Iris feature extraction using Gray Level Co-occurrence Matrix and Gabor Kernel filter its impact on iris Huffman compression image

R A Asmara 1 *, R D R Yusron 2 , F Rahutomo 1 , R Ariyanto 1 , D K P Aji 1 and P Choirina 2

1 Information Technology Department, State Polytechnic of Malang, Malang, Indonesia
2 Electrical Engineering Department, State Polytechnic of Malang, Malang, Indonesia

*rosa.andrie@polinema.ac.id

Abstract. Algorithms developed to identify people with iris image data have been tested in many field and laboratory experiment. This paper analysis some a parameters of iris image used to recognize human. Iris recognition system, which is applied based on segmentation, normalization, encoding, and matching is also describe in this paper. Circle Hough Transform segmentation module used to find the inner and outer boundaries of the iris. The experiment was carried out using CASIA v1 iris database with grayscale images. Shape, intensity, and location information for localizing the pupil or iris and normalizing the iris area a used iris segmentation by unwrapping circular area into a rectangular area. Normalized area will be used to extract the features using Gray Level Co-occurrence Matrix (GLCM) and Gabor filter, the feature compared the recognition accuracy using Support Vector Machines (SVM) and Naive Bayes classifiers. GLCM feature test results achieved 95.24% SVM classification accuracy, whereas using achieved 85.71% Naive Bayes. Gabor feature test results achieved 95.24% SVM classification accuracy, whereas using achieved 95.23% Naive Bayes. The classification process based on GLCM and Gabor features show that the SVM method have to highest recognition accuracy compare to Naive Bayes classifier.

1. Introduction
Biometrics is used to recognize a person's physical characteristics or behavior. There are various types of biometrics used such as fingerprints, retina, face recognition, sound pattern analysis etc. Among all biometrics is one of the best biometrics. Iris biometrics is for personal identification before all characteristics must be verified using various techniques. Biometrics also provides a safe authentication and identification method, because it is difficult to be forged. Iris recognition the propose of system is composed are segmentation, normalization, compression, encryption and matching, is explained briefly in the following sections.

The current research is focused on the iris feature extraction stage. The iris feature extraction was carried out with GLCM and Gabor filter kernel. This is the most important step of the recognition process. During this step, the iris feature is extracted from the input image and converted to code. Various techniques have been applied for this purpose over the past few years. John Daugman is a pioneer of the era of iris recognition [1]. He has used a 2D Gabor filter to extract iris features. After
extracting a unique feature, it converts feature vectors to binary forms and produces a 2048-bit code. Then he uses hamming distance as a classifier to compare binary feature vectors. In another experiment, Li Ma used a different Gabor filter called a circular symmetrical filter to extract iris features [2]. In this work, it uses circular symmetrical filters to extract local iris features and combine this local information to form global feature vectors. Some other researchers use wavelet transformations for feature extraction.

Jun Zhou, and team, proposed a technique for iris recognition [3]. In this work Discrete Wavelet Transform (DWT) is used for feature extraction. Where the Haar 2D wavelet is used as DWT. The work was done on the CASIA iris database which obtained the correct level of recognition.

Charles O. Ukpai, and team, proposes a new approach by using complex wavelet transformations that are rotated principally for iris recognition [4]. This approach involves PCA and Dual Tree Complex Wavelet Transform (DT-CWT) for feature extraction and SVM for classification. The CASIA database is used to make a recognize.

### 2. Method

The proposed work is modelled as shown in figure 1. As shown in the block diagram both training and testing phases will have the same procedure until feature extraction and will be compressed using Huffman. In the training phase the features are stored in the knowledge base where as features extracted in the testing phase will be compared with the features in the training phase with the help of the classifier.

![Image](image.png)

**Figure 1.** Design proposed work iris recognize.

The CASIA-IrisV1 database is used to get iris images. Images of iris in this database iris images are obtained in different ways so as to provide a versatile iris image that helps to train the system with all possible variations during taking iris. In total there are 756 images of 108 identities. Each identity has 7 images consisting of 4 training images and 3 test images.

The input iris image may not be as expected which means clarity or dimensions or other factors may not be as desired. Therefore, to achieve the desired input image property must be changed in the desired manner. Pre-processing is the step where image changes are made. This pre-processing step includes resizing an image, where the size of an inappropriate image is changed to get a fixed image size. Then the image is improved to improve image quality. Using a Canny edge detector in the image is detected [5].

**Edge Detection** The initial step of segmentation is identifying traits in an image that are related to estimating the structure and properties of objects. One characteristic of an image is the edge [6]. Edge is a characteristic that is important for image analysis and usually edges occur at the boundary between two different regions in an image. The Canny operator is a combination of the Gaussian smoothing filter and the gradient approach.

**Circle Hough transform** is used to detect parameters of simple geometric objects such as lines, circles. Here in this work Hough transform is used to detect circular objects, namely iris and pupils that are almost circular in shape. Here the edges are generated then they are thrown into the Hough to find the best fitting circle for the object. The effect of the presence of eyelashes and eyelids on the area of the iris circle is sought using linear Hough transformation. The Hough linear transformation algorithm is basically the same as the Circle Hough Transform algorithm with a difference in the equation [7],
that is, \( r = x \cos \theta + y \sin \theta \). Using these irises and pupil areas is identified and they are separated. This is shown in figure 2.

**Figure 2.** Normalized and segmented iris image.

Iris Normalization After the iris is localized the next step is normalization (iris enrolment). Using the equation, the iris data are extracted. Different circles with increasing radius and angle are drawn starting from the pupil center till it reaches near the iris coordinates. The information is extracted. \( x = c(x) - r \sin(\theta), y = c(y) + r \cos(\theta) \) where \( c(x,y) \) denotes center coordinates, \( (x,y) \) denotes coordinates of image, \( \theta \) is the angle and \( r \) denote the radius [8].

Feature Extraction after preprocessing, feature extraction will be performed using 2 algorithms, that is Gray Co-Occurrence Matrix and Gabor Kernel Filter.

Gray Level Co-Occurrence Matrix is a matrix distribution defined in the image to be a distribution of values at the given offset. At first GLCM was built based on the direction and distance between the pixels of the image. Then the feature is extracted from the matrix as a texture representation. From the GLCM various features can be extracted such as Homogeneity, Entropy, Contras, Energy, Correlation discussed below [9].

Homogeneity measure the proximity of the distribution of elements in the GLCM to the diagonal of the GLCM and also measure the degree of variation (1).

\[
H = \text{sum}(\text{sum}(p(m,n)/(1+[m-n])))
\] (1)

Entropy measuring the uncertainty of image texture, when the matrix for all values is equivalent, it reaches the minimum value (2).

\[
S = -\sum_{m} \sum_{n} p(m,n) \log p(m,n)
\] (2)

Contras measuring changes in brightness, darkness of the iris image given (3).

\[
I = \sum\sum (m-n)^2 p(m,n)
\] (3)

Energy this is a Gray scale image texture calculation of varying homogeneity, which reflects the distribution of Gray scale uniformity of weight and texture (4).

\[
S = -\sum_{m} \sum_{n} p(m,n) \log p(m,n)
\] (4)

Correlation the combined probability event of the pixel pair specified in the iris image (5).

\[
C = \text{sum}(\text{sum}((m-\mu_m)(n-\mu_n)p(m,n)/\sigma_m\sigma_n))
\] (5)

Gabor Kernel Filter act very similar to mammalian visual cortical cells so they extract features from different orientations and scales. The Impulse Response Function (IRF) is a sinusoidal wave with a Gaussian envelope [2], with the orientation of the Gabor Filter as follows (6-8)

\[
g(x,y,f_0,\theta) = \left\{- \frac{1}{2} \left[ \frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2} \right] \right\}
\] (6)
\[ x' = x \cos \theta + y \sin \theta \]  
\[ y' = y \cos \theta + x \sin \theta \]  

where \( \delta x \) and \( \delta y \) represent the spatial size of the filter, \( \theta \) is the orientation angle, \( f \) is the frequency of the filter [2]. Since a gabor filter has both its real and imaginary parts, when an input image is convolved with a bank of 32 filters, it yields 32 real and 32 imaginary outputs in Figure 3. An output feature vector of length 2048 bits is generated by combining these real and imaginary parts. This convolution process is applied to all images in the database.

The compression used is Huffman. The Huffman method is one of the compression techniques by coding in the form of bits to represent character data. How this method works or algorithm? First, first calculates the number of types of characters and the number of each character contained in a file. Next, Arrange each type of character with the least number of character types to the most number. Make a binary tree based on the sequence of characters from the smallest number to the largest, and give a code for each character. Change existing data with bit codes based on binary trees. Save the number of bits for the largest bit code, the type of character sorted from the largest exit frequency to the smallest along with the data that has been changed to the bit code as the compression data.

After the feature is extracted, it will be stored in the database. Then these features are used to train the system. In the testing phase the feature is extracted in the same way as in the training phase. Then this feature must be identified, this is done using classifiers. Classifiers are used to classify features based on the values of the class. In this work two classifiers are used in the State Vector Machine method and Naive Bayes is used as a classifier.

State Vector Machine (SVM) is a supervised learning model with a related learning algorithm that analyses the data used for classification and regression analysis. Given a series of training examples, each marked as belonging to one or the other of two categories, the SVM training algorithm builds a model that gives new examples to one category or another. Then he built an airplane or a series of airplanes in high or unlimited dimensional space, which could be used for classification. Classification is done by considering the optimal hyperplane which separates different data sets based on their features [10]. Then based on the iris classification it is identified by considering which group the feature falls.

Naive Bayes (NB) in this study will use the NB algorithm. NB is a popular machine learning tool for classification, because of its simplicity, high computational efficiency, and good classification accuracy, especially for high data classification [11].
3. Results and discussion

![Image of iris with annotations](image)

**Figure 4.** Result normalized iris image.

The image input (a) will be processed for the edge detection which is circular using Circle Hough Transform (b), then the iris will be segmented and normalized (c). Image (c) will be compressed with Huffman and take its features. Matching the Feature extraction is assessed against the Classifier with the results correctly identified or incorrectly processed. The amount of data used for the trial is 21, 56, 98 ID. The results of the recognition performance of the normalized iris comparison (c) with the GLCM and Gabor methods can be seen in figure 5.

![Graph showing recognition results](graph)

**Figure 5.** Result normalized and compressed iris recognize.

4. Conclusion

This work gives a clear understanding of the various types of algorithms used for pre-processing, feature extraction, and classification. It also gives an idea to compare algorithms for feature extraction. It is clear that the number of algorithms used increases accuracy also increases with high complexity. The results obtained in Figure 5 can be GLCM feature test results achieved 95.24% SVM classification accuracy, whereas using achieved 85.71% Naive Bayes. Gabor Filter feature test results achieved 95.24% SVM classification accuracy, whereas using achieved 95.23% Naive Bayes. The classification process based on GLCM and Gabor features show that the SVM method have to highest recognition accuracy compare to Naive Bayes classifier.
Acknowledgment
The authors would like to thank the State Polytechnic of Malang for supporting to participate in this conference.

References
[1] Daugman J 2004 How Iris Recognition Works IEEE Trans. Circuits Syst. Video Technol. 14(1) 21–30
[2] Minhas S and Javed M Y 2009 Iris feature extraction using gabor filter 252–255
[3] Zhou J, Luo T, Min, Guo S and Qing T 2010 Using 2D Haar Wavelet Transform for Iris Feature Extraction 2010 Asia-Pacific Conference on Information Theory
[4] Radu P, Sirlantzis K, Howells W G J, Hoque S and Deravi F 2013 Optimizing 2D Gabor Filters for Iris Recognition", in Emerging Security Technologies (EST), 2013 Fourth International Conference on, IEEE 2013
[5] Dharanesh C M, Prasad R and Patil C M 2017 Feature Extraction Classification for Personal Identification using Iris 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) 431-435
[6] Yang K, Du Y, Zhou Z and Beleher C 2010 Gabor descriptor based cancelable iris recognition method 2010 IEEE International Conference on Image Processing 4085-4088
[7] Kak N, Gupta R and Mahajan S 2010 Iris Recognition System Int. J. Adv. Comput. Sci. Appl. 1(1)
[8] Hariprasath S and Mohan V 2010 Biometric personal identification based on iris pattern recognition using Wavelet Packet Transform 1–5
[9] Vanthana P S and Muthukumar A 2015 Iris authentication using gray level co-occurrence matrix and hausdorff dimension 2015 International Conference on Computer Communication and Informatics (ICCCI) 1-5
[10] Asmara R A, Rahutomo F, Hasanah Q and Rahmad C 2017 Chicken meat freshness identification using the histogram color feature 2017 International Conference on Sustainable Information Engineering and Technology (SIET) 57-61
[11] Pratiwi I Y R, Asmara R A and Rahutomo F 2017 Study of hoax news detection using naïve bayes classifier in Indonesian language 2017 11th International Conference on Information & Communication Technology and System (ICTS) 73-78