Motion, Blur, Illumination based Face Recognition

Anand M.S
PG Student
Department of ECE
Satyam College of Engineering
E-mail : msanand89@gmail.com

Abstract- Existing methods for performing face recognition in the presence of blur are based on the convolution model and cannot handle non-uniform blurring situations that frequently arise from tilts and rotations in hand-held cameras. In this paper, we propose a methodology for face recognition in the presence of space-varying motion blur comprising of arbitrarily-shaped kernels. We model the blurred face as a convex combination of geometrically transformed instances of the focused gallery face, and show that the set of all images obtained by non-uniformly blurring a given image forms a convex set. We first propose a nonuniform blur-robust algorithm by making use of the assumption of a sparse camera trajectory in the camera motion space to build an energy function with $l1$-norm constraint on the camera motion. The framework is then extended to handle illumination variations by exploiting the fact that the set of all images obtained from a face image by non-uniform blurring and changing the illumination forms a bi-convex set. Finally, we propose an elegant extension to also account for variations in pose.
Keywords: **Face recognition, non-uniform blur, sparsity, illumination, pose.**

I. INTRODUCTION

It is well-known that the accuracy of face recognition systems deteriorates quite rapidly in unconstrained settings [1]. This can be attributed to degradations arising from blur, changes in illumination, pose, and expression, partial occlusions etc. Motion blur, in particular, deserves special attention owing to the ubiquity of mobile phones and hand-held imaging devices. Dealing with camera shake is a very relevant problem because, while tripods hinder mobility, reducing the exposure time affects image quality. Moreover, in-built sensors such as gyros and accelerometers have their own limitations in sensing the camera motion. In an uncontrolled environment, illumination and pose could also vary, further compounding the problem. The focus of this paper is on developing a system that can recognize faces across non-uniform (i.e., space-variant) blur, and varying illumination and pose.

Traditionally, blurring due to camera shake has been modeled as a convolution with a single blur kernel, and the blur is assumed to be uniform across the image [2], [3]. However, it is space-variant blur that is encountered frequently in hand-held cameras [4]. While techniques have been proposed that address the restoration of non-uniform blur by local space-invariance approximation [5]–[7], recent methods for image restoration have modeled the motion-blurred image as an average of projectively transformed images [8]–[12]. Face recognition systems that work with focused images have difficulty when presented with blurred data. Approaches to face recognition from blurred images can be broadly classified into four categories. (i) Deblurring-based [13], [14] in which the probe image is first deblurred and then used for recognition. However, deblurring artifacts are a major source f error especially for moderate to heavy blurs. (ii) Joint deblurring and recognition [15], the flip-side of which is computational complexity. (iii) Deriving blur-invariant features for recognition [16], [17]. For handling illumination, there have mainly been two directions of pursuit based on (i) the 9D subspace model for face [20] and (ii) extracting and matching illumination insensitive facial features [21], [22]. A subspace learning approach using image gradient orientations for illumination and occlusion-robust face recognition has been proposed in [24]. Practical face recognition algorithms must also possess the ability to recognize faces across reasonable variations in pose. Methods for face recognition across pose can broadly be classified into 2D and 3D techniques.

A sparse minimization technique for recognizing faces across illumination and occlusion has been proposed in [27], while [28], which is based on similar principles, additionally offers robustness to alignment and pose. But these works do not deal with blurred images. To the best of our knowledge, the

only attempt in the literature at recognizing faces across non-uniform blur has been made in [17] in which the uniform blur model is applied on overlapping patches to perform recognition on the basis of a majority vote.

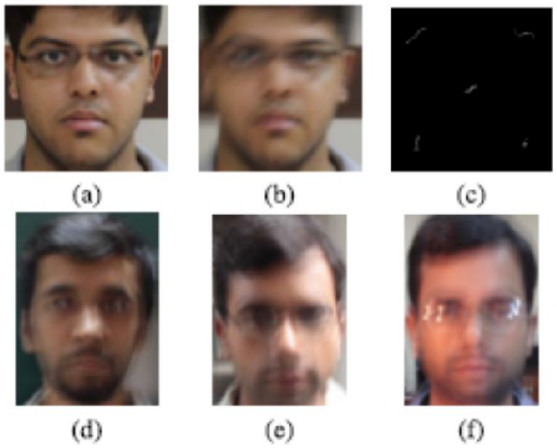

Figure 1.1: (a) Focused image, (b) synthetically blurred image obtained by applying random in-plane translations and rotations on the focused image, (c) point spread functions (PSF) at various locations in the image showing the presence of non-uniform blur which cannot be explained by the convolution model (best viewed as PDF), and (d, e, f) real blurred images from the dataset we ourselves captured using a hand-held camera.

However, they do not explicitly model illumination changes going from gallery to probe. We would like to point out that both [17] and [19] limit their discussion to frontal faces. 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.

Following [19], we assume that only a single gallery image is available. The camera transformations can range from in-plane translations and rotations to out-of-plane translations, out-of plane rotations, and even general 6D motion. An example is shown in Fig. 1.1. Observe that the blur on the faces can be significantly non-uniform. The simple yet restrictive convolution model fails to explain this blur and a space varying formulation becomes necessary. Subsequently, we also show how the proposed method can be elegantly modified to account for variations in illumination and pose. We assume a planar structure for the face [14], [17], [19] and use the geometric framework proposed in [8]–[10], [29] to model the blurred face as the weighted average of geometrically warped instances (homographies) of the focused gallery image. The warped instances can be viewed as the intermediate images observed during the exposure time. Each warp is assigned a weight that denotes the fraction of the exposure duration for that transformation. The weights corresponding to the warps are referred to as the transformation spread function (TSF) [29] in the literature.

1.1 Methodology

We develop our basic non-uniform motion blur (NU-MOB)-robust face recognition algorithm based on the TSF model. On each focused gallery image, we apply all the possible transformations that exist in the 6D space (3 dimensions for translations and 3 for rotations) and stack the resulting transformed images as columns of a matrix. We extend the convexity result proved for the simple convolution model in [19] to the TSF model and show that the set of all images obtained by blurring a particular gallery image is a convex set given by the convex hull of the columns of the corresponding matrix. To recognize a blurred probe image, we minimize the distance between the probe and the convex combination of the columns of the transformation matrix corresponding to each gallery image. The gallery image whose distance to the probe is minimum is identified as a match. We do not impose any constraints on the nature of the blur. Following [9], [11], we assume that the camera motion trajectory is sparse in the camera motion space. This allows us to construct an optimization function with l1-norm constraint on the TSF weights. Minimizing this cost function gives us an estimate of the transformations that when applied on the gallery image results in the blurred probe image. Each gallery image, blurred using the corresponding optimal TSF, is compared with the probe in the LBP (local binary pattern) [30] space. This direct method of recognition allows us to circumvent the challenging

and ill-posed problem of single image blind-deblurring. The idea of reblurring followed by LBP-based recognition has been suggested in [19], and LBP histograms have been shown to work well on blurred faces too. We have extended the formulation in [19] to space-varying situations.

Differences With [19]: The DRBF and IRBF methods proposed in [19] are restricted to the simplistic convolution blur model which is valid only when the motion of the camera is limited to in-plane translations. This assumption of uniform blur does not hold true in real settings because camera tilts and rotations occur frequently in the case of hand-held cameras [11]. The algorithms proposed in this paper, in contrast, are capable of handling any general motion of the camera which sets our work distinctly apart from [19].

In addition, we handle pose variations while the discussion in [19] is restricted to frontal faces. Our method allows for arbitrarily-shaped space-varying kernels across the image unlike [19] which seeks to explain the blur using a single PSF for the entire image. In fact, our scheme based on the TSF model subsumes the work in [19] - for the special case of only in-plane translational motion, the TSF reduces to a PSF. The work proposed in this paper advances the state-of-the art in many ways as discussed next.

- This is the first attempt to systematically address face recognition under (i) non-uniform motion blur and (ii) the combined effects of blur, illumination and pose.
- We prove that the set of all images obtained by non-uniformly blurring a given image forms a convex set. We also show that the set of all images obtained from a face image by non-uniform blurring and change of illumination forms a bi-convex set.
- We demonstrate superior performance over contemporary methods on standard face databases (FERET, PIE, Extended Yale B) in the presence of blur, illumination and pose variations, as well as a real dataset which contains, in addition to these degradations, small occlusions and expression changes.

2 EXISTING SYSTEM

2.1 Methods

The task of deblurring an image is image deconvolution; if the blur kernel is not known, then the problem is said to be .blind.. For a survey on the extensive literature in this area, see [Kundur and Hatzinakos 1996]. Existing blind deconvolution methods typically assume that the blur kernel has a simple parametric form, such as a Gaussian or low-frequency Fourier components. However, as illustrated by our examples, the blur kernels induced during camera shake do not have simple forms, and often contain very sharp edges. Similar low-frequency assumptions are typically made for the input image, e.g., applying a quadratic regularization. Such assumptions can prevent high frequencies (such as edges) from appearing in the reconstruction.

2.1.1 Multi-scale approach

The algorithm described in the previous section is subject to local minima, particularly for large blur kernels. Hence, we perform estimation by varying image resolution in a coarse-to-fine manner. At the coarsest level, **K** is a 3_3 kernel. To ensure a correct start to the algorithm, we manually specify the initial 3_3 blur kernel to one of two simple patterns. The initial estimate for the latent gradient image is then produced by running the inference scheme, while holding **K** fixed.

2.1.2 User supervision

Although it would seem more natural to run the multi-scale inference scheme using the full gradient image Ñ**L**, in practice we found the algorithm performed better if a smaller patch, rich in edge structure, was manually selected. The manual selection allows the user to avoid large areas of saturation or uniformity, which can be disruptive or uninformative to the algorithm.

Additionally, the algorithm runs much faster on a small patch than on the entire image. An additional parameter is that of the maximum size of the blur kernel. The size of the blur encountered in images varies widely, from a few pixels up to hundreds. Small blurs are hard to resolve if the algorithm is initialized with a very large kernel. Conversely, large blurs will be cropped if too small a kernel is used. Hence, for operation under all conditions, the approximate size of the kernel is a required input from the user.

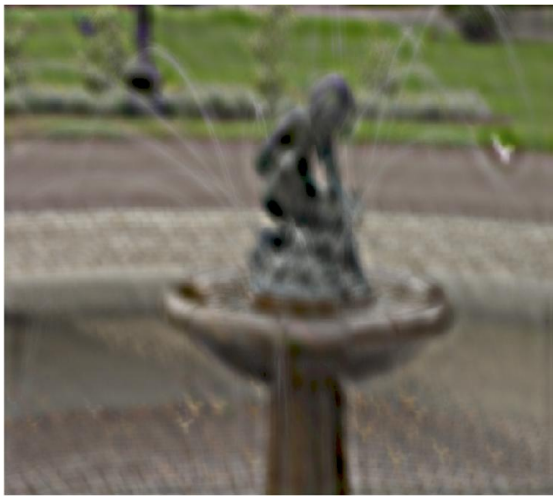

Figure 2.1: Baseline experiments, using Matlab's blind deconvolution algorithm deconvblind on the fountain image

2.2 Image Reconstruction

The multi-scale inference procedure outputs an estimate of the blur kernel **K**, marginalized over all possible image reconstructions. To recover the deblurred image given this estimate of the kernel, we experimented with a variety of non-blind deconvolution methods, including those of Geman [1992], Neelamani [2004] and van Cittert [Zarowin 1994]. While many of these methods perform well in synthetic test examples, our real images exhibit a range of nonlinearities not present in synthetic cases, such as non-Gaussian noise, saturated pixels, residual non-linearities in tonescale and estimation errors in the kernel. Disappointingly, when run on our images, most methods produced unacceptable levels of artifacts. We also used our variational inference scheme on the gradients of the whole image Ñ**B**, while holding **K** fixed. The intensity image was then formed via Poisson image reconstruction [Weiss 2001]. Aside from being slow, the inability to model the non-linearities mentioned above resulted in reconstructions no better than other approaches.

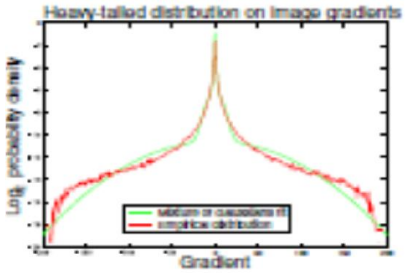

Figure 2.2: The distribution of gradient magnitudes within the scene are shown in red. The y-axis has a logarithmic scale to show the heavy tails of the distribution.

As **L** typically is large, speed considerations make simple methods attractive. Consequently, we reconstruct the latent color image **L** with the Richardson-Lucy (RL) algorithm [Richardson 1972; Lucy 1974]. While the RL performed comparably to the other methods evaluated, it has the advantage of taking only a few minutes, even on large images (other, more complex methods, took hours or days). RL is a non-blind deconvolution algorithm that iteratively maximizes the likelihood function of a Poisson statistics image noise model. One benefit of this over more direct methods is that it gives only non-negative output values. We use Matlab's implementation of the algorithm to estimate **L**, given **K**, treating each color channel independently.

3 RELATED WORK

3.1 Convolution Model For Space-Invariant Blur

As discussed in the introduction, while the convolution model is sufficient for describing blur due to in-plane camera translations, a major limitation is that it cannot describe several other blurring effects

(including out-of-plane motion and in-plane rotation) arising from general camera motion. In order to demonstrate the weakness of the convolution model in handling images blurred due to camera shake, we synthetically blur the focused gallery image to generate a probe, and provide both the gallery image and the blurred probe image as input to two algorithms- the convolution model which assumes spaceinvariant blur, and the non-uniform motion blur model (to be discussed in Section III) which represents the space-variant blurred image as a weighted average of geometrically warped instances of the gallery. Next, we compare the reconstruction errors between the probe and the gallery reblurred using the camera motion estimated by both the methods. Note that there is no change in illumination or pose between the gallery and the probe, and only the blur has to be accounted for. Observe that except for in-plane translations (row 1), where, as expected, the RMS is the same for both the models, in all the other cases, the space-variant motion blur model gives significantly smaller RMS error than the convolution model. Note that the RMS value is smaller than one, except for 6D motion (row 4) for which it is marginally higher as our algorithm needs to search through a very large set of transformations.

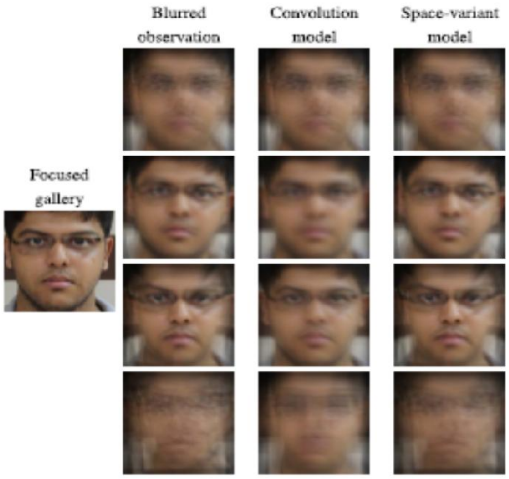

Figure 3.1: Comparing the reconstructed faces and the reconstruction errors (RMS) using the convolution model and the space-variant model. The gallery image is shown in column 1. The results are shown in row 1 for space-invariant blur and rows 2 through 4 for space-variant blur. The RMS errors for the convolution model in column 3 and the space-variant model in column 4, respectively, are- row 1: 0.04, 0.04, row 2: 12.48, 0.05, row 3: 14.88, 0.15, and row 4: 15.31, 5.15. (The blurred image in column 2 was used as the reference for RMS computation.)

3.2 Motion Blur Model For Faces

The apparent motion of scene points in the image will vary at different locations when the camera motion is not restricted to in-plane translations. In such a scenario, the space-varying blur across the image cannot be explained using the convolution model and with a single blur kernel. To get the image coordinates (x, y), the standard practice is to express **x** in homogeneous form i.e., scale **x** by its third element. At each instant of time τ during exposure, the coordinates of the 3D point **X** changes to $\mathbf{X}_\tau = \mathbf{R}_\tau\mathbf{X} + \mathbf{T}_\tau$ due to relative motion between the camera and the subject.

Here, $\mathbf{T}_\tau = [TX_\tau\ TY_\tau\ TZ_\tau]^T$ is the translation vector, and $\mathbf{R}_\tau$ represents the rotation matrix parameterized in terms of the angles of rotation θX, θY and θZ about the three axes using the matrix exponential

$$\mathbf{R}_\tau = e^{\mathbf{\Theta}_\tau} \text{ where } \mathbf{\Theta}_\tau = \begin{bmatrix} 0 & -\theta_{Z_\tau} & \theta_{Y_\tau} \\ \theta_{Z_\tau} & 0 & -\theta_{X_\tau} \\ -\theta_{Y_\tau} & \theta_{X_\tau} & 0 \end{bmatrix} \quad (1)$$

Following prior works in face recognition, [14], [17], [19], we too model the face by a flat surface i.e., all the points are at a distance do from the camera. Therefore, the depth is constant, and the point $\mathbf{x}_\tau$, at which $\mathbf{X}_\tau$ gets projected in the camera, can be obtained through a homography $\mathbf{H}_\tau$ as $\mathbf{x}_\tau = \mathbf{H}_\tau\,\mathbf{x}$ where

$$\mathbf{H}_\tau = \mathbf{K}_\mathbf{v}\left(\mathbf{R}_\tau + \frac{1}{d_o}\mathbf{T}_\tau[0\ 0\ 1]\right)\mathbf{K}_\mathbf{v}^{-1} \quad (2)$$

If g_τ denotes the transformed image (due to relative motion) captured at time instant τ , then we can write $g_\tau(\mathbf{H}_\tau\,\mathbf{x}) = f(\mathbf{x})$, or alternately, $g_\tau(\mathbf{x}) = f(\mathbf{H}_\tau^{-1}\,\mathbf{x})$ where $\mathbf{H}_\tau^{-1}$ denotes the inverse of $\mathbf{H}_\tau$. The

arguments of f in f (**H−1 τ x**), which are the image coordinates, correspond to the first two elements of **H−1**
τ x (a 3 × 1 vector) expressed in homogeneous form.

We follow this convention throughout the paper. Now the blurred face g can be interpreted as the average of transformed versions of f during exposure. Therefore, the intensity at an image point **x** on the blurred face is given by

$$g(\mathbf{x}) = \frac{1}{T_e}\int_0^{T_e} f(\mathbf{H}_\tau^{-1}\mathbf{x})d\tau \qquad (3)$$

where Te is the total exposure duration.

The blurred face can be more appropriately modeled in terms of the gallery face by averaging it over the set of possible transformations resulting from the relative motion between the camera and the subject. Let **T** denote this set of all possible transformations.

The blurred image can be equivalently written as an average of warped versions of f weighted by the TSF, hT , i.e.,

$$g(\mathbf{x}) = \int_{\lambda\in\mathbf{T}} h_T(\lambda) f(\mathbf{H}_\lambda^{-1}\mathbf{x})d\lambda \qquad (4)$$

Observe that a single TSF using (4) is sufficient to describe the observed space-varying blur. When the motion is confined to 2D translations, the TSF will have non-zero weights only for the in-plane translational components and will be identical to the PSF i.e., the convolution model for the blur is a special
case of the space-varying motion blur model.

In practice, the TSF is defined on the discrete transformation space **T** and can be considered as a vector in RNT where NT is the total number of transformations present in **T**. NT is controlled by the number of translation steps along each axis as well as the number of rotation steps about each axis. Hence,

$$g(r,c) = \sum_{k=1}^{N_T} h_T(\lambda_\mathbf{k}) f(\mathbf{H}_{\lambda_\mathbf{k}}^{-1}[r\ c\ 1]^T)) \qquad (5)$$

where g(r, c) and f (r, c) represent the pixel intensity at (r, c) for the discretized blurred image and latent image, respectively. If **g**, **f** represent the blurred image and the latent image, respectively, lexicographically ordered as vectors, then (5) can be expressed in matrix-vector notation as

$$\mathbf{g} = \mathbf{A}\mathbf{h}_\mathbf{T} \text{ such that } \mathbf{h}_\mathbf{T} \geq 0, ||\mathbf{h}_\mathbf{T}||_1 = 1. \qquad (6)$$

where **A** □ RN×NT is the matrix whose NT columns contain warped copies of **f**, **hT** denotes the vector of weights hT (**λ**), and N is the total number of pixels in the image. The warped versions of **f** are obtained by applying the homography matrix **H−1 λ** corresponding to each of the NT transformations on the gallery image. From (6), the set of all blurred images obtained from **f** can be written as

$$\mathcal{B} \triangleq \{\mathbf{A}\mathbf{h}_\mathbf{T} | \mathbf{h}_\mathbf{T} \geq 0, ||\mathbf{h}_\mathbf{T}||_1 = 1\} \qquad (7)$$

Proposition 1: The set of all images **B** obtained by blurring an image **f** using the TSF model is a convex set. Moreover, this convex set is given by the convex hull of the columns of the matrix **A**, where the columns of **A** are warped versions of **f** as determined by the TSF.

Note that **hT** is sparse since for motion blur only a fraction of the total poses NT will have non-zero weights [29]. We make use of this fact to build the following energy function

$$E(\mathbf{h}_\mathbf{T}) = ||\mathbf{g} - \mathbf{A}\mathbf{h}_\mathbf{T}||_2^2 + \beta||\mathbf{h}_\mathbf{T}||_1$$
$$\text{subject to } \mathbf{h}_\mathbf{T} \geq 0. \qquad (10)$$

The optimization problem in (10) can be solved using the nnLeastR function of the Lasso algorithm [31] which considers the additional l1-norm and non-negativity constraints. This energy function when minimized provides an estimate of the transformations that must be applied to the gallery image to
produce the blurred image.

3.2.1 Multiscale Implementation

Since we are fundamentally limited by the resolution of the images, having a very fine discretization of the transformation space **T** leads to redundant computations. Hence, in practice, the discretization is performed in a manner that the difference in the displacements of a point light source due to two different transformations from the discrete set **T** is at least one pixel.

3.2.2 Face Recognition Across Blur

Suppose we have M face classes with one focused gallery face **fm** for each class m, where m = 1, 2, . . . , M. Let us denote the blurred probe image which belongs to one of the M classes by **g**. Given **fm**s and **g**, the task is to find the identity m□ □ {1, 2, . . , M} of **g**. Based on the discussions in Section III, the first step is to generate the matrix **Am** for each gallery face. Then, since **g** belongs to one of the M classes, it can be expressed as the convex combination of the columns of one of these matrices. Therefore, the identity of the probe image can be found by minimizing the projection error of **g** onto {**Am**}s. The reconstruction error dm can be obtained by solving

$$d_m = \min_{\mathbf{h_T}} ||\mathbf{g} - \mathbf{A_m h_T}||_2^2 + \beta ||\mathbf{h_T}||_1$$
$$\text{subject to} \quad \mathbf{h_T} \geq 0. \qquad (11)$$

One could compute dm for each m = 1, 2, . . . , M and assign **g** the identity of the gallery image with the minimum dm. Note that in (11), all the pixels receive equal weight and influence the TSF estimation step equally. But not all regions in the face convey the same amount of information. Following [19], we modify the above equation by introducing a weighting matrix **W** (which weighs different regions in the face differently) when computing the reconstruction error between the probe image and the gallery images. Equation (11) then becomes

$$d_m = \min_{\mathbf{h_T}} ||\mathbf{W}(\mathbf{g} - \mathbf{A_m h_T})||_2^2 + \beta ||\mathbf{h_T}||_1$$
$$\text{subject to} \quad \mathbf{h_T} \geq 0 \qquad (12)$$

where **W** (a diagonal matrix) is learned following the procedure outlined in the appendix of [19]. This matrix has the highest weights for regions around the eyes and de-emphazises the mouth and cheeks.

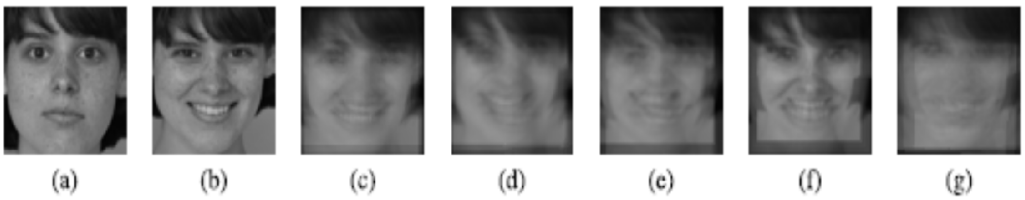

Figure 3.2: Sample images from ba and bj folders in the FERET database. (a) Gallery, (b) probe, (c)-(g) probe blurred synthetically using random transformations from the TSF intervals listed in Setting 1 - Setting 5

The steps are outlined in Algorithm 1. An alternative approach would be to use the optimal TSFs **hTm** to perform a non-blind deblurring of the probe. However, we found that, deblurring artifacts introduced in this process tend to significantly reduce the recognition accuracy (by almost 15% to 20%) which suggests that reblurring the gallery is preferable to deblurring the probe.

3.3 Algorithm Used

Algorithm 1 NU-MOB: Non-Uniform Motion Blur-Robust Face Recognition

Input : Blurred probe image g and a set of gallery images fm; m=1,2,….,M.

Output: Identity of the probe image.

1. For each gallery image fm, find the optimal TSF h_{Tm} by solving equation.
2. Blur each gallery image fm with its corresponding h_{Tm} and extract LBP features.
3. Compare the LBP features of the probe image g with those of the transformed gallery images and find the closest match.

We evaluate the proposed algorithm NU-MOB on the standard and publicly available FERET database [32]. Since this database contains only focused images, we blur the images synthetically to generate the probes. The camera motion itself is synthesized so as to yield a connected path in the motion space. The resulting blur induced mimics the real blur encountered in practical situations. In all the experiments presented in this paper, we use grayscale images resized to 64 × 64 pixels and we assume only one image per subject in the gallery. To evaluate our algorithm on different types and amounts of blur, we synthetically blur the face images with the TSF model using the following five blur settings: Setting 1 (S1): in-plane translations, Setting 2 (S2): in-plane translations and rotations, Setting 3 (S3): out-of-plane rotations, Setting 4 (S4): out-of-plane translations, and Setting 5 (S5): full 6D blur.

We also compare using a two-step non-uniform deblur [9] + recognize approach in S.Nos. 7 and 8 since neither SRC nor DFR can cope with blur. Yet another two-step baseline comparison that uses LBP features extracted from the deblurred probe for recognition is provided in S.No. 9. Recognition scores were computed for various blur kernel sizes ranging from 3 to 13 pixels for the DRBF [19] algorithm. However, we would like to add that the authors in [19] have, on their data, reported recognition rates that are, on an average, 3 to 4 percent greater using their rDRBF algorithm. For comparison with [17] for the space-varying cases in Settings 2 to 5, following the discussion in Section 4.1.2 of their paper, we divided the image into overlapping patches with sizes 75, 50 and 40 percent of the original image, performed recognition separately on each patch and used a majority vote to calculate the final recognition score. This was repeated for various blur kernel sizes ranging from 3 to 13 pixels, and the best recognition rates have been reported. In our implementation of the FADEIN algorithm, the statistical models for PSF inference were learned from 25 PSFs which included 24 motion blur kernels (length = 3, 5, 7, 9, 11, 13 pixels, and angle = 0, 0.25π, 0.5π, 0.75π) and one 'no blur' delta function.

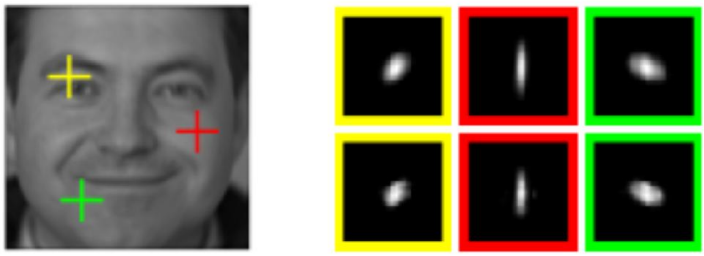

Figure 3.3: (a) Blurred probe, (b) row 1 - true PSFs at three locations (marked by crosshairs in (a)), row 2 - PSFs estimated by our NU-MOB algorithm.

The true kernels at three locations marked by crosshairs on Fig. 3.3(a) are shown in the first row of Fig. 3.3(b), while the kernels obtained from the TSF estimated by NU-MOB are shown in the second. Observe that the estimated PSFs closely resemble the true PSFs which indicates that the motion has been computed correctly.

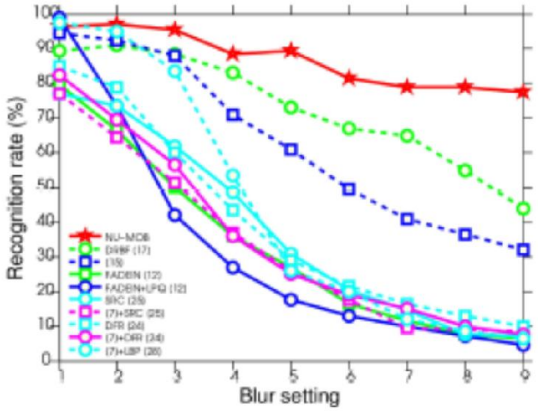

Figure 3.4: Effect of increasing the blur. (Refer to the text for blur settings along the X-axis.)

1) Effect of Increasing the Blur: We now examine our algorithm's performance as the extent of the blur is increased. The gallery, as before, is the ba folder. We select random transformations from the following nine sets of intervals to blur the images in the bj folder and generate the probes.
2) Effect of Underestimating or Overestimating the TSF Search Intervals: In all the above experiments, we have assumed that the TSF limits are known, and we used the same transformation intervals as the ones used for synthesizing the blur, while attempting recognition.

As before, the ba folder of FERET is chosen as the gallery and the probe images are generated by blurring the bj folder using random transformations lying in the intervals- in-plane translations $[-2 : 1 : 2]$ pixels, and in-plane rotations $=[-1° : 1° : 1°]$.

We note that the recognition rates are fairly stable even when the TSF search intervals are chosen to be much larger than the true TSF intervals (cases (4)-(7)) i.e., our algorithm picks the correct TSF even when the search space is large.

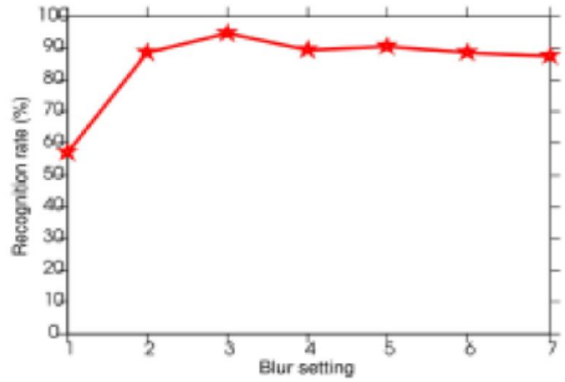

Figure 3.5: Effect of underestimating or overestimating the TSF search intervals. (Refer to the text for blur settings along the X-axis.)

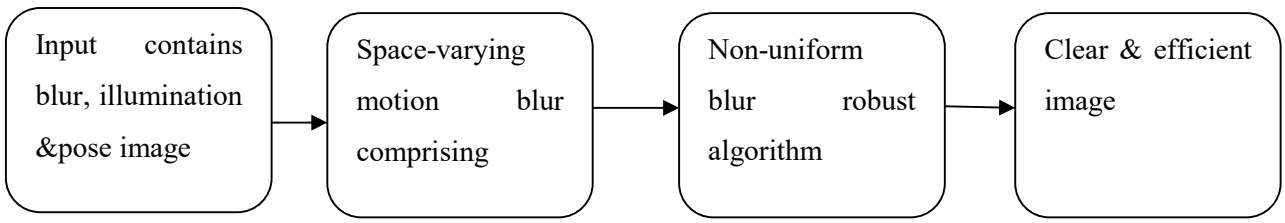

Figure 3.6: Block Diagram

3.4 Face Recognition Across Blur, Illumination, And Pose

Poor illumination is often an accompanying feature in blurred images because larger exposure times are needed to compensate for the lack of light which increases the chances of camera shake. Pose variation is another challenge for realizing the true potential of face recognition systems in practice. This section is devoted to handling the combined effects of blur, illumination and pose.

3.4.1 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 αi , i = 1, 2, . . . , 9 are the corresponding linear coefficients. The fi s, which form a basis for this 9D subspace, can be generated using the Lambertian reflectance model as where ρ and **n** are the

albedo and the surface normal, respectively, at the pixel location (r, c), and **s** is the illumination direction.

We now elaborate on the two steps involved in our AM algorithm. For any gallery image **fm**, in the first iteration, we assume the blur to be an impulse (i.e., no blur) and estimate the nine illumination coefficients αm,i by solving the linear least squares problem **g** = **Lmαm**, where **Lm** is a matrix whose nine columns contain the basis images **fm**,**i** corresponding to the subject m lexicographically ordered as vectors, and **αm** = [αm,1, αm,2, . . . ,αm,9] are its corresponding illumination coefficients. Now, we create a new relit gallery image from the basis images using the estimated illumination coefficients αm,i . This completes the first step of the alternation wherein we fixed the blur and estimated the illumination.

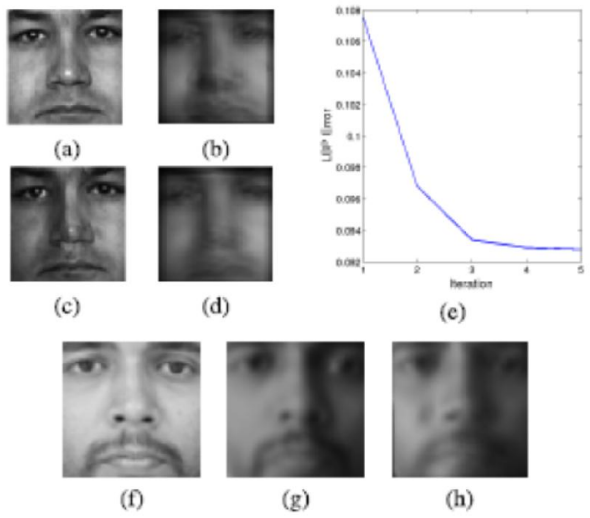

Figure 3.7: An example image from the PIE dataset illustrating the alternating minimization algorithm is shown in the first two rows. (a) Gallery, (b) probe, (c) relit gallery image, (d) reblurred and relit gallery image, and (e) a plot of the LBP cost versus iterations. Another example from our real dataset is shown in row 3. (f) Gallery, (g) probe, (h) reblurred and relit gallery image.

Next, we build **Am** using warped versions of this relit gallery image and estimate the TSF **hTm** by solving equation (13). Observe that the illumination is fixed and we are solving for the blur in this step. Following this, we blur each of the basis images using the estimated TSF **hTm** before proceeding to

the second iteration. Note that the matrix **Lm** in the second and subsequent iterations is built from the blurred basis images.

3.4.2 Handling Pose Variations

Most 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. In our experiments, we found this to be true of our MOBIL algorithm also. 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 MOBILAP algorithm which, using an estimate of the pose, matches the incoming probe with a synthesized non-frontal gallery image. To the best of the authors' knowledge, this is the first ever effort to even attempt this compounded scenario.

Figure 3.8: Example images of a subject from the PIE database under new poses.

The images in (a) and (b) are synthesized from the frontal gallery image using the average face depthmap shown in (c).

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 in Section IV-A. (See Fig. 3.7.) 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 (15) but with ρ now being the new synthesized pose and **n** being the surface normals recomputed from the rotated depthmap. Note that the motion blur model for faces discussed in Section III applies even in the event of a pose change. An overview of the algorithm is presented in Algorithm 3.

3.4.5 Recognition Across Blur, Illumination, and Pose

Finally, we take up the very challenging case of allowing for pose variations in addition to blur and illumination. We once again use the PIE dataset. We begin by selecting four near-frontal poses (pitch and yaw angles within 15°) and explore the robustness of MOBIL itself to small variations in pose. As before, the camera position c27 (frontal pose) and flash position f11 (frontal illumination) constitute the gallery. In this experiment, however, the probe set, divided into good and bad illumination subsets, contains the four nearfrontal poses c05 (−16° yaw), c07 (0° yaw and −13° tilt), c09 (0° yaw and 13° tilt) and c29 (17° yaw). Note that both camera poses and flash positions are varied. Example images are shown in Fig. 3.8 columns 1 to 4. The recognition results for GI and BI, presented respectively, are clear indicators of MOBIL's robustness to small variations in pose. However, we note that there is some drop in our algorithm's recognition accuracy in going from GI to BI due to the combined effects of blur, illumination and pose.

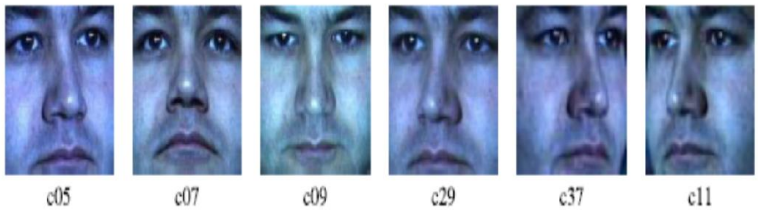

Figure 3.9: PIE database. Different camera poses under frontal illumination (f11).

Next, we select differently illuminated probes in two non-frontal poses c37 (−31° yaw) and c11 (32° yaw). See Fig. 3.8 columns 5 and 6. Once again, the frontal camera position c27 and flash position

f11 constitute the gallery. For such large changes in pose, we found that MOBIL returned recognition rates less than 15%.

3.4.6 Real Datasets

The AUC values obtained by MOBILAP along with other approaches that follow the Unsupervised protocol. The ROC curves are also shown in the plot of Fig 3.10 in order to better evaluate the performance. Note that MOBILAP ranks close to MRF-MLBP [40] which is next only to PAF [42]. LFW contains significant amounts of occlusion, expression and age variations (see Fig. 3.9(c)-(d)) which we do not model in our approach, whereas competing methods handle one or more of these effects explicitly.

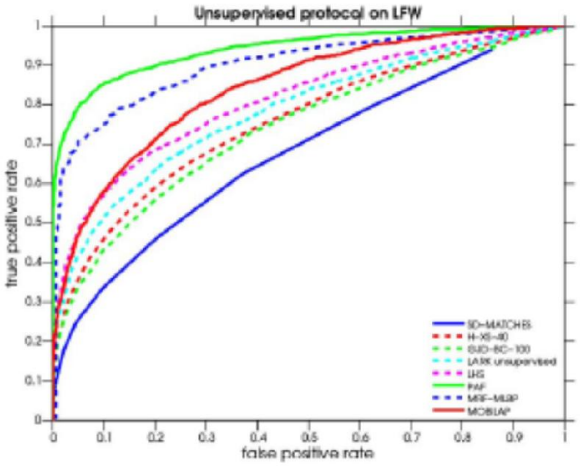

Figure 3.10: ROC curves of different approaches on the LFW dataset for the Unsupervised protocol.

Finally, we report recognition results on a real dataset that contains face images that we ourselves captured in unconstrained settings. There are 50 subjects in the dataset. The gallery contains one frontal, sharp and well-illuminated image taken outdoor under diffuse lighting. The probe images, 2,200 in number, were captured using a hand-held camera under indoor and outdoor lighting conditions.

Figure 3.11: Cropped faces of four subjects from the real dataset. The gallery images are shown in column one. The probe images have variations in illumination, facial expressions changes (column four in row one), small occlusions (missing spectacles - column four, row two) and differences in pose (columns two through four in row three).

4 CONCLUSIONS

We proposed a methodology to perform face recognition under the combined effects of non-uniform blur, illumination, and pose. We showed that the set of all images obtained by non-uniformly blurring a given image using the TSF model is a convex set given by the convex hull of warped versions

of the image. Capitalizing on this result, we initially proposed a non-uniform motion blur-robust face recognition algorithm NU-MOB. We then showed that the set of all images obtained from a given image by non-uniform blurring and changes in illumination forms a bi-convex set, and used this result to develop our 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. Extensive experiments were given on synthetic as well as real face data. The limitation of our approach is that significant occlusions and large changes in facial expressions cannot be handled.

Future Work

Future research will include methods like 3D reconstruction to generate synthetic faces based on variations in pose and expression of a target individual. An algorithm for image morphing allows to generate a set of diverse images with a smooth transition of illumination

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[2] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in Proc. 19th Int. Conf. Pattern Recognit., Dec. 2008, pp. 1–4.

[3] S. Biswas, G. Aggarwal, and R. Chellappa, "Robust estimation of albedo for illumination-invariant matching and shape recovery," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 5, pp. 884–899, May 2009.

[4] S. Cho, Y. Matsushita, and S. Lee, "Removing non-uniform motion blur from images," in Proc. Int. Conf. Comput. Vis., Oct. 2007, pp. 1–8.

[5] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," ACM Trans. Graph., vol. 25, no. 3, pp. 787–794, Jul. 2006.

[6] R. Gopalan, S. Taheri, P. Turaga, and R. Chellappa, "A blur-robust descriptor with applications to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 6, pp. 1220–1226, Jun. 2012.

[7] Gupta, N. Joshi, L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," in Proc. Eur. Conf. Comput. Vis., 2010, pp. 171–184.

[8] H. Hu and G. de Haan, "Adaptive image restoration based on local robust blur estimation," in Proc. 9th Int. Conf. Adv. Concepts Intell. Vis. Syst., 2007, pp. 461–472.

[9] Z. Hu and M.-H. Yang, "Fast non-uniform deblurring using constrained camera pose subspace," in Proc. Brit. Mach. Vis. Conf., 2012, pp. 1–11.

[10] H. Ji and K. Wang, "A two-stage approach to blind spatially-varying motion deblurring," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 73–80.

[11] K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 5, pp. 684–698, May 2005.

[12] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 12, pp. 2354–2367, Dec. 2011.

[13] M. Nishiyama, A. Hadid, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi, "Facial deblur inference using subspace analysis for recognition of blurred faces," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 4, pp. 838–845, Apr. 2011.