Skin Disease Classification by Extracting Independent Components

Dr. S. Arivazhagan, Mrs. R. Newlin Shebiah, Ms. K. Divya, Ms. M. P. Subadevi
Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi - 626 005, Tamilnadu, India.

newlinshebiah@yahoo.co.in

ABSTRACT

This paper proposes an automated system based on texture analysis for recognizing human skin diseases. The spatial distributions of melanin and hemoglobin in human skin are separated by independent component analysis of skin color image. The texture features were derived from Gray Level Run Length Matrices. Minimum Distance Classifier is used to classify the type of human skin diseases. About 350 images from DERMNET database are used for evaluation and the recognition rate is found to be satisfactory.

Keywords: Skin disease, texture, run-length features, Minimum Distance Classifier

1. INTRODUCTION

Human skin is a complex surface, with fine scale geometry that makes its appearance difficult to model. Melanin and hemoglobin pigments are contained in this structure. Slight changes of pigment construction in skin may cause a rich variation in skin color. By analyzing the skin texture, a lot of observations can be made regarding the nature and coarseness of the skin. Skin diseases, if not treated earlier might lead to severe complications in the body including spreading of the infection from one individual to the other. So it is necessary to be cautious regarding skin care. Many methodologies have been proposed to analyze and recognize textures in an automated fashion. Tsumura et al [1] proposed a technique through which melanin and hemoglobin pigment content are extracted from a single skin color image by independent component analysis (ICA). In their technique, the scattering in the skin is modeled in a simple linear form in the optical density domain in which inverse optical scattering is performed by a simple inverse matrix operation. Kumar Mittra et al [2] proposed a scheme for automated detection of three classes of skin diseases by analyzing textures obtained from a collection of medical images, using features based on Gray Level Co-occurrence Matrix (GLCM) and using neural networks as classifiers.

Yuan Tian [3] used a method in which the spatial distributions of pigments in foliage which lead to color variation are separated by independent component analysis (ICA) from a single leaf color image. Texture analysis and measurements based on a statistical approach to the pattern recognition are proposed and differences in color and coarseness of skin are quantitatively evaluated [4]. Nidhal K. Al Abbadi’s approach relied on both skin color and texture features (features derived from the GLCM) to give a more efficient recognition accuracy of skin diseases [5]. Feedforward neural networks were used to classify input images to be psoriasis infected or non psoriasis infected. A fast fixed-point type algorithm that is capable of separating complex valued, linearly mixed source signals is presented and it is claimed to be robust against outliers and computationally simple [6].

Robert M. Haralick presents an approach to extract features which contain image textural characteristics such as homogeneity, gray-tone linear dependencies, contrast, number and nature of boundaries present and the complexity of the image [7]. Hawkins (1969) has extracted the Run Length features to classify the focal lesion in ultrasound live images. The overall classification accuracy using neural networks was found to be 83% [8]. Eduardo et al reviews the non-linear Independent Component Analysis and its applications to blind source separation [9].

The paper is structured as follows. The next section discusses the Proposed Method. The Section 3 gives the Recognition Results and Discussion. Finally, Section 4 gives the Concluding remarks of the proposed method.
2. PROPOSED METHODOLOGY

The two sections that involved in this work are Training and Classification. The block diagram of the proposed method is given in Figure 1.

The input image may either an image infected by a skin disease or a clear image with no skin infection. In the training phase, the input image is analyzed by independent component analysis algorithm to obtain the independent components. Texture features are extracted from the first independent component by Run length method to characterize the image.

In the classification phase, for the test image, texture features are derived as that of the training phase and compared with corresponding feature values, stored in the feature library. The classification is done using the Minimum Distance Criterion. The image from the training set which has the minimum distance when compared with the test image says that the test image belongs to the category of that training image.

2.1 Independent Component Analysis

The Independent Component Analysis (ICA) is a technique that extracts the original signals from mixtures of many independent sources without a priori information on the sources and the process of the mixture. ICA has been applied to various problems such as array processing, communication, medical image processing etc. In our method we assume that the media is constructed by two pigments and that it is captured by an imaging system with two color channels. Before applying ICA to the input skin image pre-processing has to be done on the input image to maximize the independence and to minimize the correlation between the independent components [10].

The first step is to perform centering on the input data. Let us consider that the observed input image is represented as

\[ c = As \]

(2.1)

Here \( c \) is the observed matrix, \( s \) is the source matrix and \( A \) is the mixing matrix. The mean \( E\{c\} \) is subtracted from the observed variable \( c \) so it has zero mean. By doing so, the sources \( s \) also becomes zero mean because \( E\{s\} = AE\{s\} = 0 \). To perform whitening, transform the observed variables \( c \) so that they are uncorrelated and have unit variance.

When the mixing matrix is available, \( E\{s\} \) can be estimated to be \( A^{-1}E\{c\} \). Let \( s_1(x,y) \) and \( s_2(x,y) \) denote the quantities of the two pigments, melanin and hemoglobin, which depend on the image coordinates \((x,y)\) on the digital color image. The pure color vectors of the two pigments are denoted as \( a_1 \) and \( a_2 \) respectively. It is assumed that \( a_1 \) and \( a_2 \) are different from each other. It is also assumed that the compound color vector \( c(x,y) \) on the image coordinates \((x,y)\) can be calculated by the linear combination of the pure color vectors with the quantities \( s_1(x,y) \) and \( s_2(x,y) \) as

\[ c(x,y) = s_1(x,y)a_1 + s_2(x,y)a_2 \]

(2.2)

The relative quantity and pure color vectors of each pigment can be extracted from the compound color vectors by independent component analysis without a priori information about the quantity and color vector. Let us define the following equation by using the separation matrix \( H \) and the separation vector \( e(x,y) \):

\[ E(x,y) = He(x,y) \]

(2.3)

where \( H = [h_1, h_2] \) is the separation matrix and \( e(x,y) = [e_1(x,y), e_2(x,y)]^T \) is the extracted signal vector. By finding the appropriate separation matrix \( H \), we can extract the mutually independent signals \( e_1(x,y) \) and \( e_2(x,y) \) from the compound color vectors in the image. In our method, optimization techniques based on the fixed-point method are used to find the separation matrix \( H \). Now the extracted independent signals may actually correspond to \( s_2(x,y) \) and \( s_1(x,y) \) respectively. If the number of pigments is larger than the number of channels, it is difficult to extract the independent components by relying on the reduction of signals. If the number of pigments is smaller than the number of channels, it is possible to make the number of channels equal to the number of pigments by first finding the desired sub-space using principal component analysis.

2.2 Run-Length Features Extraction

The next process after obtaining the independent components is to analyze and extract the texture features from it. The texture features extracted from the run-length matrix can produce good classification results.

For a given image, a run-length matrix \( p(i,j) \) is defined as the number of runs with pixels of gray-level \( i \) and run length \( j \) [11]. The three features that are extracted using our method are as follows:

Short Run Emphasis

\[ \sum \sum \frac{P_r(i,j)}{P_i(i,j)} \sum \sum P_i(i,j) \]  

(2.4)

Short run lengths are emphasized by dividing each run length value by the square of its length. The total number of runs in the image is the denominator.

Long Run Emphasis

\[ \sum \sum j^2 P_r(i,j) / \sum \sum P_i(i,j) \]  

(2.5)

In order to allow higher weight to the long runs, each run length value is multiplied by the square of its length.

Gray Level Non-uniformity

\[ \sum \left( \sum P_r(i,j) \right)/ \left( \sum \sum P_i(i,j) \right) \]  

(2.6)
This feature depends on high run length values. The gray level non-uniformity feature will have its lowest value if the runs are evenly distributed over all gray levels.

These features are all based on intuitive reasoning, in an attempt to capture some apparent properties of run-length distribution.

### 2.3 Classification

Classification is a process where we categorize new samples or data and assign them to distinct classes based on the features the samples carry. In this method, we use Minimum Distance Classifier (MDC) to classify the input image.

The minimum distance classifier is used to classify unknown image data to classes in multi-feature space. The distance is defined as an index of similarity so that the minimum distance is identical to the maximum similarity [12].

### 3. EXPERIMENTAL RESULTS AND DISCUSSION

Images were obtained from Dermnet Skin disease atlas. Dermnet is the largest independent photo dermatology source dedicated to online medical education though articles, photos and video. Dermnet provides information on a wide variety of skin conditions through innovative media. From the wide availability, only 5 categories were considered for the skin disease classification system. The sample images of various diseases obtained from the database are shown in Fig 2.

These images are divided into training and testing set, where 50% of the images from each group are used to train the system and the remaining images serves as the testing set.

From the independent components 3 run length features such as Long Run Emphasis, Short Run Emphasis and Gray Level Non-uniformity were extracted.

With these set of texture features, the training and the classification is carried out. Table 2 shows the classification rate of the skin disease detection system.

<table>
<thead>
<tr>
<th>Skin Type</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acne</td>
<td>96</td>
</tr>
<tr>
<td>Basal carcinoma</td>
<td>100</td>
</tr>
<tr>
<td>Splitz Nevus</td>
<td>91.6</td>
</tr>
<tr>
<td>Venous Malformations</td>
<td>100</td>
</tr>
<tr>
<td>Clear skin images</td>
<td>76</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>92.72</strong></td>
</tr>
</tbody>
</table>

### 4. CONCLUSION

The main focus of this paper is on analyzing the texture of skin thereby using it to diagnose the skin diseases. From the experimental results discussed above, we infer that the multi-class classification can serve as an effective tool in identifying skin diseases. The future work will be based on developing algorithms to identify various other skin diseases, to improve the overall efficiency and also to further reduce the computational time.
REFERENCES


