examples of false detection. In some figures the nose tip detection is wrong so there is false detection, while others are having partial ear detection because of variation of distance between nose tip and ear compared to estimated distance.

The accuracy is calculated as: (*genuine localization/total sample*) × 100. The accuracy of the presented method on the above mentioned database is obtained to be 83%. The average time to detect an ear from a side face image using this method is 1.35 seconds with Matlab environment. It should be noted that this time is much less than the traditional template based approach [4] in which a template of ears is moved over the entire image.

3.2.2 Template Matching Method
To create ear template, a set of ear images of 17 people is considered. NCC is used to localize the ear. Points having maximum NCC values are declared as the detected ear.

This experiment is performed on 100 images of CVL dataset and 30 images of general dataset. After experimentation, if the detected region contains part of ear, it is considered to be a positive detection; otherwise it is a false detection. Figure 12 shows examples of some of the positive detections. The proposed technique is also able to detect ear in presence of little occlusion due to hair. The fourth image of fourth row of figure 12 shows such example. Figure 13 shows examples of some of

Fig 10 : Positive detection

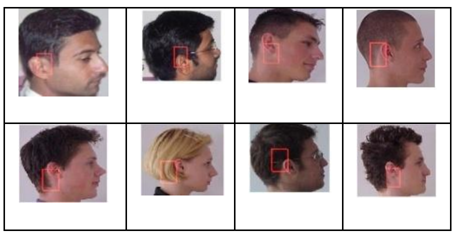

Fig 11 : False detection

the false detections. Localization method has failed in some cases, especially for the images which are of poor quality or heavily occluded due to hair (Fourth image of second row in figure 13) or face is oriented with some angle (Third image of first row and third image of second row in figure 13).

Accuracy of the localization is defined by (*genuine localization/total sample*) × 100. It is found to be 78% for the CVL dataset and 70% for general dataset. The average time to detect an ear from a side face image is approx. 3.42 seconds with Matlab environment.

3.3 Ear Verification Results

3.3.1 Edge Detection Based Method

In this work, for ear verification libsvm-2.91 tool is used. The rectangle drawn to show detection is of size 50×32. So, 1600 pixel intensity values are in one feature vector. The feature vectors are created from the pixels of the detected ear and model is created from these feature vectors. The different parameters of SVM are adjusted for accuracy measurement. The maximum accuracy is found 75.86% using polynomial, RBF and sigmoid kernel with certain parameter settings. Using polynomial kernel with degree 1 and cost 100 the found accuracy is 75.8621%. The same accuracy is found using RBF kernel with Gamma 0.0001 and cost 0.0008 and using sigmoid kernel with co efficient 1, Gamma 0.000625 and cost 1. Using Linear kernel the found accuracy is very less around 36.2069%.

3.3.2 Template Matching Method

Here also libsvm-2.91 tool is used for ear verification. The template is of size 50×32 and so the detected ear is also of the

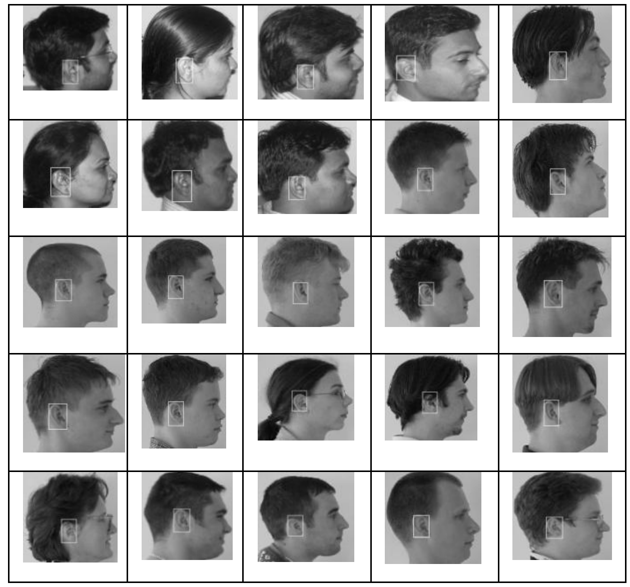

Figure 12 : Positive detection

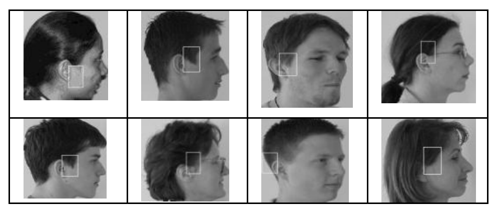

Figure 13 : False detection

same size. Therefore, 1600 pixel intensity values are in one feature vector. The different parameters are set for accuracy measurement. The maximum accuracy is found to be 93.33%. The maximum accuracy was obtained with RBF kernel using parameter values of C and gamma to be 60 and 0.000001 respectively. However, in our experimentation few other combinations were also found which gave the same accuracy.

4. COMPARISON OF METHODS

For CVL dataset, using edge detection based algorithm, the accuracy is around 83% and using template matching algorithm it is 78%. Some false positive detections are there. The comparison between the results for CVL dataset of these two methods is shown in Table 1.

Using edge detection based algorithm, satisfactory results are obtained for different environments. There are also a few false detections due to distance variation between nose tip and ear for different persons. In template matching algorithm, to detect ear in the general environment a new template has to be generated which after being used in the matching algorithm gives adequate results. The comparison between results of both the methods for general dataset is shown in Table 2.

Table 1. Comparison for the CVL Database

Name of Algori-thm	No of Images	Detected Ears	Accura-cy (%)	False Positive Rate	Detection Time In Seconds
Edge Detection Based Method	100	83	83	17	1.3
Template Matching Method	100	78	78	22	3.54

Table 2. Comparison for the general Database

Name of Algori-thm	No of Images	Detected Ears	Accura-cy (%)	False Positive Rate	Detection Time In Seconds
Edge Detection Based Method	30	24	80	20	1.4
Template Matching Method	30	21	70	30	3.3

5. CONCLUSION

From the results we can conclude that for edge detection based method the nose tip detection is very important because in this method ear detection is based on distance estimation between nose tip and ear. So if skin segmentation is not done properly then there are chances for wrong nose tip detection and hence causes false detection. For some people the distance between nose tip and ear differs than generally estimated distance, so there may be cases for partial ear detection.

Template matching algorithm does not have any effects on the detection of ears from different environments. The constraint for getting very good result is that the template has to be recreated for different datasets otherwise it degrades the performance of detection. Instead of moving template over the entire image, if it is moved over the region which has higher probability to contain ear then it takes less detection time. This method fails if ears are heavily occluded due to hair.

In our results, both methods fail, if side face is oriented with some angle.

6. REFERENCES

[1] B V Srinivasan **"**Ear Extraction From the Image of a Human Face". University of Maryland, College Park.

[2] H. Chen and B. Bhanu, "Hman Ear Detection From 3D Side Face Range Images". 3D Imaging for Safety and Security, vol.35, Springer-2007, pp.133-155.

[3] S. M. S. Islam, M. Bennamoun and R. Davies, "Fast and Fully Automatic Ear Detection Using Cascaded AdaBoost" Proc. of IEEE Workshop on Application of Computer, 2008.

[4] S. Prakash, U. Jayaraman and P. Gupta, "A Skin-Color and Template Based Technique For Automatic Ear Detection". Proc. Int'l Conf. Advances in Pattern Recognition, ICAPR' 09, Feruary 2009,pp. 213-216.

[5] K. Joshi, and N. Chauhan, "An Ear Detection and Support Vector Machine based Approach for Human Ear Detection and Varification", Int. Conf. Intellignet Systems and Data Processing (ICISD-2011), G. H. Patel College of Engineering and Technology, Gujarat, India, 24-25 January, 2011.

[6] SL Phung, A. Bouzerdoum , D.Chai, " Skin Segmentation Using Color Pixel Classification: Analysis and Comparison", IEEE trans. on Pattern Analysis and Machine Intelligence,vol.27,no.1, January 2005, pp.148-154.

[7] V. Vezhnevets, V. Sazonov, A. Andreeva: "A Survey on Pixel-Based Skin Color Detection Techniques" International Journal of Expert Systems With Applications, vol.3, April 2009, pp 4497-4507.

[8] J. D. Foley, A. Dam, S. K. Feiner, J. F .Hughes "Computer Graphics Principles and Practice", Pearson Education, 2006.

[9] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", Pearson Education, 2002.

[10] N. C. Chauhan, Y. K. Roy, Arun Kumar, A. Mittal, and M. V. Kartikeyan " SVM-PSO Based Modeling and Optimization of Microwave Components", Frequenz *62,2008.*

[11] J. A. Suykens, J. Vandewalle, and B. D. Moor, "Optimal Control by Least Squares Support Vector Machines", International Journal of Neural Networks,vol.14, no.1, Jan,2001, pp. 23-35.

[12] C. C. Chang and C. J. Lin, "LIBSVM – A Library for Support Vector Machines", 2001.

[13] K. Joshi, and N. Chauhan, "A Template Matching and Support Vector Machine based Approach for Human Ear Detection and Varification", Int. Conf. Information, Signals, and Communications (ICISC-2011), A. D. Patel Institute of Technology, Gujarat, India, 5-6 February, 2011.