Human Face Localization Using Normalized RGB Color Space

Richa Mishra and Ravi Subban
Pondicherry University, Department of Computer Science, India.

ABSTRACT

Diagnosing human faces under dynamic situations with the high accuracy has formed the base of several applications of computer vision such as human computer interaction, video surveillances, criminal identification etc. On the contrary, the performance of face detection is limited by certain factors namely pose, occlusion, cluttered background, color, lighting variation and others. To overwhelm these limitations, this paper investigates the performance of adaptive and non-adaptive skin-color models on face detection process. These models are robust against different skin-colors and cluttered background. The paper has split into two phases. The first phase builds two adaptive and two non-adaptive skin-color models respectively. The second phase evaluates the performance of these four skin-color models on the face detection process. Experimental results reveal the performances by increasing the TPR and reducing FPR rate. These skin-color models have also compared with the other state-of-the-art methods in terms of accuracy.

INTRODUCTION

Diagnosing human faces in an image or videos has created a niche for itself among several applications of computer vision. It determines the presence and location of a face or multiple faces in an image by distinguishing it from the background or from all other patterns in the image. It serves as a pre-processing step in the fields of video conference, human-computer interaction, criminal and terrorist identification etc. (Liu et al. 2010). On the contrary, the performance of it is limited by several factors such as lighting variation, pose, occlusion, complex background, color etc. To overwhelm these limitations, extensive approaches have been investigated that are categorized into different classes based on the features used to classify face and non-face in an image. These approaches are classified into different categories: feature-based and image-based by Hjelmas and Low (2001). Further, Yang et al. (2002) classified face detection techniques into four categories: knowledge-based, feature invariant, template matching and appearance-based approaches. In knowledge-based approach, the explicit knowledge of human faces is used to locate them in any arbitrary images by converting it into meaningful and well-defined rules. But the limitation is that it does not work well under varying pose or head orientation. In feature invariant approach, structural features such as local features (eyes, eyebrows, nose and mouth) or statistical models (describe the relationship between the local features) of human faces, are used to locate them in images. This approach can be a problematic if image features are severely corrupted due to illumination, noise and occlusion. In template-based approach, deformable template methods are used that can model face geometry using different elastic models. In appearance-based approach, huge number of facial and non-facial images is used to train the system so that it can identify human faces in test images.

Usage of skin-color information for performing the classification at the initial stage is one of the simplest ways as it limits the search space for face detection in an image. It depends on the selection of a suitable color space, skin distribution model and a classifier that will classify each pixel into skin or non-skin. The skin regions in a color image can be segmented by processing one pixel at a time sequentially and independently. Instead of the benefits, the performance of the skin detection has been degraded by factors such as heavy shadows, overexposure, complex background, skin races, lighting variations, highlights, camera characteristics etc.

In this paper, simplest and efficient approaches are presented through the skin-color detection using the adaptive and non-adaptive approaches to overcome the above mentioned limitations. It is the extended work of
Subban and Mishra (Subban and Mishra, 2012, 2013) to evaluate the performance of these approaches on face detection. It is initiated with the transformation of RGB color image into normalized RGB color image. Although there are many color spaces that can be used for improvising skin detection, the normalized RGB color space has been used because of its suitability for skin and facial feature detection. In addition, human skin-color is more compactly represented in its chromaticity space than other color spaces (Caetano et al., 2002). After that two adaptive and two non-adaptive skin-color models have been applied one at a time individually and independently. This will result the binary segmented images of the given input image. Morphological operator is also applied to remove noise and holes introduced at the time of processing. Finally, the segmented components are labelled, and their area and centroid are computed. The bounding box is drawn around the detected human face. This paper is the extended work of (Subban and Mishra, 2012, 2013) for testing the performance of proposed models on face detection. The rest of the paper is organized as follows: Section 2 discusses the related work. Section 3 explains the conversion of color space from RGB to normalized RGB. Section 4 and 5 give the details of the proposed skin-detection models and the face detection process respectively. Section 6 narrates the results along with their appropriate explanation. Last section concludes the paper.

2. Related Work:

In order to discriminate the human skin from non-skin pixels, several skin-color methods have been evolved and can be classified into three main categories: explicitly-defined skin region, parametric and non-parametric models. Vezhnevets et al. (2004) presented the description, comparison and numerical evaluation of popular methods for skin modeling and detection. Kakumanu et al. (2007) presented a critical review on different skin modeling and classification approaches under controlled and uncontrolled environments in addition to the classification of different color spaces used in skin modeling.

Caetano and Barone (2001) presented a multivariate statistical model in the chromatic subspace of normalized RGB color space for human skin-color identification over the skin samples of black and white people. The performance is measured using ROC and compared it with models using a single Gaussian density. Based on the extensive skin-color samples, a few researchers (Heieh et al., 2012; Hota et al., 2006; Mostafa and Abdelazeem, 2012) investigated some inequalities that are capable to encompass the skin-color distribution in adopted color space for the skin detection process. Hsieh et al. (2012) proposed a face detection-based adaptive skin-color model. They used Haar-like face detection to locate human face skin region in an image. They normalized the pixels of the extracted skin region and described it by Gaussian distributions to define the skin-color model. The accuracy rate is measured with 95.73% on average.

Hota et al. (2006) proposed a skin-color model using one class classifier named support vector data descriptor (SVDD). They found that SVDD is more flexible in defining the boundary description as compared to other skin-color models using normalized RGB color space. Mostafa and Abdelazeem (2012) proposed a face detection system based on skin algorithm and neural networks to detect upright frontal faces only under cluttered background. They found that the normalized RGB color space outperforms than other color spaces. Desai et al. (2012) implemented the face detection system using skin detection algorithm. They have extracted the skin region by defining an inequality on normalized RGB color space. If the pixel of an image satisfies the inequality, the pixel is considered as skin otherwise non-skin. Seo et al. (2002) applied the non-adaptive approach to detect the human skin-color for estimating face region. After extracting the facial region, template matching method is applied to locate faces in a given sample image. Wimmer and Radig (2005) proposed a parametric classifier for the segmentation using normalized RGB color space that is adapted to the image conditions and ethnic groups.

3. Color Space:

The primary color space used for the representation of color image is RGB. However, the presence of varying lighting conditions, complex background, skin-like objects etc. in an image limits its use in face or skin detection. Therefore, majority of color spaces (YCbCr, HSV, YUV, HSL, YIQ etc.) have been evolved or used for the identification of skin or non-skin pixels. Normalized RGB color space is one of them which is obtained by normalizing the RGB color space. It is invariant with the changes of surface orientation relatively to the light source (Skarbek and Koschan, 1994). It does not contain luminance information. Normalized RGB color space is obtained by using the following normalization procedure.

\[
\begin{align*}
    r &= \frac{R}{\text{Base}} \\
    g &= \frac{G}{\text{Base}} \\
    b &= \frac{B}{\text{Base}}
\end{align*}
\]
where, $Base = R + G + B$. As the sum of the three normalized components is known ($r + g + b = 1$), the third component can be omitted and reduced the space dimensionality as it is enough to use only two components ($r$ and $g$) to describe the skin-color space.

4. Skin Classifier:

Skin classifiers help to locate human faces within an image or a video by discriminating each pixel as skin or non-skin. It uses skin-color or texture information for the localization. Human face texture can be used to separate the faces as an auxiliary method from other similar objects present in the background. Human skin-color can be used as an effective feature in the face detection task. Several approaches have been evolved for the purpose that have discussed in the above section. After the completion of color transformation, a skin segmentation process is performed in the first phase. Explicitly-defined skin region approach has been used for labeling the pixels as skin or non-skin. It defines the explicit boundaries for clustering the skin-color within the subspace of a given color space. The main advantage of using this method is its simplicity and popularity among researchers (Mostafa and Abdelazeem, 2012). This paper has used the two approaches (adaptive and non-adaptive) of skin-color models that is already presented in (Subban and Mishra, 2013) to evaluate their performances on the face detection process.

4.1. Non-adaptive approaches:

It uses explicitly defined inequalities for performing the skin classification. A pixel is considered as skin if it will satisfy the inequalities.

**Method I:**

$$(t_1 > r > t_2) \cap (g > t_3) \cap (R > t_4)$$

**Method II:**

$$(a > t_5) \cap (c > t_6) \cap (d > t_7) \cap (g > t_8)$$

Where, $a = r/g$; $bb = (r + g + b)^2$; $c = (r^*g)/bb$; $d = (r^*b)/bb$.

4.2. Adaptive approaches:

**Method I:**

$$\sigma_r - a\mu_r < r < \mu_r + a\mu_r$$
$$\sigma_g - a\mu_g < g < \mu_g + a\mu_g$$
$$\sigma_b - a\mu_b < b < \mu_b + a\mu_b$$

**Method II:**

$$\mu_r - a\sigma_r + \frac{\mu_r}{\sigma_r} < r < \mu_r + a\sigma_r + \frac{\mu_r}{\sigma_r}$$
$$\mu_g - a\sigma_g + \frac{\mu_g}{\sigma_g} < g < \mu_g + a\sigma_g + \frac{\mu_g}{\sigma_g}$$
$$\sigma_{Base} - a\sigma_{Base} + \frac{\mu_{Base}}{\sigma_{Base}} < Base < \mu_{Base} + a\sigma_{Base} + \frac{\mu_{Base}}{\sigma_{Base}}$$

Where, $\sigma_i$ represents the standard deviation of the $i$th component of color space, $\mu_i$ represents the mean of the $i$th component of color space in an image, and $a$ represents a threshold. The details can be found in (Subban and Mishra, 2012, 2013).

5. Face Localization:

This phase involves the localization of face candidate from the skin segmented regions. Increasing true positive rate (TPR) by reducing false positive rate (FPR) for improving the localization of human faces is the main objective of this paper. The second phase is split into two steps: computing the facial area and identifying the face region. First, the holes are filled in the segmented region. After that, the regions are labeled using 8-connected components. The area of all the connected components is calculated and compared with the predefined threshold to test whether it is face or non-face. If it is a face, the centroid of the facial area is computed and the coordinates for making the bounding box around is extracted.

6. Experimental Results and Discussions:

All the methods used in this paper have implemented using IDL (Interactive Data Language) language which is ideal software for digital image processing. Two skin-color models are used for skin detection in this paper: adaptive and non-adaptive skin-color model. After detecting the human skin regions for a given sample image, morphological operators has applied to remove holes and noises occurred at the time of execution. Finally, the binary segmented image is fed into the face detection system to localize human faces available in the input image. For performing the extensive evaluation, a dataset consisting more than hundred face images
are used for skin detection and face detection. A few of the sample images are used from the extended M2VTS face database (M2VTS project) and Caltech Face Database (Weber, 2005). The performance metrics considered for estimating the performance are TPR, FPR, Precision and Accuracy. The metrics are obtained by using the following equations:

\[
TPR = \frac{TP}{TP + FN}
\]

\[
FPR = \frac{FP}{FP + TN}
\]

\[
Precision = \frac{TP + FP}{TP + TN}
\]

\[
Accuracy = \frac{TP + TN}{TP + TN + FP + FN}
\]

Here, TP (True Positive) counts the number of pixels correctly detected as skin. FP (False Positive) counts the number of non-skin pixels detected as skin. TN (True Negative) counts the number of pixels correctly detected as non-skin. FN (False Negative) counts the number of skin pixels detected as non-skin. The manual ground truth images are also formed for the entire images of the dataset to evaluate these parameters.

Table I illustrates the experimental results of the proposed methods along with the comparisons with the existing methods. Figure 1 shows that the first proposed method in the adaptive skin-color model produces better results in terms of facial feature for most of the images. Background is classified correctly. The adaptive skin-color model depends on the value of alpha (\(\alpha\)). The second approach of adaptive skin-color model also produces better result for all the images but as compared with the previous approach it has some FPs. In the non-adaptive skin-color model, both the methods produce better result in terms of facial feature detection for most of the testing images. The problem encountered is in detecting the skin-like objects available in sample images. In the figure, the skin segmented along with their face detected images is shown to the corresponding approaches.

![Table I: Performance Metrics (%)](image)

<table>
<thead>
<tr>
<th>Performance Metrics (%)</th>
<th>Non-adaptive approaches</th>
<th>Adaptive Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Proposed Method II</td>
</tr>
<tr>
<td></td>
<td>Seo et al. (2002)</td>
<td>Proposed Method I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proposed Method II</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proposed Method II</td>
</tr>
</tbody>
</table>

![Fig. 1: Experimental results of the state-of-the-art methods and proposed methods along with their corresponding skin segmented image.](image)
7. Conclusion:

The performance of the adaptive and non-adaptive skin-color models on the face detection process is studied in this paper. The experimental results reveal that all the proposed methods yielded better result as compared to their existing approaches. We found that the non-adaptive skin-color model produces better result as compared to the adaptive skin-color model. After extracting the skin region, the face localization process has performed and makes the bounding box around the human faces available in the given image. To improvise the result by identifying multiple faces is the future scope of the paper and also to incorporate it into the face tracking system as a pre-processing step.

Table 1: Experimental result of the skin detection approaches.

<table>
<thead>
<tr>
<th>Performance Metrics (%)</th>
<th>Non-adaptive approaches</th>
<th>Adaptive Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed Method II</td>
<td>Proposed Method I</td>
</tr>
<tr>
<td></td>
<td>Seo et al. (2002)</td>
<td>Proposed Method II</td>
</tr>
<tr>
<td></td>
<td>Hsieh et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>False Positive Rate (FPR)</td>
<td>9.00</td>
<td>2.39</td>
</tr>
<tr>
<td>False Negative Rate (FNR)</td>
<td>98.83</td>
<td>12.57</td>
</tr>
<tr>
<td>Accuracy</td>
<td>97.02</td>
<td>95.05</td>
</tr>
</tbody>
</table>

REFERENCES


Extended M2VTS database – www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/.


Hota, R.N., V. Venkoparao and S. Bedros, 2006. Face Detection by Using Skin Color Model Based on One Class Classifier. In Proceedings of the 9th International Conference on Information Technology (ICIT’06), IEEE.


