Is the Eye Region More Reliable Than the Face? A Preliminary Study of Face-based Recognition on a Transgender Dataset

Gayathri Mahalingam and Karl Ricanek Jr.
Institute of Interdisciplinary Studies in Identity Sciences (IISIS),
University of North Carolina Wilmington
{mahalingamg, ricanekk}@uncw.edu

Abstract

In this work we investigate a truly novel and extremely unique biometric problem: face-based recognition for transgender persons. A transgender person is someone who undergoes gender transformation via hormone replacement therapy; that is, a male becomes a female by suppressing natural testosterone production and exogenously increasing estrogen. Transgender hormone replacement therapy causes physical changes in the body and face. This work provides a preliminary investigation into the effects of these changes on face recognition systems: commercial matcher as well as established texture-based matchers (LBP, HOG, SIFT). The performance of the full-face matchers are compared with that of the periocular region for the same feature sets. The results indicate that periocular recognition outperforms the full face on the transgender dataset under real-world conditions. In addition, we introduce a novel dataset for researchers: transgender dataset, which was organized from public sources.

1. Introduction

The face is the most expressive part of the human body. From the face one can determine a number of attributes or characteristics of a person. For example, the face conveys identity, lineage, sex, ethnicity, mood, feelings, etc. The face is powerfully expressive, and hence, it can be challenging to extract information in a robust and efficient way. Over the years the face has provided unique challenges to the biometrics community. These challenges are often described as A-PIE which stands for aging, pose, illumination, and expression. There has been a wealth of research conducted over the last 30 years to resolve these challenges. In this work, we introduce a very unique challenge, which is recognition under gender transformation.

Face recognition performance from the Good, the Bad, and the Ugly problem [8] indicates that more work is needed for face recognition to address the non-ideal–PIE–scenarios. Face verification research in the literature [3], [5], and [9] has shown that the verification performance decreases with an increase in the age span between the match pairs, and that the main–anecdotally derived–causes are due to large facial shape and texture changes. Coincidentally, changes in facial shape and texture are also observed for subjects who undergo gender transformation through hormone replacement therapy (HRT).

Gender transformation through HRT effects face fat distribution thus causing a change in the overall shape and texture of the face. Reduced fat distribution can allow for fine wrinkles and lines to become apparent whereas an increase in fat distribution stretches the dermis removing fine wrinkles and lines. For example, female to male gender transformation causes the face to become more angular (masculine) by reducing the fat distribution in the face. (The reduction in fat cells is caused by a shrinkage of the cells not an eradication of the fat cells.) In addition, the skin is either thinned (male to female) or thickened (female to male), thus introducing texture variations to the face region. It has also been shown that the factors influencing skin aging process are significantly improved as a result of HRT [11]. The result of HRT gender transformation is an increase in the within-class variation and reduction in between-class variation. Figure 1 shows sample images used in this work to illustrate the significant changes that occur during the transformation.

It can be argued that gender transformation can be considered a variant of face disguise, however, disguise falls under the broader category of biometric obfuscation [14], which refers to the deliberate alteration of the face for the purpose of masking one’s identity. Transgender persons do not undergo HRT for the purpose of biometric obfuscation, however, the question that still remains is, “will someone use HRT for the purpose of masking or creating a new identity?” The juror is out, but researchers will now have data with which they can use to develop face matching systems for this disguise variant. One has to only review Figure 2 to
ascertain the problem presented by gender transformation. The figure clearly illustrates reduction in similarity score over the course of time (HRT). This graphic was obtained by comparing the first image in the sequence to the remaining images. The similarity scores are obtained using the Pittsburgh-Pattern Face Recognition SDK [10]. It is clearly evident that the similarity between the images of the same subject decrease indicating an increase in the within-class variation. Hence, a recognition/verification system should take into account the variations caused by gender transformations or gender invariant features of the face in order to provide better recognition performance.

Despite the high adoption rate of face recognition for a number of security and consumer applications, face recognition under gender transformation has not been researched by the community. This paper introduces the problem of face recognition under the presence of this new covariate, gender transformation. Gender transformation occurs by down selecting the natural sex hormone of a person in replacement for its opposite. This is known medically as hormone replacement therapy; however, more broadly this can be described as hormone alteration or medical alteration. A periocular-based recognition framework is proposed in a hope to improve the recognition performance. The periocular region is being investigated because we believe that this region, which is anchored by the eye orbit and brow ridge, absent of fat paddings (exception, fat pads around the eyeball), and where the epidermis, skin, is sufficiently thin for both male and female, will be more stable to transgender changes.

It is to be noted that the periocular region as defined in this paper refers to the region that includes the eyes, the eyebrows, and the periorbital region (soft tissue region contained around the eye-orbit). The framework is evaluated with a transgender dataset (organized by the authors) consisting of images of 11 subjects taken under unconstrained environment across time and gender transformation. The source of the images comes from YouTube©videos. The images are extracted from frames of video sequences and only those frames that present a frontal face, where the face is dominant object in the frame. Figures 1 and 5 show sample images from the dataset. It is unknown whether these subjects have undergone facial surgery in addition to HRT. In addition to these variations, the images include other covariates such as pose, expressions, illumination, aging, etc. that are not expressly handled in this research work.

1.1. Contributions

The following are the main contributions of this research work.

• Propose the problem of face recognition under the covariate, gender transformation, hormone alterations/medical alterations.

Figure 1. Sample images of subjects showing their facial appearance before and after gender transformation (HRT). Row 1 contains pre-HRT images while row 2 contains post-HRT examples. Subjects have undergone at least 1-year of hormone replacement therapy (HRT).

• Propose a recognition framework based on the periocular region of the face for improved recognition performance when compared with full face.

• A transgender video face database with real-world images across gender transformation and other covariates.

• Performance evaluation of the proposed framework through recognition and verification experiments and comparison with various feature descriptors.

The paper is organized as follows: Section 2 provides a detailed description of the framework that includes periocular-region extraction and representation. Section 3 describes the experimental setup, the database used, the experiments, and a detailed analysis of the results. Section 4 concludes this paper.

2. Periocular Region Extraction and Representation

The extraction of feature vectors from the periocular region involves the following steps:

• Alignment and cropping of the periocular-region by registering the face image using the eye center coordinates.

• Representation of periocular region using TPLBP, LBP, HOG, and SIFT feature descriptors.

This section provides details of each of these steps in the extraction and representation of the periocular region from a face image. Figure 3 shows the framework for periocular region extraction and representation of an image.

2.1. Alignment and Cropping

The first step in aligning the face region is to extract the location of the eye centers in the face image. The center of the two eyes are automatically detected using PittPatt’s Face Recognition SDK [10].
Figure 2. Similarity scores (from PittPatt v5.2.2 SDK) indicating performance degradation as subject underwent hormone alteration for gender transformation. X-axis shows the nth image matched to the original (pre-HRT) and Y-axis is similarity score.

Figure 3. Framework for periocular region extraction and representation.

The alignment and cropping of the periocular region is achieved by means of geometric normalization. The face region in the original face image is assumed to be frontal or near frontal pose, those that were non-frontal were not utilized. The coordinates of the eye centers are used to scale, rotate, and crop the face region to a specified size. These geometric transformations are performed such that the centers of the eyes are horizontally aligned and placed on standard pixel locations. Cropping of the left and right periocular region is accomplished by defining a cropping boundary around each eye center. The extracted periocular images are then resized to 64 x 64 pixels.

2.2. Representation

The aligned and cropped periocular images are represented individually using the Three-Patch Local Binary Patterns (TPLBP) [13], Local Binary Patterns (LBP) [7], Histogram of Oriented Gradients (HOG) [1], and Scale-Invariant Feature Transformation (SIFT) [4]. While the LBP, HOG, and SIFT are pixel based feature descriptors, the TPLBP is a patch-based feature descriptor that extracts features from local patches around a central patch.

The TPLBP is a variant of LBP, where a central patch encompassing a pixel location at its center is compared with its neighboring patches to generate the feature descriptor. The TPLBP descriptor is produced by comparing the values of three neighboring patches to produce a single bit value in the descriptor code. For each pixel in the image, a patch of size $w \times w$ centered on the pixel, and $S$ neighboring patches uniformly distributed in a circle of radius $r$ around it considered. Two neighboring patches that are $\alpha$ patches apart are compared with the center patch and the descriptor code bit is set based on the neighboring patch that is more similar to the center patch. The TPLBP code is given by the following formula:

$$TPLBP_{r,m,w,\alpha}(p) = \sum_{i} (f(d(C_i, C_p) - d(C_i \text{ mod } m, C_p)))^2$$

where $C_i$ and $C_{i+\alpha}$ are two patches along the ring and $C_p$ is the central patch. The function $f(\cdot)$ is any distance function between two patches (e.g., $L_2$ norm of their gray level differences) and $f$ is defined as:

$$f(x) = \begin{cases} 1 & \text{if } x \geq \tau \\ 0 & \text{if } x < \tau \end{cases}$$

where $\tau$ is set to a value slightly larger than zero in order to provide stability in uniform regions [2]. Figure 4 shows the computation of TPLBP code for a pixel.

Patch-based approaches have been shown to provide state-of-the-art capabilities in similarity learning of faces and of general images [6]. On the other hand, LBP is invariant to minimal pose and illumination variations and HOG is invariant to geometric and photometric transformations, except for object orientation. All the above mentioned descriptors are local descriptors, in the sense that they operate on a local region around each pixel in the image.

Figure 4. Figures (a) and (b) show the computation of the three-patch LBP.
3. Experiments

This section discusses in detail the experimental setup, the dataset and the experimental results from the periocular-based face verification experiments.

3.1. Dataset

To the best of our knowledge, there exists no dataset in the literature that includes images of a subject taken during the period of gender transformation. Hence, we collected these images from Youtube videos that are compilations of the images taken during various time periods of the gender transformation. In one case it was a picture a day for three years and in others it was a picture a week or a random sampling over a year or more. The transgender dataset, as we call it includes 3056 face images of 11 subjects that span at least a year of their life showing the gender transformation variations on their face. The dataset includes an average of 278 images per subject that are taken under real-world conditions, and hence, include variations in pose, illumination, expression, and occlusion. Figure 5 shows sample images of two subjects across their gender transition.

3.2. Preprocessing

The images undergo a preprocessing stage that includes the extraction of eye center coordinates, alignment by geometric transformation using the eye center coordinates, and extracting the left and right periocular region. The extracted periocular images are then resized to $64 \times 64$ pixels. Finally, Wiener filter [12] is applied to each of these images to remove derivative noise. No other preprocessing is included, that is we don’t attempt to correct for pose, or perform any advanced illumination mitigation, etc.

3.3. Experimental Setup

In verification, the task is to determine whether a pair of images belong to the same subject or not, whereas in recognition, the task is to determine the identity of the subject (the probe) by matching it with all the images in the gallery (the enrolled). Face verification experiments on the left and right periocular region were performed individually and fused using a simple weighted score level fusion approach. Face verification experiments on the full face provide a baseline for the periocular-based face verification experiments. The full face representation refers to the extraction of feature descriptors from globally aligned full face images where the verification is performed using these feature descriptors: TLBP, LBP, HOG, and SIFT.

For the periocular-based verification experiments, feature vectors are extracted using TPLBP, LBP, HOG, and SIFT feature descriptors individually applied to the aligned and cropped left and right periocular regions. The similarity between two feature vectors is measured as the Euclidean distance between two feature vectors. For the task of verification, two periocular images are considered to be from the same subject if the Euclidean distance between their feature vectors is below a threshold value. The normalized Euclidean distance is converted to a similarity score by simply subtracting the distance from one. The fusion of the left and right periocular region is obtained at the score level, which is obtained by a weighted combination of the similarity scores from the left and right periocular region. Equal weights were used in our experiments.

The verification performance is measured in terms of Receiver Operating Characteristic (ROC) curve that is plotted with the False Acceptance Rate (FAR) as the x-axis and the True Acceptance Rate (TAR) as the y-axis. The verification rate is also measure in terms of Equal Error Rate (EER), which is defined as the error rate when FAR and TAR is equal. The ROC curve is generated by varying the threshold value for the similarity score.

In addition to verification experiments, periocular-based recognition is also performed on the dataset and evaluated using the four feature descriptors mentioned above. The performance is evaluated in terms of rank-1 recognition accuracies. The gallery includes the first one-third of images of each subject. The remaining images of the subject are used as probes. This setup of gallery and probe ensures an appropriate analyses on the effectiveness of the periocular region in matching images across gender transformation.

3.4. Verification Results

The effectiveness of periocular region under gender transformation is studied through verification experiments on left and right periocular regions. Verification is performed using each descriptor on the left, right, and the full face region. Verification performance from the score-level fusion is obtained by fusing the verification scores of both the left and the right periocular region using weighting factor of 0.7 for our experiments.

Figures 6, 7, 8, and 9 show the ROC curves generated from the performance of the TPLBP, LBP, HOG, and SIFT descriptors. The average EER is reported in Table 1. From the EERs it can be observed that the best performance is achieved using the TPLBP descriptor when compared...
Table 1. Periocular-based verification and Full face verification results. Table shows the average EER.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Left</th>
<th>Right</th>
<th>Fusion</th>
<th>Full Face</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPLBP</td>
<td>0.3254</td>
<td>0.3067</td>
<td>0.2963</td>
<td>0.3292</td>
</tr>
<tr>
<td>LBP</td>
<td>0.3933</td>
<td>0.3849</td>
<td>0.3733</td>
<td>0.4037</td>
</tr>
<tr>
<td>HOG</td>
<td>0.3540</td>
<td>0.4020</td>
<td>0.3529</td>
<td>0.3704</td>
</tr>
<tr>
<td>SIFT</td>
<td>0.4138</td>
<td>0.4087</td>
<td>0.4000</td>
<td>0.4000</td>
</tr>
</tbody>
</table>

Figure 6. ROC curves showing the verification performance of TPLBP descriptor.

Figure 7. ROC curves showing the verification performance of LBP descriptor.

Figure 8. ROC curves showing the verification performance of HOG descriptor.

Figure 9. ROC curves showing the verification performance of SIFT descriptor.

3.5. Recognition Results

The periocular-based recognition is performed using the extracted gallery and probe sets from the verification experiment. Euclidean distance measure is used to compute the match score between a pair of images and the pair with the least distance is determined as the match. A similar matching procedure is followed for full-face and score level fusion approaches.

The rank 1 recognition accuracy for all the approaches is shown in Table 2 and the CMC curve is shown in Figure 10. The results affirm the reliability of the periocular region for recognition of transgender subjects when compared with full face for the texture-based local descriptors (TPLBP, LBP, and HOG). However, the use of full face region provides better performance than the periocular region for the SIFT key-point descriptor, which is probably due to the availability of more key-points on the face than a peri-
Table 2. Periocular-based face recognition. Table shows rank-1 recognition accuracy of various descriptors on the transgender dataset.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Left</th>
<th>Right</th>
<th>Fusion</th>
<th>Full Face</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPLBP</td>
<td>71.70%</td>
<td>67.93%</td>
<td>74.50%</td>
<td>67.88%</td>
</tr>
<tr>
<td>LBP</td>
<td>54.05%</td>
<td>46.74%</td>
<td>56.40%</td>
<td>42.03%</td>
</tr>
<tr>
<td>HOG</td>
<td>61.94%</td>
<td>53.31%</td>
<td>63.56%</td>
<td>53.27%</td>
</tr>
<tr>
<td>SIFT</td>
<td>41.64%</td>
<td>42.23%</td>
<td>48.11%</td>
<td>51.05%</td>
</tr>
</tbody>
</table>

Table 2: Periocular-based face recognition. Table shows rank-1 recognition accuracy of various descriptors on the transgender dataset.

Figure 10. Cumulative Match Characteristic Curve showing performance of periocular-based recognition on transgender dataset.

Figure 11. Similarity scores for both periocular and full face using TPLBP from matching images undergoing HRT transformation with the image before transformation for all subjects. X-axis shows the image number and Y-axis shows the average similarity score.

Figure 12. Similarity scores for both periocular and full face using LBP from matching images undergoing HRT transformation with the image before transformation for all subjects. X-axis shows the image number and Y-axis shows the average similarity score.

Figure 13. Similarity scores for both periocular and full face using HOG from matching images undergoing HRT transformation with the image before transformation for all subjects. X-axis shows the image number and Y-axis shows the average similarity score.

3.6. Effects of Gender Transformation

It is evident from Figure 2 that the similarity between two images of a subject decreases across gender transformation. In order to evaluate the performance of the periocular-based recognition framework with the presence of this covariate, a recognition experiment was conducted that included the earliest image of a subject in the gallery and the rest of the images in the probe. The match scores obtained for each image in the probe can provide an insight on the invariant nature of the periocular region across gender transformation. The experiment is conducted for both the full face and the periocular region using all the descriptors. Figures 11, 12, 13, 14 shows the average similarity score obtained for subjects for both periocular and full face region by matching every image (during gender transformation) with their earliest image in the sequence. The similarity scores from matching full face images show significant variations across the timeline for all the descriptors, which potentially indicate the effect of significant face variations. However, it is to be noted that there are insignificant variations in the similarity scores for all the descriptors (except SIFT) on the periocular region, possibly indicating the invariant nature of the periocular region to gender transforma-

5. Acknowledgements

This work was made possible due to support from the FBI Biometric Center of Excellence and the Army Research Lab in conjunction with CASIS Center for Advanced Studies in Identity Science.

References