Enhancement of Illumination Scheme for Adult Image Recognition

Sasan Karamizadeh
Iran Telecommunication Research Center (ITRC) Information and Communications Technology Research Institute, Tehran, Iran s.karamizadeh@itrc.ac.ir

Abouzar Araborskh* Faculty member of Iran Telecommunication Research Center (ITRC) Information and Communications Technology Research Institute, Tehran, Iran abouzar_arab@itrc.ac.ir

Received: May 10, 2017 - Accepted: August 29, 2017

Abstract—Biometric-based techniques have emerged as the most promising option for individual recognition. This task is still a challenge for computer vision systems. Several approaches to adult image recognition, which include the deep neural network and traditional classifier, have been proposed. Different image condition factors such as expressions, occlusion, poses, and illuminations affect the facial recognition system. A reasonable amount of illumination variations between the gallery and probe images need to be taken into account in adult image recognition algorithms. In the context of adult image verification, illumination variation plays a vital role and this factor will most likely result in misclassification. Different architectures and different parameters have been tested in order to improve the classification’s accuracy. This proposed method contains four steps, which begin with Fuzzy Deep Neural Network Segmentation. This step is employed in order to segment an image based on illumination intensity. Histogram Truncation and Stretching is utilized in the second step for improving histogram distribution in the segmented area. The third step is Contrast Limited Adaptive Histogram Equalization (CLAHE). This step is used to enhance the contrast of the segmented area. Finally, DCT-II is applied and low-frequency coefficients are selected in a zigzag pattern for illumination normalization. In the proposed method, AlexNet architecture is used, which consists of 5 convolutional layers, max-pooling layers, and fully connected layers. The image is passed through a stack of convolutional layers after fuzzy neural representation, where we used filter $8 \times 8$. The convolutional stride is fixed to 1 pixel. After every convolution, there is a subsampling layer, which consists of a $2 \times 2$ kernel to do max pooling. This can help to reduce the training time and compute complexity of the network. The proposed scheme will be analyzed and its performance in accuracy and effectiveness will be evaluated. In this research, we have used 80,400 images, which are imported from two datasets - the Compaq and Poesia datasets - and used images found on the Internet.

Keywords—adult image; illumination; fuzzy deep neural network segmentation; Histogram truncation and stretching; DCTII; AlexNet; Convolutional

I. INTRODUCTION

Nowadays information on the Internet is increasing. Children and teenagers have to be prevented from gaining access to adult information like adult images[1]. However, several techniques have been utilized for the recognition and detection of adult images in numerous fields, which are statistics, security, and military [2]. The illumination variable is assumed to be one of the crucial issues in facial recognition [4]. Lighting conditions differ indoors and outdoors, so therefore one of the important goals for facial recognition would be the creation of incentives for facial recognition related to understanding and improving computer vision vis-à-vis illumination changes [5]. Reasonable amounts of illumination variation between the gallery and probe images need to be taken into account in adult image recognition.
algorithms [6]. Therefore, the research area’s range will be limited, which will allow us to conclude that people with different illumination can appear to have similar features [7]. In the context of adult image verification, the factor of illumination variation is very important [8]. This factor will most likely result in misclassification during adult image recognition [9].

The illumination normalization scheme introduced is employed to minimize illumination variation in adult image datasets and decrease misclassifications due to the similarities in the facial features of adult images with illumination [10].

This paper consists of an introduction and then other sections as follows: related works are described in Section 2, the research methodology detailed in Section 3, the experiment itself explained in Section 4, and the conclusion in Section 5.

II. RELATED WORK

[11,12] used an SVM algorithm as a classifier. SVM is a binary classifier. To do a multi-class classification, pairwise classifications can be used (one class against all others, for all classes) but they are computationally expensive, thus they run slowly. They also used features or the dual-tree complex wavelet (DTCWT)-FFT features from the enhanced face images, but the high-computational complexity and high memory capacity requirement are important disadvantages.[13] Anisotropic diffusion illumination normalization technique and DCT were used for face recognition. The Anisotropic diffusion was employed as a preprocessor before applying the DCT as a feature extractor [14]. Performance metrics were generated and evaluated by verification and identification rates using nearest neighbor classifier. In this research, preprocessing before applying DCT II is done by employing Anisotropic diffusion, which has a linear function [15]. Although this technique is suitable for noise reduction while keeping object boundaries, linear functions are not suitable in complex situations. Another drawback in this research is applying illumination normalization over the entire image, which is less accurate compared to segmented illumination normalization techniques. [16] utilized an LDA algorithm to develop a pipeline for dealing with variable illumination. He minimized the effects of illumination variations on face-recognition performance. Additionally, [17] proposed an SOM (self-organizing map) to tackle illumination variations problems. An SOM is used for feature extraction and dimensionality reduction and histogram equalization is used for resolving the issue of illumination variation in an image. However, both studies were unable to address the challenges of applying multiple reference subspaces to different poses. The algorithm is used by a fuzzy cluster for segmenting the candidate region according to common colors of breast feature. A forward propagation neural network has been used for judging whether the candidate region is a breast feature or not [18]. A mixture of CNN has been utilized for adult image recognition. The proposed method has been developed as a linear regression problem wherever the weight has been calculated utilizing a normal least square [19].

III. RESEARCH METHODOLOGY

The research framework is represented in Figure 1. This framework involves two parts, the first of which is pre-processing and the second of which is auto contrast balancing and then using Histogram Equalization to balance the images’ contrast. After auto contrast balancing, we need to reduce the noise on the input image because noises such as salt-and-pepper and speckle noises can affect the feature extraction. This can be done using a Wiener filter, and Adult image processing includes the next step, which is illumination normalization. For the illumination variation of the Fuzzy Deep Neural Network, the Segmentation technique is applied to remove illumination variation; which keeps the main facial feature unimpaired. The next step in the proposed scheme is the face detection algorithm, which is employed to detect multiple faces in an image frame. It extracts the Haar-like features in the training images and applies these features to a Cascade Classifier. In order to extract facial features, we utilized a Singular Value Decomposition (SVD) feature extraction technique, and the extracted features are then utilized to train an HMM classifier.

![Fig 1. Research Framework](image_url)

A. Auto Contrast Balancing

The Histogram Equalization method is utilized to balance color channels in this step. The whole
histogram has been distributed with a range of pixel intensity values that have been balanced. Auto Contrast Balancing stretches the color distribution equally on RGB channels. As an outcome of this step, we have an image that has a broader range of color.

B. Noise Reduction

We discussed the method adopted for reducing ambient noise, especially noises created as an effect of Auto Contrast Balance, in the previous section as part of the second step. We have applied Wiener Filter techniques to investigate their respective potentials in being applied to adult image recognition systems.

C. Illumination Normalization

Illumination normalization is the first step of carrying out adult image recognition. Illumination variation in an adult image recognition system may cause reduced performance. To solve this performance issue, we have proposed utilizing illumination normalization based Fuzzy Deep Neural Network Segmentation technique. The proposed framework, which is shown in Figure 2, includes four phases; starting with Fuzzy Deep Neural Network Segmentation, which is utilized to segment an image based on illumination normalization. For improving histogram distribution in the segmented area during the second phase, we have used Histogram Truncation. Stretching Contrast Limited Adaptive Histogram Equalization (CLAHE) is utilized in the third phase to enhance the contrast of the segmented area, and in the last phase, DCT-II is used for illumination normalization and a zigzag pattern is utilized for extracting low-frequency coefficients.

![Fig 2. Illumination Normalization Phases](image)

- **Fuzzy Deep Neural Network Segmentation**
  A Fuzzy Fully Convolutional Network includes five parts. In a nutshell, the input image (purple) is processed in the Fuzzy Membership Function Layer (black) and enters the Fuzzy Neural Representation Layer (green), which prepares it to go through the deep neural network. The FCN network consists of a Convolutional Layer (red) and a Pooling Layer (maroon). Finally, classification is done in the Prediction Layer (gray). All these steps are shown in Figure 3.

- **Fuzzy Membership Function Layers**
  In this layer, each node has been connected with a multiple membership function, which assigns labels to each input variable. The fuzzy membership function is calculated by the degree that an input node belongs to a certain fuzzy set. We have utilized Gaussian membership with mean $\mu$ and variance $\sigma^2$ for fuzzy rule layer performance.

- **Fuzzy Neural Representation**
  This part exploits the neural learning concept to transform the input into some high-level representations.

  AlexNet architecture is used, which is made up of 5 convolutional layers, max-pooling layers, and fully connected layers. We have explained each of them:

  - **Convolutional Layer**
    The image is passed through a stack of convolutional layers after fuzzy neural representation, where we utilized filter $8 \times 8$. The convolutional stride is fixed to 1 pixel. This layer is applied to the image like a filter that extracts features from low level to high level. Different activation functions in different layers are applied to extract different features. The Convolutional Layer is the core building block of a CNN.

  - **Pooling layer**
    A Pooling Layer is typically used by Deep Neural Networks to ensure generalization of a model and to avoid overfitting. A Pooling Layer is a form of non-linear down-sampling. In this part, we have utilized spatial pooling, which is carried out by five max-pooling layers, which follow some of the convolutional layers. Max-pooling is performed over a 2x2 pixel window, with stride 2.
- **Prediction Layer**
  The Prediction Layer is the last layer that assigns the fused representation into its corresponding category. Soft-max function has been utilized to classify the data points into different classes.

  We have utilized fuzzy learning because of three important points which are: first, the greatest way of reducing the uncertainties of input data is provided by fuzzy learning; the second is that fuzzy learning obviously leads to soft logistic values (fuzzy representations) in the range of (0, 1) and moreover, each dimension of the neural outputs is also in the range of (0, 1); and the last reason why we used fuzzy learning is that the fuzzy learning part allows task-driven parameter learning in order to achieve the highest accuracy. The FCN-AlexNet was pre-trained over a SYNTHIA dataset that contains over 200,000 HD images. Afterwards, transfer learning is done on illumination dense segmentation by data augmentation and a MSRA weight initialization scheme.

- **Histogram Truncation and Stretching**
  The next step for illumination normalization is histogram truncation and stretching, which is a histogram remapping technique. It is used for redistributing the intensities on the scale of values available as well as possible.

- **Contrast Limited Adaptive Histogram Equalization**
  The third step is Contrast Limited Adaptive Histogram Equalization. This technique is utilized widely in image processing as an improved version of histogram equalization.

- **DCT-II Method**
  DCT-II normalization is utilized to remove the illumination normalization’s effects while keeping the main facial features unimpaired. The main idea of the proposed approach is that illumination normalization can be expressively compacted by truncating low-frequency Discrete Cosine Transform (DCT-II) coefficients in the logarithm DCT-II domain. Figure 4 shows the illumination normalization pseudo code.

  ![FCNN Illumination Normalization](image)

  **Fig 4.** Shows the illumination normalization pseudo code.

  Figure 4 shows the illumination normalization used in this algorithm, utilizing the FFCN method. First, fuzzy degrees of nodes are calculated. Secondly, fuzzy neural features are extracted by the FCN model. Data points are classified by soft-max function and the image is illumination segmented. Afterward, HTS is applied over the segmented area. The next step is applying Contrast Limited Adaptive over the segmented area. Finally, Discrete Cosine Transformation (DCT-II) is applied over the segmented area, followed by the zigzag method of coefficient traversing. This function sets the DCT-II low-frequency coefficients to zero in order to increase illumination invariance. The flowchart of the DCT Illumination invariant is illustrated in Figure 5 as follows.

  ![Illumination normalization flowchart](image)

  **Fig 5.** Illumination normalization flowchart

  Figure 5 illustrates the required steps for performing FFCN in order to normalize illumination. The first step is illumination segmentation, which is done by the fuzzy neural network algorithm. First of all, the degree to which an input node belongs to a certain fuzzy set is calculated, and is followed by extracting fuzzy neural features. To finalize the illumination segmentation, classification of data points is done by soft-max function. After that, HTS is applied over the segmented area, followed by applying Contrast Limited Adaptive
over the segmented area. The final step is applying DCT-II over the segmented area and processing the zigzag function of DCT-II low-frequency coefficients to zero.

D. Face Detection

Another important step in adult image recognition is face detection. Occasionally an image without a head has a lot of naked parts, but it’s not an adult image. Therefore, we review an adult image by utilizing the naked part; firstly, we need to remove the face part and because of this, we need to detect the face. We have utilized the AdaBoost algorithm for face detection which is a more efficient algorithm for face detection. The key point is to train different weak classifiers with the same training set. Then, these weak classifiers can be combined together to constitute a final strong classifier. The AdaBoost algorithm has a good effect, but it can be only applied to frontal face detection.

E. Feature Extraction

Singular Value Decomposition is an excellent unsupervised learning tool to process source data in order to find clusters, reduce dimensionality, and find latent variables. This means that field or experiment data can be converted into a form which is free from redundancy and can be better organized to reveal hidden information about the algorithm extraction’s feature.

F. Adult Image Recognition

The Hidden Markov Model (HMM), which is a discrete-time hidden model, states that the observation shall not be in any way explicit[13][13][13]. It has rehashed the association of probability distribution function, as well as modeled the probability of emitting symbols’ availability in each state. HMM is a very well-known application, especially when the data is in one dimension, such as speech recognition. However, the computation of two-dimensional data would become complex whenever the extension is fully connected to two-dimensional HMM data. HMM could be defined in a manner similar to the entities as described below:

- \( S = \{ S_1, S_2, \cdots, S_N \} \) a finite set of hidden states,
- The transition matrix \( A = \{ a_{ij} \in [0, 1] \leq i \leq j \leq N \} \) representing the probability of going from state \( S_i \) to state \( S_j \), \( a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \) \( 1 \leq i, j \leq N \) with \( a_{ij} \geq 0 \) e \( P N = 1 \), \( a_{ij} = 1 \); with \( a_{ij} = 0 \) e \( P N = 1 
\); the emission parameters \( B = \{ b(o|S_j) \} \), indicate the probability of emission of the symbol \( o \) when the system state is \( S_j \). In short, HMM is denoted as a triplet \( \lambda = (A, B, \pi) \). In order to determine the parameters \( (A, B, \pi) \), given that the set of sequences is \( \{ O_i \} \), the Baum Welch re-estimation standard will be used to maximize the probability of this training set of sequences model \( P(\{ O_i \} | \lambda) \). In this paper, the training procedure would be immediately stopped upon convergence. All of the evaluation steps involve the computation of the probability \( P(O|\lambda) \), with \( \lambda \) as a given model, whereas \( O \) is the new observation sequence, and is carried out through the forward-backward procedure.

IV. EXPERIMENT

This paper has utilized 80,400 images including adult images and non-adult images from two datasets, which are the Compaq and Poesia datasets, and images found on the Internet. 70 percent of these images are used for training and the rest are utilized for testing. We are focused on the problem of illumination normalization in this paper.

A. Comparing F-measures

As pointed out previously, the algorithm uses an improved illumination normalization to increase the F-measure rate. The algorithm’s performance is evaluated with and without illumination normalization and the results are shown in Table 1.

| Table 1: Performance is evaluated with and without illumination normalization |
|-----------------------------|-----------------------------|
| Percent | With illumination normalization segmentation | without illumination normalization segmentation |
| 20% | 0.41 | 0.4 |
| 30% | 0.56 | 0.34 |
| 40% | 0.67 | 0.66 |
| 50% | 0.88 | 0.78 |
| 60% | 0.93 | 0.85 |
| 70% | 0.96 | 0.87 |
| 80% | 0.93 | 0.82 |
| 90% | 0.92 | 0.81 |

As is shown in Table 1, the best training mark was achieved at 70%, which shows the highest F-measure rate reached the best possible result with illumination normalization and misclassification at a minimum rate of 70%. Among all the training values, the classifier experiment shows the worst result was at a 20% training rate. At 20 percent dataset training, the result was not conclusive. Based on our experiments, training datasets should be 30% or higher to get more conclusive and reliable results. If we have more training, testing data is not sufficient or not reliable at a 20% or lower percentage of dataset training. In different tests, the variance of results is very high due to insufficient training datasets, so the results at 20% and lower are not reliable. The data for testing is not enough for testing. Due to inadequate training data, the latter shows a wide fluctuation. After conducting the experiment 20 times, the results prove that a 70% training rate consistently achieved the best results.

B. Analysis of Noise Reduction Experiments

Figure 6 shows the Wiener noise reduction filter shows 96.9 percent, which is the best result compared to other noise reduction filters. The Wiener filter achieved the best results because of its nature which means it tailors itself to local image variance. It performs the best smoothing when local variance is small, which is applicable to our scenario. Although all three applied noise reduction filters show significant improvement to the F-measure values, the achieved F-measure values with a Wiener filter was 0.969, which...
is the highest between the tested filters. Therefore, the accuracy of without noise reduction is 0.893, and tests with a Wiener filter reveals a 7.9 percent improvement to the F-measure.

![Graph showing F-measure values for different filters](image)

**Fig 6: Result with and without Noise Reduction**

### C. Comparative Study

Table 3 shows the result of our experiment step by step, before and after applying the filters to the dataset.

<table>
<thead>
<tr>
<th>Number of experiments</th>
<th>Phase 1 (Pre-processing)</th>
<th>Phase 2 (Adult image Processing)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Auto Contrast Balancing</td>
<td>Noise Reduction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Illumination normalization</td>
</tr>
<tr>
<td>1</td>
<td>➡️</td>
<td>➡️</td>
</tr>
<tr>
<td>2</td>
<td>➡️</td>
<td>➡️</td>
</tr>
<tr>
<td>3</td>
<td>➡️</td>
<td>➡️</td>
</tr>
<tr>
<td>4</td>
<td>➡️</td>
<td>➡️</td>
</tr>
<tr>
<td>5</td>
<td>➡️</td>
<td>➡️</td>
</tr>
<tr>
<td>6</td>
<td>➡️</td>
<td>➡️</td>
</tr>
<tr>
<td>7</td>
<td>➡️</td>
<td>➡️</td>
</tr>
</tbody>
</table>

As shown in the first row of Table 3, the F-measure before applying the auto contrast was 94.320, whereas after applying the function, the F-measure improved by 0.4%. In the next row, we avoided applying the noise reduction algorithm. As shown in the table, the noise reduction algorithm improved our F-measure significantly by 8.52 percent from 88.387 to 96.904. In the next experiment, we removed all the pre-processing steps. The outcome shows that pre-processing steps can improve the F-measure by 11.34 percent. In the next round of the experiment, we removed illumination normalization. The result showed a 13.2 percent decline in F-measure performance. An experiment with noise reduction but lack of auto contrast balancing and illumination normalization showed an F-measure with value of 13.65%. The result of an experiment with auto contrast balancing but without noise reduction and illumination normalization reveals a decline of 13.5 percent in the F-measure. In the final round of our experiment, we applied all steps including illumination normalization. All the filters combined gave the highest F-measure improvement. Our final result showed a 96.904 percent improvement in the F-measure.

### V. COMPARISON OF UTILIZED METHOD

Table 4 compares the results between three available enhancements to the accuracy of adult image recognition schemes.

<table>
<thead>
<tr>
<th>Researchers</th>
<th>Auto contrast balancing</th>
<th>Normalization</th>
<th>Illumination</th>
<th>Noise reduction</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>[12]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>0.934</td>
</tr>
<tr>
<td>[20]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>0.921</td>
</tr>
<tr>
<td>[21]</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>0.95</td>
</tr>
</tbody>
</table>

**proposed scheme** ✓ ✓ ✓ ✓ 0.969

The comparison of F-measures for three methods and the proposed scheme is illustrated in Table 4. The scheme of [12] shows high precision, but its recall is low and therefore the achieved F-measure is 0.934. On the contrary, [20] scheme shows a low difference between precision and recall but due to lower values of precision and recall, it achieved a F-measure with a value of 0.921. That scheme benefits from having almost close values of recall and precision but hence they are lower than the previous scheme. [21] scheme and the proposed scheme with high precision and high recall gained F-measure values of 0.95 and 0.969 respectively, and so are higher than the other schemes. But the proposed scheme got an F-measure that is 1.9 percent higher than the others.

### VI. CONCLUSION

This paper has presented a combination of Fuzzy Logic and Deep Neural Network in illumination normalization by utilizing the integration and enhancing of fuzzy normalization in adult image recognition. This paper has aimed to improve accuracy. The new enhancement is the application of the Fuzzy Deep Neural Network Segmentation technique in order to enhance illumination normalization on the adult image. In this paper, image processing is divided into two phases. The first is the pre-processing phase. There are two pre-processing phases; the first of which comprises auto contrast balance and noise reduction. The second phase is adult image processing which contains four steps; illumination normalization, face detection, feature extraction, and image recognition. We utilized 80,400 images which were selected randomly from the Compaq and Poesia datasets and the Internet. The proposed scheme improved accuracy by
1.5% compared to the highest accuracy resulting from the existing state-of-the-art scheme.

ACKNOWLEDGMENT

This research has been supported by the Iran Telecommunication Research Center, and this method was designed and implemented as part of the applied research activities of the Socio-cultural Protection Plan by using intelligent systems in the Information and Communications Technology Research Institute.

REFERENCES


