FACE RECOGNITION BY FUSING BLUR-INVARIANT TEXTURE AND STRUCTURE FEATURES ACROSS NON-UNIFORM MOTION BLUR, ILLUMINATION, AND POSE

H.Srivigneshwari¹, U.Neeraja²
¹(UG Scholar, Department of ECE, IFET College of engineering,srivigneshwari@et@gmail.com)
²(UG Scholar, Department of ECE, IFET College of engineering, Villupuram, Tamilnadu, India)

Abstract—Face recognition under blur, illumination and pose is a difficult task in image processing. The existing method performs good under illumination and pose, but fail in case of blurring. So we propose a new method for blurred face recognition. First the structure and texture blur-invariant features are extracted and the complete description on blurred image is generated by fusing those features. LPQ is extracted in a densely sampled way and to enhance its performance a vector of locally aggregated descriptors (VLAD) is employed for texture blur-invariant feature. The histogram of oriented gradient (HOG) is used for structure blur-invariant feature. Then the improved HOG is extracted and then fused with the original HOG by canonical correlation analysis (CCA). For handling pose and illumination variations, we follow MOBILAP algorithm. The expected results demonstrate our improvements and performance in blurred face recognition.

Keywords— Face Recognition; Illumination; Invariant Texture; Non-Uniform Motion;

1. INTRODUCTION

Face recognition has become one of the most active topics in computer vision research due to its many potential applications [1]. However, despite significant progress in the last decade, the design of recognition algorithms that are effective over a wide range of viewpoints, occlusions, aging of subjects, and complex outdoor lighting is still a major area of research. While significant number of works had proposed to addressing these issues, problems caused by image degradations due to the factors such as pose, illumination, blur, noise, and sampling are mostly overlooked. Such image degradations significantly affects the performance of face recognition systems and are often present in images and videos in real-time applications such as watch-list monitoring and video surveillance. Only recently has the research community started to look at facial image degradations, e.g., through facial denoising [2]. The focus of this paper is therefore on coping with blur, illuminations and pose variations and, in particular, automatic deblurring of face images for enhancing the recognition performance. Blur affects the appearance of faces in images, causing two main problems for face recognition: 1) Blur changes the facial appearance of an individual drastically 2) when blurred, different individuals tend to appear more similar. A few existing methods attempt to handle these problems. However, they are not yet satisfactory when facing the significant amount of blur that is common in many real-time settings. For instance, Stainvas and Intrator [3] match a query image to artificially blurred copies of the original sharp target images registered for identification. The method can alleviate the first problem of dissimilarity caused by blur, but the second problem of similarity exits. In our approach, first we remove the blur from facial appearances using blind image deconvolution [4]. The deblurred images can then be used to perform more robust recognition. Obviously, such an approach can solve both problems 1) and 2) simultaneously, but requires a Point Spread Function (PSF) that represents the blurring process. In the field of blind image deconvolution, many methods have been proposed for deploring from a single image [5]. For instance, Chan and Wong [6] simultaneously infer a PSF and deblur an image using total variation regularization. Other methods attempt to model the smoothness of intensity changes around edges. A PSF is inferred using information derived from this smoothness using the variation of Gaussian scale [7], [8], wavelet coefficient [9], [10], the summation of image derivatives [11], or alpha values representing the object boundary transparency [12]. These methods have to solve an ill-posed problem because they do not adequately exploit the prior knowledge of the image content, and so the quality of the deblurred image using an inferred PSF is often poor. As our experiments will reveal, the deblurred images using these methods are insufficient for accurate face recognition. Yuan et al. [13] and Ancuti et al. [14] infer a PSF using multiple images captured from the same scene. This setting limits face recognition applications. Fergus et al. [15] infer a PSF using heavy-tailed natural image priors, but these priors are very generic and so fairly weak, and the PSF inference is computationally expensive. It appears then that the majority of the previous methods for blind image deconvolution use the smoothness of intensity changes around edges to infer the PSF from a single image without prior knowledge of the image contents. These methods thus often infer a poor quality PSF as it is difficult to distinguish between blurred edges and smooth object surfaces.

In this paper, we propose a face recognition algorithm that is robust to non-uniform (i.e., space-varying) motion blur arising from relative motion between the camera and the subject, pose and illuminations variations. The texture and
structure blur-invariant features are extracted and the complete description on blurred image is generated by fusing those features. In case of pose and illumination variations, the gallery image is transformed into the given probe image’s pose and illumination and then the face is recognized.

2. TEXTURE BLUR-INVARIANT FEATURE EXTRACTION

The Eq. 1 formulated the image blurring process, where \( f(x, y) \) is the original image, \( g(x, y) \) is the observed image. \( h(x, y) \) and \( n(x, y) \) represent the point spread function (PSF) and the additive noise respectively. The additive noise is often ignored in theoretical analysis.

\[
g(x, y) = f(x, y) \ast h(x, y) + n(x, y)
\]

The texture blur-invariant information is extracted by the LPQ descriptor. It uses the short term Fourier transform (STFT) to transform the image to its Fourier domain representation, which is defined as below:

\[
G(u, v) = \sum_{x} \sum_{y} g(x, y) e^{-j2\pi(xu+yv)/M}
\]

The STFT is calculated in a \( M \times M \) neighborhood centered at pixel \( (x, y) \). \( Nx \) and \( Ny \) represent the neighborhood region. LPQ extracts the texture information by calculating the STFT on frequency points set

\[
V = [G(u_1, x), G(u_2, x), G(u_3, y), G(u_4, x)]
\]

The representation can be further refined by \( W \) as shown below:

\[
W = [\text{Re}(V), \text{Im}(V)]
\]

LPQ encodes the calculated phase information using the equation defined as Eq.

\[
b = \sum_{i=1}^{S} q_i 2^{i-1}
\]

where \( q_i \) is the quantization of the \( i \)-th elements in \( W \), which is calculated as below:

\[
q_i = \begin{cases} 1, & \text{if } W_j \geq 0 \\ 0, & \text{otherwise} \end{cases}
\]

Finally, the histogram of the encoded values for all pixels in the image is obtained as the LPQ descriptor representation.

We firstly improve LPQ by extracting it in a densely sampled way. For a given image, a sliding sub-window having the size of \( PW \times PH \) is used. The sliding sub-window screens the image with vertical stride of \( SY \) and horizontal stride of \( SX \).

After screening the local patches are obtained. To further extract the structure information among each patch, we divide the patch into \( 2 \times 2 \) small cells. LPQ is extracted from the cells and concatenated into a 1024-dimensional local patch representation.

After the local patch LPQ representation is obtained, VLAD is used to encode the local patches to a final texture blur-invariant image representation. VLAD aggregates descriptors based on the accumulated differences of the local descriptor and its assigned visual word. It is able to characterize the distribution of the local descriptor with respect to the visual word.

For VLAD, a codebook with \( k \) visual words is learned from training samples by k-means. The codebook is represented as below, where \( ci \) represents the learnt visual word:

\[
C = \{c_1, ..., c_k\}
\]

Each given local patch \( x \) is assigned with its nearest visual word \( c_i \)

\[
c_i = \text{NN}(x)
\]

where \( \text{NN} \) finds the nearest neighbor of the given \( x \) in the codebook. VLAD uses the following approach to obtain the accumulated differences:

\[
v_{ij} = \sum_{r \in \text{NN}(x)} (x_j - c_{ij})
\]

In Eq. 9, \( x_j \) and \( c_{ij} \) represent the \( j \)-th component of the local descriptor \( x \) and its assigned visual word \( c_i \). Assuming the dimensionality of each local descriptor is \( d \), each \( v_{ij} \) can be calculated for \( i = 1..k \) and \( j = 1..d \). The vector of \( v \), which is a concatenation of \( v_{ij} \), can be obtained, which is the final texture blur-invariant representation of the image and its dimension is \( k \times d \). The final VLAD representation can be then normalized with square rooting and \( L_2 \) normalization.

3. STRUCTURE BLUR-INVARIANT FEATURE EXTRACTION

The HOG descriptor extracts the structure information of the image by calculating the gradient of the given image as shown below:

\[
m(x, y) = \sqrt{G_x^2 + G_y^2}
\]

where \( x \) and \( y \) are the gradients of horizontal and vertical direction, which can be calculated by the mask \([-101] \). \( m(x, y) \) is the gradient magnitude of the given pixel \((x, y)\). The direction of the gradient can be also calculated as below:

\[
\theta(x, y) = \arctan\left(\frac{G_y}{G_x}\right)
\]

\((x, y)\) is the gradient direction of the given pixel \((x, y)\), whose range can be set. After the calculation of the gradient of each pixel, the image is divided into several image blocks, which are further split into cells. There may be overlaps between image blocks. For each cell, the histogram of the gradient direction is built by the calculated gradient magnitude and direction. Each pixel in the given cell has a weighted vote for the given channel. The weighted vote is
usually the gradient magnitude and the channel is based on the split of the gradient direction. After the histogram of each cell is obtained, the histograms of cells of the same block are concatenated to form the image’s HOG representation. As we can see, HOG is based on the histogram of image gradient information. Although image blur may affect both the magnitude and direction, from the feature extraction’s point of view, HOG is somewhat blur invariant. We can notice that after image is blurred, the strong gradient is not affected so much. The weak gradient, which is more sensitive with image blur, is affected much more than strong gradient. We can further enhance the HOG’s blur invariance by eliminating the magnitude of the weak gradient as shown below:

\[ H_i = \begin{cases} H_i & H_i \geq \text{mean}(H) \\ 0 & H_i < \text{mean}(H) \end{cases} \]

where \( H \) is the original HOG descriptor, \( i \) \( H \) is the \( i \)-th component of the HOG descriptor. The above Eq. sets the magnitudeless than the mean magnitude to zero, which eliminates the weak gradient and enhances the HOG’s blur invariance. Furthermore, to enhance its blur invariance without losing the discriminative power, we use CCA to fuse the eliminated HOG with the original HOG, which keep the discriminative power with enhanced blur invariance at the same time. The descriptor after fusion is the final improved HOG, which extracts the structure blur-invariant information. After the extraction of texture blur-invariant information and structure blur-invariant information, we again use CCA to fuse the two different blur-invariant features to form the final representation of the blurred image. The final representation contains a complete description of the blurred image, from both the texture and structure side. The representation can be used for blurred face recognition.

4. HANDLING ILLUMINATION VARIATIONS

To handle illumination variations, we modify our basic blur-robust algorithm (NU-MOB) by judiciously utilizing the following two results:

In the seminal work of [20], it has been shown that if the human face is modeled as a convex Lambertian surface, then there exists a configuration of nine light source directions such that the subspace formed by the images taken under these nine sources is effective for recognizing faces under a wide range of lighting conditions. Using this “universal configuration” of lighting positions, an image \( f \) of a person under any illumination condition can be written as

\[ f = \sum_{i=1}^{9} \alpha_i f_i \]

where \( \alpha_i, i = 1, 2, \ldots, 9 \) are the corresponding linear coefficients. The fis, which form a basis for this 9D subspace, can be generated using the Lambertian reflectance model as

\[ f_i(r, c) = \rho(r, c) \max(n(r, c)^T s_i, 0) \]

Where \( \rho \) and \( n \) are the albedo and the surface normal, respectively, at the pixel location \( (r, c) \), and \( s_i \) is the illumination direction. Following [19], we approximate the albedo \( \rho \) with a frontal, sharp, and well-illuminated gallery image captured under diffuse lighting, and use the average (generic) 3D face normals from [33] for \( n \). In [19], it has been shown that for the case of illumination, the set of all images under varying illumination forms a bi-convex set, i.e., if we fix the illumination, the resulting subsets convex. According to the illumination model for faces, the set of illuminated images obtained by illuminating a focused gallery image using the illumination model also forms a convex set. Based on these results, we develop illumination robust face recognition algorithm. The solution that we seek can be posed as the minimization of the following cost function given by

\[
\begin{align*}
|b_{Rm} \cdot \alpha_{m,i}| = \arg \min \left[ W(g - \sum_{i=1}^{9} \alpha_i A_m, b_{Rm}) \right]_1^2 + \beta |b_{Rm}|_1 \\
\text{subject to } b_{Rm} \geq 0
\end{align*}
\]

We adopt the alternating minimization strategy outlined in [19] to solve the above equation. We first obtain the nine basis images \( f_m, i = 1, 2, \ldots, 9 \) for each gallery image \( f_m, m = 1, 2, \ldots, M \). Next, for each gallery image \( f_m \), we estimate the optimal illumination coefficients \( \alpha_{m,i} \) by solving the above equation. To determine the identity of the probe, we transform (re-illuminate) each of the gallery images fusing the estimated illumination coefficients \( \alpha_{m,i} \), compute the LBP features from these transformed gallery images and compare them with those from the probe g to find the closest match.

We now elaborate on the two steps involved in our AM algorithm. For any gallery image \( f_m \), in the first iteration, we are estimating the nine illumination coefficients \( \alpha_{m,i} \) by solving the linear least squares problem \( g = Lm \alpha \), where \( Lm \) is a matrix whose nine columns contain the basis images \( f_m \), corresponding to the subject \( m \) lexicographically ordered as vectors, and \( \alpha \) \( = \) \( \{ \alpha_{m,1}, \alpha_{m,2}, \ldots, \alpha_{m,9} \} \) is its corresponding illumination coefficients. Now, we create a new relit gallery image from the basis images using the estimated illumination coefficients \( \alpha_{m,i} \). This completes the first step of the alternation wherein we fixed the estimated blur and the illumination.

Handling Pose Variations

Most of the face recognition algorithms are robust to small variations in pose (~15°) [25], but the drop in performance is severe for greater yaw and pitch angles. The reason behind this drop in accuracy is that intra-subject variations caused by rotations are often larger than inter-subject differences. Clearly, there is no overstating the formidable nature of the problem at hand—recognizing faces across blur, illumination and pose. To this end, we next propose our modified MOBILAP algorithm which, using an estimate of the pose, matches the incoming probe with a synthesized on-frontal
gallery image. A recent work [34] that unifies face detection, pose estimation, and landmark localization has also adopted to 15° discretization. This method, suited for focused, cluttered images, detects the face(s) and returns a quantized estimate (between −90° to 90° every 15°) of the pose(s).

We use this technique to obtain an estimate of the pose of the blurred probe image. We note that there are errors in landmark localization due to blur, and the method in [34] can then yield inaccurate pose estimates with the true pose being returned only about 45−55% of the time. However, it almost always returns an estimate which is within ±15° of the true pose. Using this estimate, we synthesize, from each frontal gallery, the image of the subject under the new pose with the help of the average depthmap used. These synthesized poses now form the new gallery set.

Although the shape of each subject’s face may vary from the generic depthmap, the algorithm retains its simplicity and the increase in computational time due to this step is only minimal. The nine illumination basis images are estimated as before using Lambertian reflectance model, with ρ now being the new synthesized pose and n being the surface normals recomputed from the rotated depthmap.

5. CONCLUSION

We proposed a novel method for face recognition under blur, illumination and pose variations by the fusion of texture and structure blur-invariant features and modified MOBIL algorithm. We showed that the set of all images obtained by non-uniformly blurring a given image using the fusion of texture and structure blur-invariant features. Capitalizing on this result, we initially proposed face recognition under non-uniform motion blur. We then showed that the set of all images obtained from a given image by illumination forms a bi-convex set, and used this result to develop our modified non-uniform motion blur and illumination-robust algorithm MOBIL. We then extended the capability of MOBIL to handle even non-frontal faces by transforming the gallery to a new pose. The experiments are performed on different datasets. The results demonstrate that our improvements and proposition can outperform the other face recognition algorithms.

REFERENCES