ENCODING AND DECODING LOCAL BINARY PATTERNS FOR HARSH FACE ILLUMINATION NORMALIZATION

Felix Juefei-Xu and Marios Savvides
Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

ABSTRACT
In this work, we propose a new illumination normalization technique based on a simple, yet widely used descriptor: local binary patterns (LBP). We capitalize on the fact that LBP retains tolerance to illumination changes and use the LBP mapping to remove illumination variations cast on face images. Through learning a reverse mapping from the LBP domain to the pixel domain, we are able to recover the illumination normalized face with high fidelity. The reverse mapping step is made possible via a joint dictionary learning framework between the LBP domain and the pixel domain. The illumination normalized faces using our proposed LBP encoding and decoding method not only exhibit very high fidelity against neutrally illuminated face, but also allow for a significant improvement in face verification experiments using even the simplest nearest-neighbor classifier. These conclusions are drawn after benchmarking our algorithm against 22 prevailing illumination normalization techniques on Extended YaleB database which has been widely adopted for challenging face illumination problems.

Index Terms— Illumination Normalization, LBP

1. INTRODUCTION
With the rising expectations on the robustness of face recognition systems for both mobile applications and surveillance purposes, the study of face illumination normalization problem has gained a lot of attention in the past decade. This is because illumination still is one of the biggest challenges that can severely jeopardize the performance of face recognition systems. In this work, we propose a simple, yet very effective tool for harsh face illumination normalization. Motivated by the fact that the local binary patterns (LBP) [1] exhibit tolerance to illumination changes, we demonstrate a novel illumination normalization method through the encoding and decoding of the LBP.

A recent and very interesting work by Han et al. [2] explicitly exploits maximizing the separability of different subjects’ faces as the objective for illumination preprocessing. The authors first decompose the input face via scale-space decomposition and then the normalized face is computed as a linear combination of the decomposed coarse-to-fine bands.

Fig. 1. The flowchart of the proposed illumination normalization method based on encoding and decoding LBP. Illumination variations are first removed by the forward mapping $F$. Illumination normalized face is then recovered through a reverse mapping $F^{-1}$ obtained via joint dictionary learning. The reconstructed face displays high fidelity compared to the well-illuminated face of the same subject from the YaleB database.

Most importantly, the combining coefficients are learned from a training set by maximizing the Fisher criterion. The entire optimization is solved by simulated annealing. Chen et al. [3] extract illumination invariant features based on natural images statistics, with which they derive a Wiener filter approach to separate the illumination-invariant features from an image. Matsukawa et al. [4] base their method on small- and large-scale features [5]. They focus on the large-scale feature (LF) in the presence of cast shadows where large-scale features are decomposed using PCA basis, and then the normalized LF is the quotient image of the original LF and the approximated LF using top-$K$ PCA basis. In the end, small- and large-scale features are combined to produce a final output. Han et al. [6] has provided a comparative study on 12 representative illumination preprocessing methods and grouped them into 3 categories: (1) grey-level transformation, e.g., histogram equalization, logarithmic transform, (2) gradient or edge extraction, e.g., Laplacian of Gaussian, and (3) reflectance field estimation, e.g., work of Tan and Tiggs [7].

2. PROPOSED METHOD
LBP is tolerant to illumination changes [1] because the encoding remains stable as long as the partial ordering of the pixels in the local patch remains unchanged. When harsh il-
lumination/shadow is cast on face images, each local patch changes intensity non-uniformly. However, as long as the relative intensities between the neighboring pixels and the center pixel remain the same, the LBP feature would stay stable [8, 9, 10, 11, 12, 13, 14, 15, 16]. Therefore, the first step in our proposed method is to apply LBP to encode the input dark face in order to largely remove the illumination variations. Second, we recover from the intermediate LBP image to obtain final illumination normalized one. This is carried out by a dictionary learning approach. The flowchart of the proposed method is shown in Figure 1.

The LBP operation is a nonlinear 1-to-1 mapping $F$ from the pixel domain to the LBP domain. The non-linearity is due to the thresholding of the neighboring pixels when compared to the center pixel within each local patch. It is also obvious that there could be many images that share the same LBP representation, as long as the local partial ordering of the pixels stays the same. Therefore, the mapping from LBP to pixel domain is N-to-1. Thus, obtaining the reverse mapping $F^{-1}$ to obtain one single illumination normalized image requires more constraints.

Explicitly learning such a reverse mapping $F^{-1}$ is infeasible and not necessary. We can capitalize on the joint learning framework to implicitly learn the mapping.

Jointly, we can learn an overcomplete dictionary for the pixel domain faces and another for the LBP faces such that the sparse approximation coefficients for the intermediate LBP face can be shared by the pixel domain face which can lead to the reconstruction of the illumination normalized face.

The first objective is to learn a dictionary for the pixel domain faces by solving: $\min_{D, X} \| Y - DX \|^2_F$ such that $\forall i, \| x_i \|_0 < K$. Similarly, the second objective is to learn a dictionary for the LBP faces: $\min_{D_{LBP}, X} \| Y_{LBP} - D_{LBP}X \|^2_F$ such that $\forall i, \| x_i \|_0 < K$.

Combining these two objectives and solving them jointly allows us to force a common $K$-sparse representation. Our primary problem is therefore:

$$\arg \min_{D, X, D_{LBP}} \| Y - DX \|^2_F + \| Y_{LBP} - D_{LBP}X \|^2_F$$

such that $\forall i, \| x_i \|_0 < K$ (1)

Here $D$ and $D_{LBP}$ are the two overcomplete dictionaries for the pixel domain and the LBP domain faces, $Y$ and $Y_{LBP}$ are the pixel domain and the LBP domain training images respectively, and $X$ is the sparse coefficient matrix shared between the two domains. Solving the formulation is achieved by a simple rearrangement before using the standard K-SVD [17] dictionary learning approach as previously observed [18]:

$$\arg \min_{D, D_{LBP}, X} \left\| \begin{pmatrix} Y \\ Y_{LBP} \end{pmatrix} - \begin{pmatrix} D \\ D_{LBP} \end{pmatrix} X \right\|_F^2$$

such that $\forall i, \| x_i \|_0 \leq K$ (2)

This method is open set, enabling reconstruction of any face that is not present in the training set. The dictionaries are pre-trained using 200,000 mugshot-type face images. Figure 2 shows the learned dictionaries for both the pixel and the LBP domain faces.

3. EXPERIMENTS

In this section, we will evaluate the effectiveness of the proposed method. We benchmark it against 22 prevailing illumination normalization techniques with open-source implementations consolidated in [19]. Parameters are tuned towards the most optimal ones. Unfortunately, implementation of some of the recent work discussed in Section 1 cannot be found. We also omit the basic algorithms like histogram equalization, logarithmic transform, gamma intensity correction, Gaussian high-pass, directional gray-scale derivative, etc., for the sake of brevity. The 22 algorithms are: (1) single scale retinex (SSR), (2) multi scale retinex (MSR), (3) adaptive single scale retinex (ASR), (4) homomorphic filtering (HOMO), (5) single scale self quotient image (SSQ), (6) multi scale self quotient image (MSQ), (7) discrete cosine transform (DCT), (8) retina modeling (RET), (9) wavelet (WA), (10) wavelet denoising (WD), (11) isotropic diffusion (IS), (12) anisotropic diffusion (AS), (13) steering filter (SF), (14) non-local means (NLM), (15) adaptive non-local means (ANL), (16) modified anisotropic diffusion (MAS), (17) gradientfaces (GRF), (18) single scale Weberfaces (WEB), (19) multi scale Weberfaces (MSW), (20) large and small scale features (LSF), (21) Tan and Triggs (TT), and (22) difference of Gaussian filtering (DOG).

3.1. Face Illumination Databases

Among existing databases targeting face illumination problems, Yale Face Database B (YaleB) [20] and Extended Yale Face Database B (ExtYaleB) [21] contain the most extreme and challenging illumination conditions. The YaleB contains single light source images of 10 subjects each seen under 576 viewing conditions (9 poses × 64 illumination conditions).
Table 1. Results including verification rate (VR) at 0.1% false accept rate (FAR), equal error rate (EER), and peak-signal-to-noise ratio (PSNR) are tabulated for experiments on YaleB+. VR/PSNR/overall rankings are also shown. ∑ column is the sum of RkVR and RkPSNR.

<table>
<thead>
<tr>
<th>Method</th>
<th>VR (EER)</th>
<th>RkVR</th>
<th>PSNR</th>
<th>RkPSNR</th>
<th>∑</th>
<th>RkALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>0.0750 (0.4319)</td>
<td>22</td>
<td>10.3763</td>
<td>20</td>
<td>42</td>
<td>23</td>
</tr>
<tr>
<td>SSR</td>
<td>0.0755 (0.4319)</td>
<td>22</td>
<td>10.3763</td>
<td>20</td>
<td>42</td>
<td>23</td>
</tr>
<tr>
<td>MSR</td>
<td>0.0749 (0.4252)</td>
<td>23</td>
<td>10.3984</td>
<td>21</td>
<td>43</td>
<td>22</td>
</tr>
<tr>
<td>ASR</td>
<td>0.1332 (0.2942)</td>
<td>4</td>
<td>6.4917</td>
<td>23</td>
<td>27</td>
<td>16</td>
</tr>
<tr>
<td>HOMO</td>
<td>0.0906 (0.4343)</td>
<td>18</td>
<td>11.5667</td>
<td>16</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>SSR</td>
<td>0.1181 (0.3160)</td>
<td>6</td>
<td>10.1418</td>
<td>21</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>MSQ</td>
<td>0.1173 (0.3497)</td>
<td>8</td>
<td>10.9452</td>
<td>17</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>DCT</td>
<td>0.0972 (0.3374)</td>
<td>16</td>
<td>12.3461</td>
<td>9</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>WA</td>
<td>0.1096 (0.4546)</td>
<td>11</td>
<td>12.1002</td>
<td>14</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>WD</td>
<td>0.1382 (0.2963)</td>
<td>2</td>
<td>12.7586</td>
<td>6</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>IS</td>
<td>0.1047 (0.4117)</td>
<td>12</td>
<td>12.6833</td>
<td>8</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.1346 (0.3841)</td>
<td>3</td>
<td>13.1269</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 3. Visual results of 2 subjects from the YaleB database under all 64 illumination variations. (top) the original images, (middle) the LBP encoded image, and (bottom) the LBP decoded images. The original images show severe illumination variations with many faces being almost entirely dark for human perception. Please zoom in for details.

The ExtYaleB contains images of 28 human subjects under 9 poses and 64 illumination conditions. We combine the unique subjects in both database and call it YaleB+. So we have 10 + 28 = 38 subjects in total. We only choose the frontal image with all the illumination variations. The entire database contains 38 × 64 = 2432 images.

3.2. Fidelity Experiments

We investigate the fidelity of the illumination normalized image against the neutrally-illuminated face of the same subjects in the databases. We adopt the broadly used peak signal-to-noise ratio (PSNR) as the fidelity measurement [22, 23, 24]. All the dark images are resized to be 32 × 32 and then various algorithms are applied for illumination normalization. We report the average PSNR for each of the algorithms in the PSNR columns in Table 1. In addition, some visual results of the proposed method are shown in Figure 3. The original images show severe illumination variations with many faces being almost entirely dark for human perception. However, after the LBP encoding, the intermediate representations in the LBP domain show much higher stability. It captures minute variations even in those almost pure black patches. Therefore, LBP is a good candidate for removing illumination variations. We can also see that the recovered illumination normalized faces after LBP decoding show high level of fidelity with minimum amount of artifacts.

3.3. Face Verification Experiments

We also conduct face verification experiments by matching all the illumination normalized images to themselves. We utilize the simplest nearest neighbor classifier based on the normalized cosine distance (NCD) between each query and gallery image [25, 26, 27, 28, 29, 30]. The features are the raw representation out of each algorithm. Thus, the resulting similarity matrix will be of size 2432 × 2432 for YaleB+, whose entry SimM(i, j) is the NCD between query i and gallery j. The experimental results are reported using receiver operating characteristic (ROC) curves. Both the verification rate (VR)
We have presented a practical and effective method for harsh face illumination normalization. Such a method exhibits very high level of fidelity, which is outstanding among many competing algorithms. The formulation of the method is based on the encoding and decoding of the LBP, a widely used descriptor that is tolerant to illumination changes. We capitalize on this fact and use the LBP mapping to remove illumination variations cast on the face image. Through learning a reverse mapping from the LBP domain to the pixel domain, we are able to recover the illumination normalized face with high fidelity. The reverse mapping step is made possible via a joint dictionary learning framework between the LBP domain and the pixel domain. The illumination normalized faces using our proposed LBP encoding and decoding method not only exhibit very high fidelity against neutrally illuminated face, but also allow for a significant improvement in face verification experiments using even the simplest classifier. The effectiveness of the approach is confirmed after benchmarking our algorithm against 22 prevailing illumination normalization techniques on Extended YaleB database which is widely adopted for harsh face illumination problems.

4. CONCLUSIONS

We have presented a practical and effective method for harsh face illumination normalization. Such a method exhibits very high level of fidelity, which is outstanding among many competing algorithms. The formulation of the method is based on the encoding and decoding of the LBP, a widely used descriptor that is tolerant to illumination changes. We capitalize on this fact and use the LBP mapping to remove illumination variations cast on the face image. Through learning a reverse mapping from the LBP domain to the pixel domain, we are able to recover the illumination normalized face with high fidelity. The reverse mapping step is made possible via a joint dictionary learning framework between the LBP domain and the pixel domain. The illumination normalized faces using our proposed LBP encoding and decoding method not only exhibit very high fidelity against neutrally illuminated face, but also allow for a significant improvement in face verification experiments using even the simplest classifier. The effectiveness of the approach is confirmed after benchmarking our algorithm against 22 prevailing illumination normalization techniques on Extended YaleB database which is widely adopted for harsh face illumination problems.

5. REFERENCES


