Research Article Vol. 33, No. 9 / September 2016 / Journal of the Optical Society of America A 1887

Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

Recognizing blurred, nonfrontal, illumination, and expression variant partially occluded faces

ABHIJITH PUNNAPPURATH* AND AMBASAMUDRAM NARAYANAN RAJAGOPALAN

Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai 600036, India
**Corresponding author: jithuthatswho@gmail.com*

Received 25 April 2016; revised 26 June 2016; accepted 2 August 2016; posted 11 August 2016 (Doc. ID 259294); published 31 August 2016

The focus of this paper is on the problem of recognizing faces across space-varying motion blur, changes in pose, illumination, and expression, as well as partial occlusion, when only a single image per subject is available in the gallery. We show how the blur, incurred due to relative motion between the camera and the subject during exposure, can be estimated from the alpha matte of pixels that straddle the boundary between the face and the background. We also devise a strategy to automatically generate the trimap required for matte estimation. Having computed the motion via the matte of the probe, we account for pose variations by synthesizing from the intensity image of the frontal gallery a face image that matches the pose of the probe. To handle illumination, expression variations, and partial occlusion, we model the probe as a linear combination of nine blurred illumination basis images in the synthesized nonfrontal pose, plus a sparse occlusion. We also advocate a recognition metric that capitalizes on the sparsity of the occluded pixels. The performance of our method is extensively validated on synthetic as well as real face data.

OCIS codes: (100.0100) Image processing; (100.5010) Pattern recognition; (100.3008) Image recognition, algorithms and filters; (150.0150) Machine vision.

http://dx.doi.org/10.1364/JOSAA.33.001887

1. INTRODUCTION

State-of-the-art face recognition (FR) systems can outperform even humans when presented with images captured under *controlled* environments. However, their performance drops quite rapidly in unconstrained settings due to image degradations arising from blur, variations in pose, illumination, and expression, partial occlusion, etc. Motion blur is commonplace today owing to the exponential rise in the use and popularity of lightweight and cheap handheld imaging devices and the ubiquity of mobile phones equipped with cameras. Photographs captured using a handheld device usually contain blur when the illumination is poor because larger exposure times are needed to compensate for the lack of light, which increases the possibility of camera shake. On the other hand, reducing the shutter speed results in noisy images while tripods inevitably restrict mobility. Even for a well-lit scene, the face might be blurred if the subject is in motion. The problem is further compounded in the case of poorly lit dynamic scenes because the blur observed on the face is due to the combined effects of the blur induced by the motion of the camera and the independent motion of the subject. In addition to blur and illumination, practical face recognition algorithms also must possess the ability to recognize faces across reasonable variations in pose. Partial occlusion and facial expression changes, common in real-world applications, further escalate the challenges. Yet another factor that governs the performance of face recognition algorithms is the number of images per subject available for training. In many practical application scenarios such as law enforcement, driver's license, or passport identification, where there is usually only one training sample per subject in the database, techniques that rely on the size and representation of the training set suffer a serious performance drop or even fail to work. Face recognition algorithms can broadly be classified into either discriminative or generative approaches. While the availability of large labeled data sets and greater computing power recently has boosted the performance of discriminative methods [1,2], generative approaches continue to remain popular [3,4], and there is concurrent research in both directions. The model we present in this paper falls into the latter category. In fact, generative models are even useful for producing training samples for learning algorithms.

Literature on face recognition from blurred images can be broadly classified into four categories. It is important to note that all of them (except our own earlier work in [4]) are restricted to the convolution model for uniform blur. In the first approach [5,6], the blurred probe image is first deblurred using standard deconvolution algorithms before performing recognition. However, deblurring artifacts tend to significantly lower recognition accuracy for moderate to heavy blur. An exemplar-based deblurring algorithm specifically designed for face images was proposed in [7]. But their method, too, is limited to space-invariant blur. Moreover, they do not explicitly address the task of face recognition. The second approach performs joint deblurring and recognition [8]. However, this is computationally expensive. The third category is based on extracting blur-invariant features and using them for recognition [9,10], but these are effective only for small blurs. Finally, the fourth and more recent trend is to attempt direct recognition [3,4] by comparing reblurred versions from the gallery with the blurred probe image in the local binary pattern (LBP) [11] space. This is the strategy that we also employ in this work.

The seemingly unrelated area of image matting has been employed for face segmentation [12], face and gait recognition [13], image deblurring [14], etc. The idea of using the transparency map or the alpha matte to estimate space-invariant blur was first mooted in [14]. Transparency, in the context of motion blur, is induced by the movement of the camera/object during image capture. The sharp boundary of an opaque foreground object gets smeared against the background due to motion. For object motion, the fractional matte at a pixel can be physically interpreted as the fraction of the total exposure duration during which the foreground object was imaged. This argument can be extended to camera shake because the motion results in the mixing of foreground and background colors at the boundaries of the foreground object. Thus, the alpha matte provides a robust and simple model to explain the effect of motion blur at the boundaries of the foreground object, which in our case is the face.

Pose variations are a major bottleneck for most recognition algorithms. According to the survey paper by [15] on pose, methods for FR across pose can be broadly classified into 2D and 3D techniques. A recent survey paper by [16] tracks the developments in pose-invariant face recognition in the past six years after the survey by [15]. There have mainly been two kinds of pursuits for handling illumination in face recognition. The first is based on extracting illumination-insensitive features from the face image and using them for matching [17,18]. The second is based on the linear subspace model of [19], which states that each face can be characterized by a nine-dimensional subspace. A face recognition system has to be robust to occlusion, too, in order to guarantee reliable real-world operation. The traditional approach while dealing with occlusions or large expression changes is to discard the occluded pixels during the matching step [20,21]. In contrast, methods based on sparse representation [22,23] model the occluded face image as a combination of the unoccluded face plus the occlusion and seek the sparse representation jointly over a training sample dictionary and an occlusion dictionary.

Although it is quite challenging to perform recognition even when one of these degradations—blur, pose, illumination, or occlusion—is present, a few attempts have been made to jointly address some of these issues under one roof. A sparse minimization technique for recognizing faces across illumination and occlusion was proposed in [22]. But this method requires multiple images of the same subject for training. A dictionary-based approach to recognizing faces across illumination and pose has been proposed by [24]. But neither of these works deal with blurred images. The role of sparse representation and dictionary learning in face recognition has been reviewed in [25]. The problem of recognizing faces across blur and illumination has been formally addressed by [3]. A recent work [26] presents a domain adaptive solution for face recognition across blur, illumination, and 2D registration. But the formulation in [3] and [26] is based on the restrictive convolution model for uniform blur. The formulation does not address the more challenging and practically common scenario of space-varying blur. The problem of recognizing faces across nonuniform blur was first addressed by [10]. The authors applied the uniform blur model on overlapping patches and performed recognition based on a majority vote. However, their method did not explicitly model illumination changes between the gallery and the probe images. Moreover, [3,10,26] limit their discussion to frontal faces.

The focus of this paper is on developing a system that can recognize faces across nonuniform (i.e., space-varying) blur (due to relative motion between the camera and the face), varying pose, illumination, and expression as well as partial occlusion when only a single image per subject is available in the gallery. The gallery images are assumed to be sharp, frontal, well-illuminated, unoccluded, and captured under a neutral expression. The motion blur in the probe can be a result of object motion and incidental camera shake. We do not assume any parametric or special form for the blur but show how our method generalizes to camera and/or object motion. However, we assume that the camera trajectory is sparse in the camera motion space [27]. We demonstrate how the use of the matte instead of the intensity image, to estimate the motion, simplifies our model and allows for accurate blur estimation. Matting algorithms, however, require a trimap (a pre-segmented image consisting of three regions, namely, sure foreground, sure background, and unknown region) as input from the user. To avoid the need for user interaction, we even develop a method to *automatically* generate the trimap. Having computed the motion from the matte, we model the other degradations using the intensity images. We show how an estimate of the pose of the probe can be used to synthesize nonfrontal gallery images that match the probe's pose, and perform pose-invariant recognition. The probe lighting itself can be uncontrolled. To handle illumination changes, we approximate the face to a convex Lambertian surface and use the nine-dimensional subspace model of [19]. We solve for the illumination coefficients by modeling the blurred and differently lit nonfrontal probe as a linear combination of nine blurred basis images in the synthesized nonfrontal pose. Our final modification to the proposed framework aims at explicitly accounting for partial occlusion and expression changes based on the *a priori* knowledge that these changes affect only a sparse number of pixels. This is achieved by appending an occlusion vector and jointly solving for the illumination and occlusion components. To perform recognition, we select a few potential matches from the gallery by examining the sparsity of the estimated occlusion vectors, transform the images from this selected set to match the blur, pose, and illumination of the probe and match the probe with the transformed gallery in the LBP [11] space to determine the closest match.

Differences with our earlier work in [4]**:** We had focused on the problem of recognizing faces across nonuniform motion blur, illumination, and pose in our recent work [4]. The alternating minimization (AM) scheme for jointly handling nonuniform blur and illumination can only guarantee local minima. This poses problems when the probe lighting is poor and/or if the blur increases beyond a certain extent because the blur and illumination coefficients will be incorrectly estimated. The framework proposed in this paper does not suffer from this drawback because the blur is directly estimated from the matte (and not from an intensity image pair as in [4]) independent of illumination and other degradations. Furthermore, occlusions and changes in facial expressions can throw the AM framework of [4] completely off course. In comparison, the technique proposed in this paper explicitly models occlusion.

To summarize, the main contributions of this paper are as follows:

- This is the first attempt of its kind to *systematically* address face recognition under the combined effects of nonuniform blur, pose, illumination, and occlusion under the very challenging scenario of a single image per subject in the gallery.
- We propose a methodology that judiciously harnesses the alpha matte for inferring the relative motion between the camera and the face. We successfully employ the projective motion blur model on mattes for general motion estimation.
- We also present a pipeline that requires absolutely no user interaction for generating the trimap paving the way for a fully automated face recognition system.
- We show how the probe can be modeled as a linear combination of nine blurred basis images in the synthesized nonfrontal pose, plus a sparse occlusion. We also show how the sparsity of the estimated occlusion aids the recognition task.
- We demonstrate superior performance over state-of-the-art methods using publicly available face databases as well as on a data set we ourselves captured, which contains significant amounts of blur, variations in pose, lighting, and expression and partial occlusion.

The organization of the rest of the paper is as follows: we first explore the advantages of estimating motion from the matte in Section 2. We also examine the projective motion blur model from the perspective of transparency maps. In Section 3, we provide detailed and systematic analyses of how we model and estimate blur, pose, illumination, and occlusion. We also discuss how to perform recognition using these estimated values. In Section 4, our framework is used to perform face recognition on standard publicly available data sets and also on our own data set. Section 5 concludes the paper.

2. MOTION FROM MATTE

In this section, we first discuss the advantages of using the matte instead of the intensity image to estimate the blur. We then review the nonuniform motion blur model or the projective motion blur model but from a new perspective involving the transparency map. Unlike in [14], where the blur is assumed to be space-invariant, our framework can handle even nonuniform blur arising from general motion of the camera.

It is natural to ask why the intensity image itself cannot be provided as input to standard nonuniform blind deblurring techniques to recover the motion. Face images have less texture than natural images, and existing deblurring methods do not perform well on faces [7]. This is because the success of these deblurring algorithms hinges on implicit or explicit extraction of salient edges for kernel estimation, and, for blurred images with less texture, the edge prediction step is less likely to provide robust results. The transparency map, on the other hand, is not related to the underlying complex image structure or features of the face because all alpha values on the face (excluding the pixels on the boundary) are equal to 1. While an intensity-based blur estimation technique would typically make use of *all pixels* in the image, a matte-based method only requires the fractional *boundary pixels* because they neatly encode the motion information. These properties of the transparency map and the advantages of estimating blur via the matte (instead of the intensity image) are illustrated through the following experiment. We selected 10 different face images without any blur and generated 20 random motion blur kernels. Next, we synthesized a set of 200 uniformly blurred observations by convolving each of these 10 images with the 20 kernels. In the case of camera shake, the blur kernel or the point spread function (PSF) reveals how a point light source spreads under the effect of the camera motion. Note that the PSF is the same for the entire image (as in this experiment) under the convolution model for space-invariant blur. To quantitatively demonstrate the advantages of transparency maps over intensity images for kernel estimation, we extracted the mattes of the foreground face region from each of the 200 blurred observations. Three representative examples from this set are shown in Fig. 1. In each of the three columns, the first row contains a pair of images with the one on the left being the blurred observation, while the one on the right is its corresponding matte. We then estimated the PSF from the intensity image using the state-of-the-art blind deconvolution algorithm in [28] (only the cropped face region was provided as input to [28] because some of the intensity images had saturated pixels in the background) and also from the matte using the approach in [14] because blur is space-invariant. We compared both the PSF estimated from the intensity image and the PSF from the matte with the ground truth kernel using normalized cross correlation [29], which is a standard metric for kernel similarity. We obtained a cross-correlation value of 0.601 averaged over all 200 images for the kernels estimated using the intensity images. The same measure calculated from the PSFs computed using the transparency maps was 0.849. Row two of Fig. 1 shows the ground truth kernels, and the PSFs estimated using the intensity images and the transparency maps, respectively, for the three examples in row one. It can be observed that, even visually, the shapes of the matte-based kernels closely resemble the ground truth PSFs.

A. Projective Motion Blur Model for Transparency Maps

The convolution blur model or the uniform blur model is valid only when the motion of the camera is limited to in-plane translations. However, tilts and rotations occur frequently in the case of handheld cameras [30] resulting in blur that is significantly nonuniform across the face. The need for a space-varying blur model for faces has already been elaborately discussed in [4]. The space-varying model or the projective

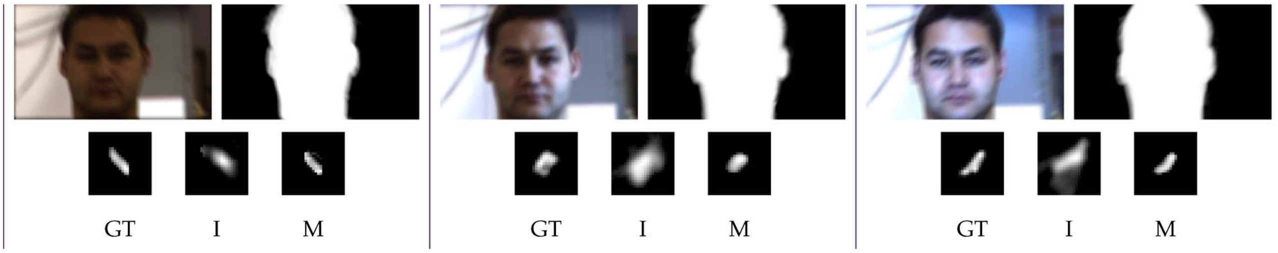

Fig. 1. Synthetic experiment demonstrating the superiority of mattes over intensity images for blur kernel estimation. GT = ground truth kernel, I = PSF estimated from the intensity image, M = PSF estimated from the matte.

model [27,30,31] assumes that the blurred image is the weighted average of warped instances of the underlying focused image. While previous works were based on intensity images, we propose to apply the projective motion blur model to transparency maps for the reasons discussed in the first part of this section. Following other works in face recognition that handle blur [3,6,10], we, too, model the face as planar. The matte $\boldsymbol{\alpha_b}$ extracted from the blurred probe can be represented as

$$\boldsymbol{\alpha_b} = \sum_{k\in\mathbf{S}} \omega_k \alpha_{l_k}, \tag{1}$$

where $\boldsymbol{\alpha_l}$ is the latent unblurred matte of the probe, and $\boldsymbol{\alpha_{l_k}}$ is $\boldsymbol{\alpha_l}$ warped by the homography $\mathcal{H}_{\boldsymbol{k}}$. Each scalar ω_k denotes the fraction of the total exposure duration for which the camera stayed in the position that caused the transformation $\mathcal{H}_{\boldsymbol{k}}$. Akin to a PSF, $\sum_{k\in\mathbf{S}}\omega_k = 1$ and $\omega_k \geq 0$. The parameter ω_k depicts the motion, and it is defined on the discrete transformation space $\mathbf{S}$, which is the finite set of sampled camera poses.

The homography $\mathcal{H}_{\boldsymbol{k}}$ corresponding to $\boldsymbol{\alpha_{l_k}}$ in Eq. (1) in terms of the camera parameters is given by

$$\mathcal{H}_{\boldsymbol{k}} = \mathbf{K_v}\left(\mathbf{R_k} + \frac{1}{d_0}\mathbf{T_k}[0\ 0\ 1]\right)\mathbf{K_v^{-1}}, \tag{2}$$

where $\mathbf{R_k}$ is a rotation matrix [27] parameterized in terms of θ_X, θ_Y, and θ_Z, which are the angles of rotation about the three axes. $\mathbf{T_k} = [\,T_{x_k} \quad T_{y_k} \quad T_{z_k}\,]^T$ is the translation vector, and d_0 is the scene depth. The camera intrinsic matrix $\mathbf{K_v}$ is assumed to be of the form $\mathbf{K_v} = \mathrm{diag}(v, v, 1)$, where v is the focal length. Six degrees of freedom arise from $\mathbf{T_k}$ and $\mathbf{R_k}$ (three each). In this discussion, we assume that v is either known or can be extracted from the image's EXIF tags, and the weights ω_k are what need to be estimated.

3. MODELING THE DEGRADATIONS

In this section, we formally introduce our approach for estimating the space-varying blur across the face using the alpha matte extracted from the probe. We also present an automated method for trimap generation that requires absolutely no user intervention. After computing the motion from the matte, we return to the intensity images to model the other degradations. We discuss in detail how each of these variations—pose, illumination, partial occlusion, and expression—are modeled. Finally, we show how recognition can be performed based on the extent of the estimated occlusion and LBP matching between the probe and the transformed gallery images.

A. Blur

Our objective is to calculate the nonuniform motion blur from the matte of the probe. The weights ω_k, which encapsulate this global motion information, can be computed from a few locally estimated PSFs. The PSFs themselves are determined from small patches lying on the boundary of the probe matte $\boldsymbol{\alpha_b}$. The assumption is that the blur is uniform within each patch (we used small patches of size 51×51 pixels for all our experiments), although it can be space-varying across the image. As few as four PSFs are sufficient for the accurate estimation of ω_k provided the locations of their corresponding patches are spatially spread out across the image. To automatically identify four such patches on the matte boundary, we first draw a rough bounding box around the face region using the landmark points detected by the method in [32] [see Fig. 2(a)]. The box is also divided into four quadrants. This region is then isolated from the matte extracted from the probe, as shown in Fig. 2(b). Because the matte entries are mostly 0 and 1 with fractional values lying only at the transition from background to foreground, the boundaries of the transparency map can easily be identified by a simple column/row sum operation, and four patches [each lying in one of the four quadrants in Fig. 2(a)] are selected, as illustrated in Fig. 2(b). Next, the PSFs corresponding to these patches are estimated, and the estimated kernels are stacked as a vector $\mathbf{h}$. Then the relationship between the PSFs and $\boldsymbol{\omega}$ (which denotes the vector of weights ω_k) is given in [33] as

$$\mathbf{h} = \boldsymbol{\Lambda\omega}. \tag{3}$$

Here $\boldsymbol{\Lambda}$ is a matrix whose entries are determined by the location of the blur kernels and the bilinear interpolation coefficients [33]. Note that $\boldsymbol{\omega}$ is a sparse vector because the blur is

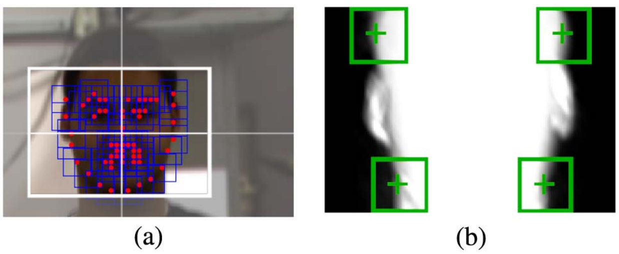

Fig. 2. (a) Rough bounding box drawn around the face region using the landmark points detected by the method in [32]. (b) Four spatially separated patches selected on the matte boundary.

typically due to incidental camera shake, and only a small fraction of the poses in $\mathbf{S}$ will have non-zero weights in $\boldsymbol{\omega}$ [27].

The optimal sparse $\tilde{\boldsymbol{\omega}}$ is then computed by minimizing the following cost [33]:

$$\tilde{\boldsymbol{\omega}} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \|\mathbf{h} - \boldsymbol{\Lambda}\boldsymbol{\omega}\|_2^2 + \beta_1 \|\boldsymbol{\omega}\|_1 \text{subject to } \boldsymbol{\omega} \geq \mathbf{0}. \quad (4)$$

The optimization problem in Eq. (4) can be solved using the *nnLeastR* function of the Lasso algorithm in [34], which considers the l_1 norm and non-negativity constraints. While computing $\tilde{\boldsymbol{\omega}}$, which encodes the space-varying motion directly using the entire matte $\boldsymbol{\alpha_b}$, would typically require a multiscale pyramidal scheme similar to [27], the approach we adopt of estimating it via PSFs computed from patches is simpler and more robust [33]. We also would like to point out that the authors of [33] estimate motion from intensity images while we apply their technique to transparency maps.

Unlike [4], our alpha matte framework allows us to solve for the blur independent of other degradations. To highlight the advantages of this approach over the AM framework of [4], consider the example in Fig. 3. Observe that the probe in column two is well-lit, while the illumination is poor for the probe in column four. Also note that the probe in column two, though well-illuminated, is heavily blurred. It can be seen that the PSFs estimated via the matte (mattes extracted from the probes have not been shown in Fig. 3) are more accurate compared with the PSFs computed from the gallery-probe pair using the AM framework in [4]. This is because blur has been decoupled from illumination in our matte-based approach, whereas the AM scheme employed by [4] is susceptible to local minima.

1. Automatic Trimap Generation

Because matting is an ill-posed problem, matting algorithms require a trimap as an additional input from the user. However, manually generating the trimap can be cumbersome. Hence, we present an automated method that requires no user interaction in trimap generation. To this end, we effectively use the landmark points (shown in blue in column one of Fig. 4) detected by the method in [32]. The convex hull (shown in red) of these landmark points is first computed. We note that the method of [32] is designed for focused images, and there can be errors in landmark point estimation when a blurred probe is processed. Therefore, to ensure that the sure foreground/background are correctly labeled, we shrink this polygon (shown in yellow) and label the region inside it as sure foreground. Likewise, the region lying outside the expanded

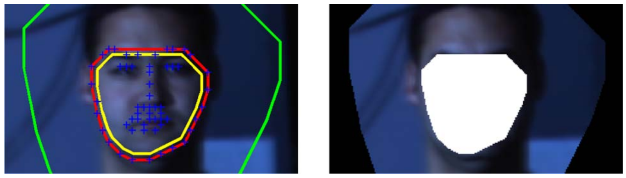

Fig. 4. Example depicting our automatic trimap generation step on a synthetically blurred face image.

polygon (shown in green) is labeled as sure background. To estimate the matte, the trimap (column two of Fig. 4), thus generated, and the blurred probe image are given as input to the closed-form matting technique of [35]. Closed-form matting works even on scribbles, which are essentially sparse trimaps [36], and its performance, therefore, is not adversely affected by a broad trimap. Note that, in addition to [14], other works such as [37,38] also have successfully used closed-form matting [35] on blurred images. The facial landmark localization code of [32] and the closed-form matting code of [35] are both publicly available.

B. Pose

Having estimated the blur from the matte independent of all other degradations, we now return to the intensity image to perform face recognition. For the sake of discussion, in this section, let us assume that the probe has the same lighting and neutral expression as the gallery and is unoccluded, i.e., only blur and pose changes need to be modeled in going from the gallery to the probe. We will relax these assumptions in the next section. Although small changes in pose can be handled by our nonuniform motion blur model itself, explicit strategies are needed to model larger pose variations. In particular, we found from our experiments that, although matching near-frontal poses (pitch and yaw angles within 15°) with the frontal gallery returned good results, there was a drastic fall in recognition accuracy for larger rotation angles (up to 20% drop for ±30° yaw). To perform recognition across nonfrontal faces, we judiciously utilize the method in [32], which we had earlier used for automatically generating the trimap. In addition to detecting landmark points, the algorithm of [32] also returns a quantized estimate of the pose of the face (between −90° to 90° yaw in intervals of 15°). Following [4], we use this pose estimate Ψ to synthesize from each frontal gallery the image of the subject under the new pose with the help of the average (generic) 3D face depth map in [39]. Thus, the knowledge of $\tilde{\boldsymbol{\omega}}$ and Ψ allows us to transform the gallery so as to match the

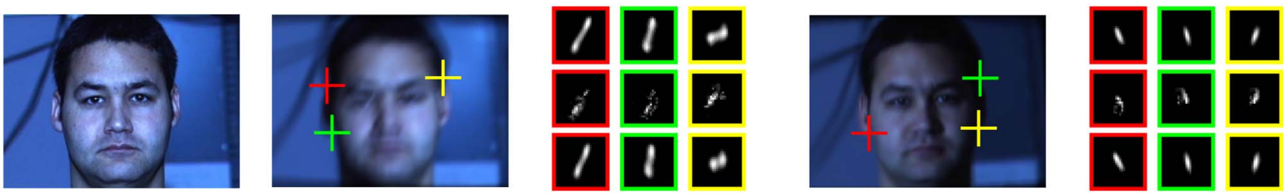

Fig. 3. Column one: gallery image. Columns two and four: a well-lit image and a poorly lit image, respectively, of the same subject synthetically blurred by applying random in-plane translations and rotations. Columns three and five: PSFs at the three locations marked by crosshairs on the images in columns two and four, respectively. Row one: true PSFs. Row two: PSFs obtained using the AM framework in [4]. Row three: PSFs obtained using our matte-based approach.

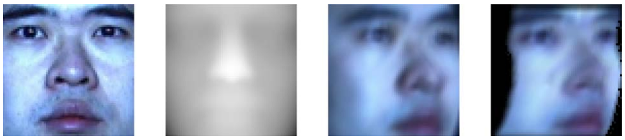

Fig. 5. Frontal gallery, average depth map, blurred probe of the same subject in a nonfrontal pose, and transformed gallery using the frontal image in column one, the average depth map in column two, and the motion estimate $\tilde{\omega}$.

blurred nonfrontal probe (see Fig. 5). Although the method in [32] has been designed to work on focused images, we found from our experiments that it returned an estimate Ψ, which is within $\pm 15°$ of the true pose 98% of the time.

C. Illumination, Partial Occlusion, and Expression Changes

Let us now consider a probe that is blurred, nonfrontal, and differently illuminated. To handle changes in illumination between the gallery and the probe, we apply the result in the well-known work of [19]. Using the "universal configuration" of lighting positions discussed in [19], a face image $\mathbf{I}$ of a person in a given pose under any illumination condition can be written as

$$\mathbf{I} = \sum_{p=1}^{9} \gamma_p \mathbf{I_p}, \tag{5}$$

where $\mathbf{I_p}$, $p = 1, 2, \ldots, 9$ forms a basis for the nine-dimensional subspace, and γ_p is the corresponding linear coefficient. The $\mathbf{I_p}$s can be generated using the Lambertian reflectance model as

$$\mathbf{I_p}[r, c] = \boldsymbol{\rho}[r, c] \max(\mathbf{n}[r, c]^T \mathbf{s_p}, 0), \tag{6}$$

where $\boldsymbol{\rho}$ and $\mathbf{n}$ are the albedo and the surface normal, respectively, at pixel location $[r, c]$, and $\mathbf{s}$ is the illumination direction. If the pose estimate Ψ is zero (i.e., frontal probe), following [3], we approximate $\boldsymbol{\rho}$ with our frontal, sharp, and well-illuminated gallery image and use the average 3D face normals from [39] for $\mathbf{n}$. However, if Ψ is non-zero, the synthesized gallery image in the new pose (following the discussion in Section 3.B) serves as $\boldsymbol{\rho}$, and the surface normals recomputed from the rotated depth map are used for $\mathbf{n}$.

We extend the result in [19] to the case of blur, and model a face in a given pose characterized by the nine basis images $\mathbf{I_p}$, $p = 1, 2, \ldots, 9$ under all possible lighting conditions and blur as

$$\sum_{p=1}^{9} \gamma_p \sum_{k \in \mathbf{S}} \omega_k \mathbf{I_{p_k}}, \tag{7}$$

where $\mathbf{I_{p_k}}$ denotes the basis image $\mathbf{I_p}$ warped according to the homography $\mathcal{H}_{\boldsymbol{k}}$.

Let us consider an M class problem with $\{\mathbf{f_m}\}_{m=1}^{M}$ denoting the gallery images, with one face per subject. Let $\mathbf{g}$ denote the probe image, which belongs to one of the M classes. The problem we are looking at is, given $\mathbf{f_m}$s and $\mathbf{g}$, to find the identity $m^* \in \{1, 2, \ldots, M\}$ of $\mathbf{g}$. From each gallery image $\mathbf{f_m}$, $m = 1, 2, \ldots, M$, we first synthesize the image corresponding to the pose of the probe $\mathbf{f_{syn_m}}$ based on the pose estimate Ψ. Following the above discussion, for each synthesized gallery image $\mathbf{f_{syn_m}}$, we obtain the nine basis images $\mathbf{f_{syn_{m,p}}}$, $p = 1, 2, \ldots, 9$. Next, we blur each $\mathbf{f_{syn_{m,p}}}$ using the estimated $\tilde{\omega}$ (see row one of Fig. 6). For each subject in the gallery, we can solve the linear least-squares problem

$$\mathbf{g} = \mathbf{L_m} \boldsymbol{\gamma_m} \tag{8}$$

to estimate the illumination. Here $\mathbf{L_m}$ is a matrix whose nine columns contain the blurred basis images in the synthesized pose corresponding to the subject m lexicographically ordered as vectors, and $\boldsymbol{\gamma_m} = [\gamma_{m,1}, \gamma_{m,2}, \ldots, \gamma_{m,9}]$ are its corresponding illumination coefficients.

Now consider a probe that has, in addition to the above degradations, expression changes and partial occlusion. Based on the observation in [22] that occlusion and expression changes only affect a sparse number of pixels, we show how the introduction of an occlusion vector to the proposed scheme allows us to model these changes, too, and develop an algorithm that is robust to nonuniform blur, pose, illumination, and occlusion. To the best of our knowledge, this is the first ever effort to even so much as to attempt this compounded scenario.

In order to account for partial occlusion and facial expression changes, we now modify, in the following manner, the framework in Eq. (8):

$$\mathbf{g_{occ}} = \begin{bmatrix} \mathbf{L_m} & \mathbf{I_N} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma_m} \\ \boldsymbol{\chi_m} \end{bmatrix} = \mathbf{B_m} \mathbf{x_m}. \tag{9}$$

Here $\mathbf{g_{occ}}$ is the blurred and occluded probe face under a different illumination and pose, and $\mathbf{I_N}$ is the $N \times N$ identity matrix that represents occlusions (N denotes the number of pixels in the face image). $\mathbf{x_m}$ is the combined vector, the first nine elements $\boldsymbol{\gamma_m} = [\gamma_{m,1}, \gamma_{m,2}, \ldots, \gamma_{m,9}]$ of which represent the illumination coefficients and the remaining N elements represent the occlusion vector $\boldsymbol{\chi_m}$ corresponding to the subject m. In Eq. (9), $\mathbf{g_{occ}}$ can be viewed as the unoccluded blurred and differently lit nonfrontal probe, plus the occlusion. To solve this underdetermined system, we leverage the prior information that the occlusion is sparse. Observe that, in the combined vector $\mathbf{x_m}$, only nine elements correspond to the illumination component, while the number of elements corresponding to the sparse occlusion component is typically much larger. Thus, we can impose l_1-norm prior on the whole vector $\mathbf{x_m}$. We estimate the combined vector $\widetilde{\mathbf{x_m}}$ by solving the following optimization problem

Fig. 6. Row one: the nine blurred basis images. Row two: the gallery image; the blurred and partially occluded probe of the same subject under a different illumination and pose; the transformed gallery image, which is a linear combination of the nine blurred basis images in row one, and the estimated occlusion.

$$\widetilde{\mathbf{x_m}} = \underset{\mathbf{x_m}}{\operatorname{argmin}} \|\mathbf{g_{occ}} - \mathbf{B_m x_m}\|_2^2 + \beta_2 \|\mathbf{x_m}\|_1. \tag{10}$$

We solve Eq. (10) using the *LeastR* function of the Lasso algorithm in [34]. This energy function when minimized provides an estimate of the lighting coefficients that take the gallery close to the illumination of the probe. In addition, it furnishes information about the location and intensity of the occluded pixels. This joint formulation for illumination and occlusion is one of our contributions in this work. Note that the locations of occlusions differ for different input images and are not known *a priori* to the algorithm. Thus, the knowledge of $\tilde{\boldsymbol{\omega}}$, Ψ and $\widetilde{\mathbf{x_m}}$ allows us to transform each of the gallery images to match the probe (see row two of Fig. 6). It is worth mentioning that the AM framework in [4], which uses a sharp/blurred image pair to estimate blur and illumination, cannot be directly extended to handle occlusion. The optimization cost in [4] is formulated based on the bi-convexity of the set of all blurred and differently lit faces. The addition of a third unknown, i.e., occlusion, violates this bi-convex property.

D. Recognition

To determine the identity of the probe, we examine the sparsity of the estimated vectors $\boldsymbol{\chi_m}$ (extracted from $\widetilde{\mathbf{x_m}}$) for the m gallery images, i.e., we rank these m vectors from most sparse to least sparse. If the difference in the extent of sparsity between the rank-1 and rank-2 vectors is greater than or equal to a threshold, then we directly declare the identity of the probe as the rank-1 subject in the gallery. The intuition behind so doing is that the algorithm will have to introduce only a few non-zero entries in the vector $\boldsymbol{\chi_m}$ at the actual locations of the sparse occlusion for a correct match. If the above difference, however, is less than the specified threshold, then we flag the rank-2 subject. Now, we compute the difference between the rank-1 and rank-3 vector. If the difference, once again, is less than the threshold, then we flag the rank-3 subject too. We proceed in this manner and flag all those vectors whose difference in the amount of sparsity between the rank-1 vector is less than the specified threshold. For this select group, we first transform each of the gallery images to match the blur, pose, and illumination of the probe using the estimated values of $\tilde{\boldsymbol{\omega}}$, Ψ, and $\gamma_{m,p}$ (first nine elements extracted from $\widetilde{\mathbf{x_m}}$), respectively. For each transformed gallery image and probe, we divide the face into nonoverlapping rectangular patches, extract LBP histograms independently from each patch, and concatenate the histograms to build a global descriptor. We perform recognition with a nearest neighbor classifier using chi-square distance [9] with the obtained histograms as feature vectors.

To maintain stable performance even for probes with high occlusion, we adopt a block partitioning approach similar to [22]. Once the estimated camera motion has been applied on the basis images corresponding to each subject in the gallery, we partition the blurred basis images and the probe image into q nonoverlapping blocks and rewrite Eq. (9) as $\mathbf{g_{occ_q}} = \mathbf{B_{m_q} x_{m_q}}$. The optimization problem in Eq. (10) then becomes

$$\widetilde{\mathbf{x_{m_q}}} = \underset{\mathbf{x_{m_q}}}{\operatorname{argmin}} \|\mathbf{g_{occ_q}} - \mathbf{B_{m_q} x_{m_q}}\|_2^2 + \beta_2 \|\mathbf{x_{m_q}}\|_1. \tag{11}$$

For a given probe image, we find the match in the gallery, as explained earlier *independently for each block*, and finally aggregate the results by voting. These steps are outlined in Algorithm 1. For all our experiments, the threshold for the difference in the extent of sparsity between the rank-1 and subsequent occlusion vectors was selected as 5% of the total number of pixels in the given block.

Algorithm 1. Motion blur, pose, illumination, and occlusion-robust face recognition

Input: Blurred and differently illuminated probe image $\mathbf{g_{occ}}$ in a nonfrontal pose with partial occlusion and facial expression changes, and a set of gallery images $\{\mathbf{f_m}\}_{m=1}^{M}$.
Output: Identity of the probe image.

1. Find the optimal $\tilde{\boldsymbol{\omega}}$ from the matte of the blurred probe $\mathbf{g_{occ}}$ by solving Eq. (4) (Section 3.A).
2. Obtain an estimate Ψ of the pose of the probe $\mathbf{g_{occ}}$ using the method in [32] (Section 3.B).
3. For each gallery image $\mathbf{f_m}$, synthesize the image in the new pose $\mathbf{f_{syn_m}}$ based on the estimated Ψ (Section 3.B).
4. For each synthesized gallery image $\mathbf{f_{syn_m}}$, obtain the nine basis images $\mathbf{f_{syn_{m,p}}}$, $p = 1, 2, \ldots, 9$, and blur each of these images using the estimated $\tilde{\boldsymbol{\omega}}$ (Section 3.C).
5. Partition the blurred basis images and the probe image into nonoverlapping blocks. For each block, solve for the nine illumination coefficients $\gamma_{m,p}$ and the occlusion vector $\boldsymbol{\chi_m}$ using Eq. (11) (Sections 3.C and 3.D).
6. For each block, find the closest match based on the extent of the estimated occlusion and LBP matching between the probe and the transformed gallery images. Determine the identity of the probe based on a majority vote across all blocks (Section 3.D).

4. EXPERIMENTS

We first evaluate the effectiveness of the proposed method using two publicly available databases: PIE [40] and AR [41]. Because these databases contain only focused images, we blur the images synthetically to generate the probes. This represents a quasi-real setting because, although the blur is synthetically added, the changes in pose, illumination, and expression and occlusions are real (Sections 4.A and 4.B). Next, we report results on the Labeled Faces in the Wild (LFW) [42] dataset using the "Unsupervised" protocol (Section 4.C). We also evaluate our algorithm on a data set we ourselves captured using a handheld camera that contains significant blur, pose, and illumination variations, in addition to large occlusions and facial expression changes (Section 4.D).

We also compare our results with the following state-of-the-art face recognition techniques:

1. MOBILAP algorithm of [4], which performs recognition across nonuniform blur, illumination, and pose.
2. IRBF [3] technique where uniform blur and illumination are jointly modeled.
3. Gopalan *et al.* [10] where recognition across both uniform and spatially varying blur is performed using blur invariants on a Grassmann manifold.
4. FADEIN [6], which infers the PSFs using learned statistical models of the variation in facial appearance caused by blur. The blurred probe is first deblurred using the inferred PSF and the deblurred image is used for recognition.
5. FADEIN + LPQ [6] where LPQ [9] features extracted from the deblurred image produced by FADEIN are used for recognition.

Table 1. Summary of Comparison Techniques[a]

S. No.	Comparison Technique	Approach	Methods Compared with	Code	Degradations Modeled: UB	NUB	I	P	O	Remarks
1	MOBILAP	Nonuniformly blurred probe as a weighted average of geometrically warped gallery	IRBF, FADEIN+LPQ BIM, SRC, DFR	Our work	✓	✓	✓	✓	✗	AM framework is susceptible to local minima for poor lighting and/or very heavy blur.
2	IRBF	Direct recognition using LBP	FADEIN, LPQ, FADEIN +LPQ	Shared by authors	✓	✗	✓	✗	✗	Targeted at recognizing faces acquired from distant cameras where the blur is well-approximated by convolution.
3	BIM	Blur invariants on a manifold for recognition	FADEIN, LPQ, HH	Shared by authors	✓	✓	✗	✗	✗	Space-varying blur is handled using overlapping patches, where the blur in each patch is assumed to be uniform.
4	FADEIN	Deblurring using inferred PSF followed by recognition	Eigen faces, Laplacian faces LBP, LPQ, HH	Our implementation	✓	✗	✗	✗	✗	Limited to learned blur kernels only. Cannot capture the entire space of PSFs. LPQ's ability to
5	FADEIN +LPQ	Recognition using LPQ features extracted from probe deblurred using FADEIN		LPQ code downloaded from authors' webpage	✓	✗	✓	✗	✗	handle illumination is governed by FADEIN correctly inferring the PSF when there is a change in lighting.
6	SRC	l_1-minimization based on sparse representation	Nearest Neighbor, Nearest Subspace, Linear SVM	Our implementation	✗	✗	✓	✗	✓	Dictionary is built using basis images of all subjects. Cannot cope with blur in the images.
7	DFR	Dictionary-based approach	SRC, CDPCA	Shared by authors	✗	✗	✓	✓	✗	
8	DSV+SRC	Probe deblurred using space-varying blind deconv. Code in DSV passed to SRC, DFR, LBP for recognition	Not applicable	DSV, LBP codes downloaded from the authors' webpage	✓	✓	✓	✗	✓	Deblurring artifacts are a major source of error.
9	DSV+DFR				✓	✓	✓	✓	✗	
10	DSV+LBP				✓	✓	✓	✗	✗	

[a]UB, uniform blur; NUB, nonuniform blur; I, illumination; P, pose; O, occlusion; BIM, blur-invariants on manifold [10]; DSV, deblur space-varying [27]; HH, [5].

6. SRC [22], which uses an l_1 minimization technique and seeks the sparsest linear representation of the probe image in terms of an overcomplete dictionary built from several training examples of each subject. SRC can handle occlusion and changes in illumination and facial expressions.

7. DFR [24], which uses a dictionary-based approach for recognition, and is designed to handle illumination and pose variations.

8. Because SRC cannot cope with blur, the probes are first deblurred using the nonuniform blind deblurring technique in Whyte *et al.* [27] before recognition is performed. The method in [27] is selected for its ability to handle spatially varying blur.

9. A similar two-step deblur [27] + recognize approach for DFR since DFR too is not designed to handle blurred images.

10. Yet another two-step comparison that uses LBP features extracted from the deblurred probe returned by [27] for recognition.

While we select methods S.Nos. 1–5 for their ability to handle blurred faces, S.Nos. 6 and 7 were chosen because comparisons in [22] and [24] with contemporary methods suggesting that the SRC and DFR algorithms are among the best for classical face recognition applications. An overview of all the comparison methods is provided in Table 1.

For our synthetic experiments in Sections 4.A and 4.B, we blur the images using the following four blur settings: in-plane translations (IP− > T), in-plane translations and rotations (IP− > T + R), out-of-plane translations (OP− > T), and in-plane and out-of-plane rotations (IP + OP− > R). We do not select a 6D camera space in view of the observation [27,30,31] that, in most practical scenarios, a 3D search space is sufficient to explain the general motion of the camera. While [30,31] use IP− > T + R and [27] uses IP + OP− > R, we select the transformation intervals on the image plane for generating blurred images as follows: in-plane translations range = [−6 : 1 : 6] pixels, out-of-plane translations range = [0.8 : 0.005 : 1.2] as scaling, and in-plane-rotations range = [−3° : 0.5° : 3°]. The focal length (in pixels) is set to 800, and the out-of-plane camera rotations range is selected as [−2° : 0.5° : 2°]. The cardinality of the camera pose space $\mathbf{S}$ for a particular setting, say IP + OP− > R, can be determined as $|\mathbf{S}|$ = Number of rotation steps about each of the three axes = ([−2° : 0.5° : 2°] about X axis) × ([−2° : 0.5° : 2°] about Y axis) × ([−3° : 0.5° : 3°] about Z axis) $= 9 \times 9 \times 13 = 1053$. The camera motion is synthesized such that it forms a sparse but connected path in the motion space, and the blur induced mimics the real blur incurred due to camera-shake. We varied the percentage sparsity (the ratio of the number of poses having a non-zero weight to the total number of poses in the motion space) of the *randomly* generated motion path from ∼6% for IP− > T + R, to 12% for IP + OP− > R, 18% for IP− > T, and 24% for OP− > T. The transformation intervals and sparsity levels above were chosen such that synthetically blurring the image using transformations lying in these intervals results in moderate to heavy blur on the face, making it a challenging problem from a face recognition perspective (see the four sample synthetically blurred probe images in Fig. 7 for severity of blur). For IP− > T, the PSFs and $\boldsymbol{\omega}$ are identical because the motion is comprised of only in-plane translations. For the remaining three settings, the search space for $\boldsymbol{\omega}$ was chosen to

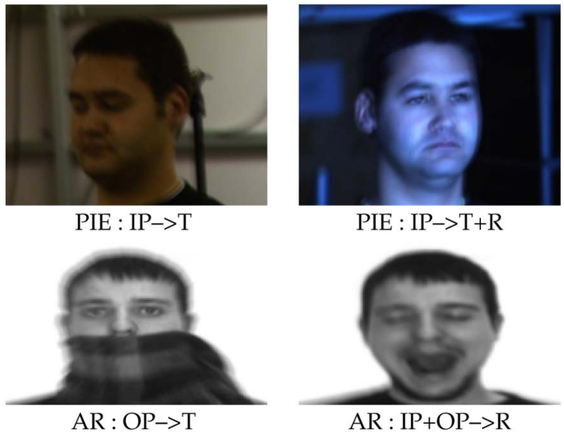

Fig. 7. Synthetically blurred probes from PIE and AR data sets.

be same as the transformation intervals used for generating the blurred probes. We used a value of 0.01 for β_1, and 100 for β_2 for all our experiments.

A. Results on PIE Database

The PIE data set [40] consists of images of 68 individuals under different pose, illumination, and expression. We first test our algorithm's ability to recognize faces across blur, pose, and expression using the images in the *Expression* folder (Section 4.A.1). Next, we use the *Illumination* folder and take up the more challenging case of performing recognition across blur, pose, illumination, and synthetic random block occlusion (Section 4.A.2).

1. Recognition Across Blur, Pose, and Expression

We select images having a frontal pose (c_{27}) and neutral expression (N W) from the *Expression* folder as our gallery. We begin with the simple case where the probes also are in the frontal pose (c_{27}). The probe set is comprised of the images labeled B W (blink without glasses), S W (smile without glasses), N G (neutral expression with glasses), B G (blink with glasses), and S G (smile with glasses). We blur all the probe images using the four different blur settings discussed in Section 4 and run our Algorithm 1. The recognition results of various methods are presented in the first plot of Fig. 8.

We used the IRBF algorithm to compare our results with [3]. Recognition scores were computed for various blur kernel sizes ranging from 3 to 13 pixels. We report the best recognition rates in the plots of Fig. 8. We wish to point out that the authors of [3] have, in their data, reported recognition rates that are, on an average, 3%–4% higher using their rIRBF algorithm. For comparison with [10] for the space-varying cases (the last three blur settings), following the discussion in Section 4.1.2 of their paper, we divided the image into overlapping patches with sizes 75%, 50%, and 40% of the original image, performed recognition separately on each patch and used a majority vote to calculate the final recognition score. (For IP− > T, the algorithm in Section 4.1.1 of their paper was used.) This was repeated for various blur kernel sizes ranging from 3 to 13 pixels, and the best recognition rates have been reported. Since Gopalan *et al.* [10] do not model variations due to illumination, we followed the approach taken in their paper and histogram

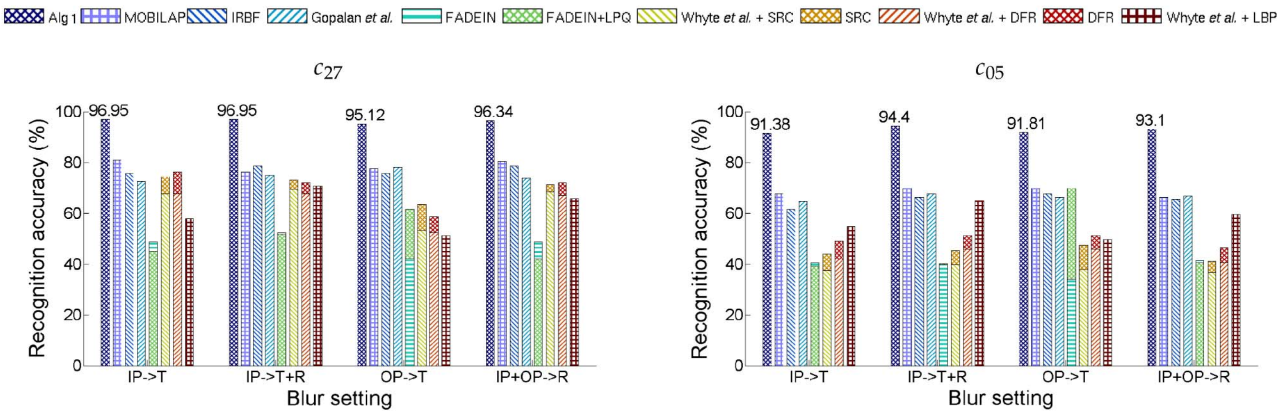

Fig. 8. Recognition results of Algorithm 1 on the *Expression* folder of the PIE database for the frontal c_{27} pose and the near-frontal c_{05} pose, along with comparisons.

equalized both the gallery and the probe images before running their algorithm. 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 $= 0, 0.25\pi, 0.5\pi, 0.75\pi$) and one "no blur" delta function. Because there is only one image per subject in the current scenario, and SRC and DFR work best in the presence of multiple images for each subject, to be fair, we provide as input to the algorithms in [22] and [24] the nine basis images of each subject in the database. It can be seen from the first plot of Fig. 8 that our method generalizes well to all types of camera motion and consistently performs better than contemporary techniques. In our approach, motion computation and pose estimation are performed only once for a given probe; hence, the overhead does not increase with the number of subjects in the gallery. On a 3.4 GHz processor running MATLAB, landmark detection and pose estimation take around 5–6 s, while trimap generation, matting, and camera motion estimation typically add another 16 to 18 s. Note that all basis images in all discretized poses can be generated and stored offline. This step, therefore, does not add to the runtime at testing. The illumination coefficients and the occlusion vector have to be determined for each gallery image. However, not only is this process fast (forward blurring using the estimated $\tilde{\boldsymbol{\omega}}$ and solving Eq. (11) take less than a second for a given gallery image), it also can be parallelized because the computation is independent for each gallery-probe pair. The final recognition step does not contribute significantly to the runtime because most matches are typically eliminated via the sparsity check, and LBP values have to be computed only on a few gallery-probe pairs. Even the LBP matching step is amenable to parallelization.

In the next experiment, we test how our technique fares when there are small changes in pose by selecting probe images that are near-frontal. To this end, we use probes images with labels B W, S W, N G, B G, and S G in the near-frontal c_{05} pose (−16° yaw). The results are presented in the second plot of Fig. 8, which shows that our method yet again scores over others.

Last, we take up the nonfrontal case where the superiority of Algorithm 1 over competing methods is more clearly revealed. We use as probes images with labels N W, B W, and S W in the nonfrontal poses c_{37} (−31° yaw) and c_{11} (32° yaw). Our algorithm, which uses a synthesized nonfrontal gallery, exhibits stable performance even under considerable pose variations and expression changes, as can be seen from the results in Table 2. However, for such large changes in pose, the accuracy of competing methods falls drastically to below 25%. Hence, their scores have not been reported.

2. Recognition Across Blur, Illumination, Pose, and Synthetic Random Block Occlusion

We now select faces with a frontal pose (c_{27}) and frontal illumination (f_{11}) from the *Illumination* folder as our gallery. For the first experiment, the probe set comprises of subsets $f_{06}, f_{07}, f_{08}, f_{09}, f_{12}$, and f_{20} (six different illumination conditions) in the frontal c_{27} pose. We simulate various levels of contiguous occlusion, from 0% to 50%, by replacing a randomly located square block of each probe face with an unrelated image. Note that the location of occlusion is randomly chosen for each image and is unknown to the algorithm. Next, we blur all the occluded probes using the last three space-varying blur settings, as discussed in Section 4. For example, the entire center of the face is occluded in the image in the third column of Fig. 9; this is a tough recognition task even for humans. Yet, our Algorithm 1 performs well, as can be seen from the plots in the first row of Fig. 10. Observe that the performance of all competing techniques, including [4], degrades quickly as the percentage of occlusion increases.

Next, we perform the same experiment but now using probes in the near-frontal c_{05} pose. The results are presented in the plots in the second row of Fig. 10, and it can be seen that our method is easily able to tolerate up to 40% occlusion even with pose variations.

Table 2. Recognition Results (%) of Algorithm 1 across Blur, Pose, and Expression on the *Expression* Folder of the PIE Database for Two Nonfrontal Poses c_{37} and c_{11}

Blur Setting	IP− > T	IP− > T + R	OP− > T	IP + OP− > R
c_{37}	81.76	86.49	87.84	81.76
c_{11}	77.70	78.38	77.70	75.68

Fig. 9. Blurred probe faces from the PIE database under varying levels of synthetic random block occlusion.

B. Results on AR Database

The AR database [41] consists of frontal images of 126 individuals with different facial expressions and occlusions. Each person in this data set participated in two sessions, separated by two weeks' time. No restrictions on wear (clothes, glasses), makeup, hair style, etc. were imposed on the participants. We choose a subset of the data set consisting of 120 subjects. The images with neutral expression from Session 1 form the gallery. We demonstrate our algorithm's ability to tackle facial expression changes between the gallery and the probe images (Section 4.B.1) and real occlusion (Section 4.B.2).

1. Recognition Across Blur and Expression

In this experiment, the probe set comprises of the images with expression changes labeled *Smile*, *Anger*, and *Scream* from both Session 1 and Session 2. The probe images are blurred using the four different blur settings discussed in Section 4. The advantages of explicitly modeling blur and expression changes are evident from the recognition results presented in the first plot of Fig. 11.

2. Recognition Across Blur and Real Occlusion

For the next experiment, the probe set consists of images with occlusion where the subjects are wearing either *Sunglasses* or *Scarf*. We note that, although our algorithm's performance with and without block partitioning was nearly the same for small levels of occlusion, there was a marked improvement in recognition accuracy for probes with larger occlusion (such as in this experiment) while using the partitioning scheme. The recognition results of Algorithm 1 are presented in the second plot of Fig. 11, and it can be observed that we outperform competing approaches by a significant margin.

C. Results on LFW Database

LFW [42] is a challenging data set designed to study the unconstrained *face verification* problem in which a pair of face images is presented, and it is required to classify the pair as either "same" or "different" depending upon whether the images are of the same person or not. The data set contains 13,233 images of 5749 subjects in which the faces have large pose, illumination, expression changes, partial occlusion, etc. However, as pointed out in [43,44], the images in LFW are typically posed and framed by professional photographers and are known to contain little or no blur. Even so, an evaluation on this data set is quite useful because, in real applications, the extent of blur in the probe images is not known *a priori*.

Because our method does not involve any training, we report results under the "Unsupervised" protocol on "View 2" of the data set consisting of 3000 matched and 3000 mismatched pairs divided into 10 sets. Note that this protocol is considered the most difficult [45] of all because no training data are available. In the Unsupervised paradigm, the area under the ROC curve (AUC) is the scalar-valued measure of accuracy according

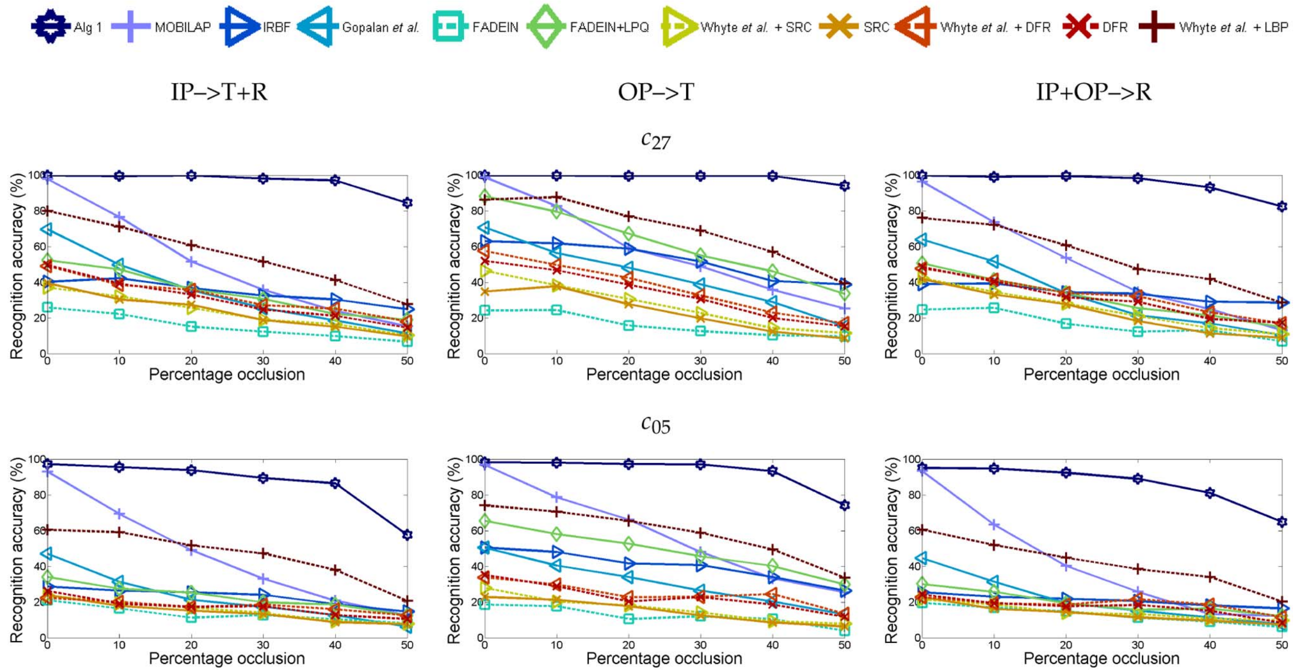

Fig. 10. Recognition results of Algorithm 1 on the *Illumination* folder of the PIE database for the frontal c_{27} pose and the near-frontal c_{05} pose under varying block occlusion, along with comparisons.

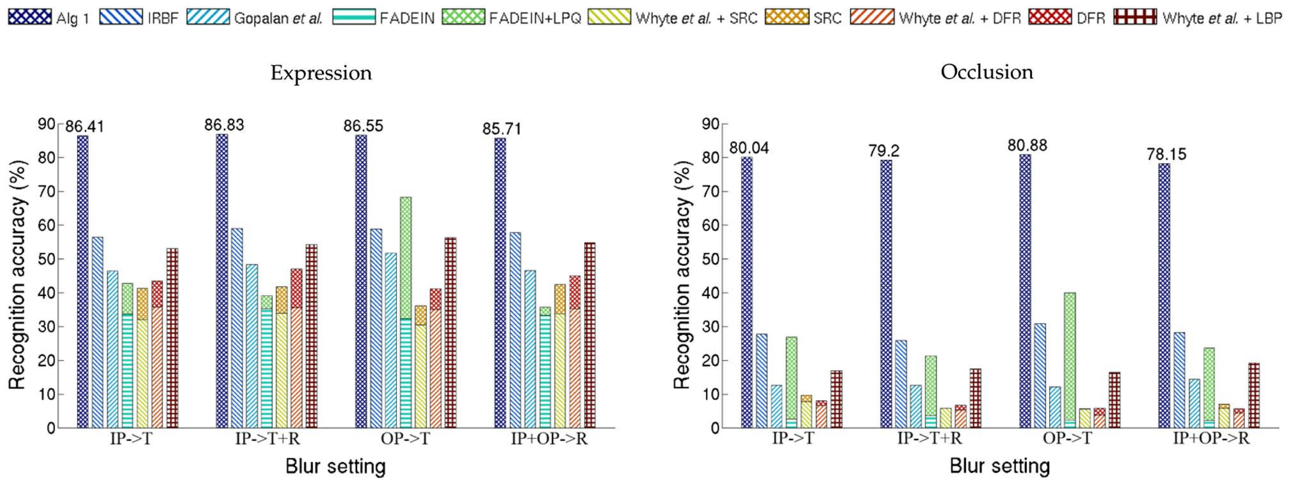

Fig. 11. Recognition results of Algorithm 1 on the AR database, along with comparisons.

to the score reporting procedures for LFW. We used the LFW-aligned version [46] of the database to report our scores. Because the images in this data set contain very little or no blur, the search intervals for $\boldsymbol{\omega}$ were kept small (in-plane translations of [−2 : 1 : 2] pixels, and in-plane rotations of [−1° : 0.5° : 1°]). For all pairs, we transform the first image to match the second using the estimated values of blur, pose, illumination, and occlusion and compare them. Following [45], we then exchange the roles of the first and second image and compare again. This procedure is then repeated for horizontally mirrored versions of both images. Of these four combinations, we select the one that gives the minimum error as our final similarity measure. Table 3 reports the AUC values obtained by our method along with other approaches that follow the Unsupervised protocol. The scores have been reproduced from the LFW results page (http://vis-www.cs.umass.edu/lfw/results.html#Unsupervised). Our AUC value is close to Spartans, which is next only to MRF-Fusion-CSKDA. The ROC curves also are shown in the plot of Fig. 12 in order to better evaluate the performance. Note that our framework can explicitly model blur, if present, while competing methods in Table 3 are not tailored to deal with it.

Table 3. Performance Comparison for Different Methods on LFW Database under the Unsupervised Protocol

Method	AUC
SD-MATCHES, 125 × 125, funneled	0.5407
H-XS-40, 81 × 150, funneled	0.7547
GJD-BC-100, 122 × 225, funneled	0.7392
LARK unsupervised, aligned	0.7830
LHS, aligned	0.8107
Pose Adaptive Filter	0.9405
MRF-MLBP, aligned [45]	0.8994
MRF-Fusion-CSKDA	0.9894
Spartans	0.9428
Ours	0.9416

D. Recognition in Unconstrained Settings

Finally, we report recognition results on a challenging data set where we captured images in an unconstrained manner. The data set has 50 subjects and 2500 probe images (this data set is expanded from our earlier work [4]; we added more images with occlusions and facial expression changes to form a larger and more challenging database). One frontal, sharp, well-illuminated, and unoccluded image taken outdoors under diffuse lighting comprises the gallery. The probe images, captured using a handheld camera under indoor and outdoor lighting, suffer from varying types and amounts of blur, changes in illumination, pose, and expression as well as occlusion. Although the blur was predominantly due to camera shake, no restriction was imposed on the movement of the subjects during image capture; therefore, a subset of these images have both camera and object motion.

Compared with the gallery, the probes were either overlit or underlit depending on the time of the day and the setting under which they were captured. For indoor scenes, the exposure time in some cases was as large as 1 s resulting in images with

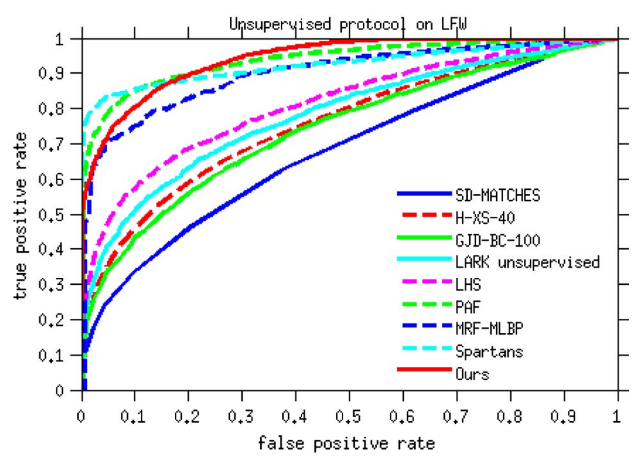

Fig. 12. ROC curves of different approaches on the LFW database for the Unsupervised protocol.

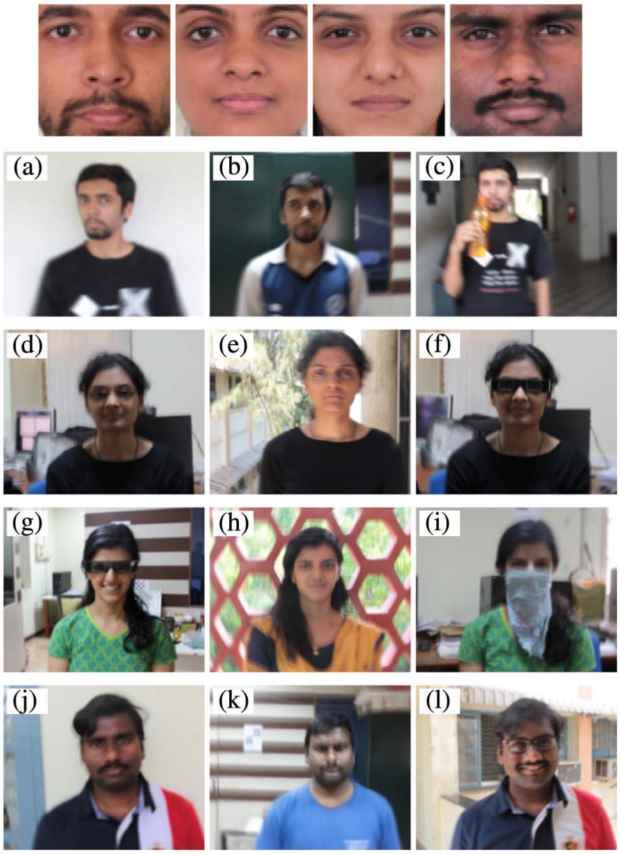

Fig. 13. Cropped gallery faces of four subjects from our own data set are shown in row one. Sample blurred probe images shown in (a)–(l) have variations in illumination, (a) pose, (g,l) facial expressions changes, and (d), (f), (g), (i), (l) occlusions.

significant blur. The distance between the camera and the subject was also allowed to vary. The background in most cases had significant clutter. Sample images from our dataset are shown in Fig. 13. We selected the search intervals for $\boldsymbol{\omega}$ as $-10:1:10$ pixels for in-plane translations and $-2°:0.25°:2°$ for in-plane rotations for running our Algorithm 1. The recognition rates, presented in the plot of Fig. 14, are a clear indicator of the efficacy of the proposed framework in advancing the state-of-the-art in unconstrained face recognition.

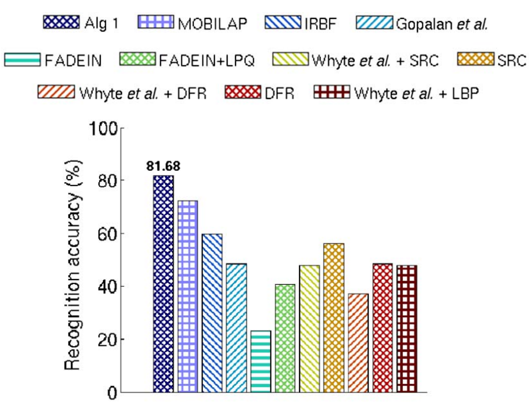

Fig. 14. Recognition results on our own data set using Algorithm 1, along with comparisons.

5. DISCUSSION AND CONCLUSIONS

We proposed a methodology that harnesses the alpha matte extracted from the probe to solve for the relative motion between the camera and the face independent of all other degradations. We demonstrated how matching across poses is possible by transforming the gallery to the nonfrontal pose of the probe. We also showed how applying the estimated motion on each of the nine basis images in the synthesized nonfrontal pose and taking their weighted sum allows us to model illumination changes, too. Finally, variations in expression and partial occlusion were subsumed into the proposed framework by appending an occlusion dictionary to model the sparse occlusion. Illumination and occlusion were jointly solved for, and the sparsity of the estimated occlusion vector was itself judiciously used to aid the recognition task. In summary, we have addressed in this paper the challenging scenario of performing face recognition across nonuniform blur, varying pose, illumination, and expression as well as partial occlusion, when only a single image of each subject is available in the gallery. We also presented a completely automated pipeline demonstrating our algorithm's viability for practical use. Our experiments revealed that our method significantly outperforms contemporary techniques and is well-equipped to handle complex motion trajectories, yaw angles as large as $\pm 30°$, harsh illumination conditions, and considerable occlusion.

REFERENCES

1. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: closing the gap to human-level performance in face verification," in *Proceedings of Conference on Computer Vision and Pattern Recognition* (2014), pp. 1701–1708.
2. F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: a unified embedding for face recognition and clustering," in *Proceedings of Conference on Computer Vision and Pattern Recognition* (2015), pp. 815–823.
3. P. Vageeswaran, K. Mitra, and R. Chellappa, "Blur and illumination robust face recognition via set-theoretic characterization," IEEE Trans. Image Process. **22**, 1362–1372 (2013).
4. A. Punnappurath, A. N. Rajagopalan, S. Taheri, R. Chellappa, and G. Seetharaman, "Face recognition across non-uniform motion blur, illumination, and pose," IEEE Trans. Image Process. **24**, 2067–2082 (2015).
5. H. Hu and G. De Haan, "Adaptive image restoration based on local robust blur estimation," in *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS)* (2007), pp. 461–472.
6. 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. **33**, 838–845 (2011).
7. J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring face images with exemplars," in *Proceedings of European Conference on Computer Vision (ECCV)* (2014), pp. 47–62.
8. H. Zhang, J. Yang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang, "Close the loop: joint blind image restoration and recognition with sparse representation prior," in *Proceedings of International Conference on Computer Vision (ICCV)* (2011), pp. 770–777.
9. T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in *Proceedings of International Conference on Pattern Recognition (ICPR)* (2008), pp. 1–4.
10. R. Gopalan, S. Taheri, P. Turaga, and R. Chellappa, "A blur-robust descriptor with applications to face recognition," IEEE Trans. Pattern Anal. Mach. Intell. **34**, 1220–1226 (2012).

11. T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Anal. Mach. Intell. **24**, 971–987 (2002).
12. H. Li, K. N. Ngan, and Q. Liu, “FaceSeg: automatic face segmentation for real-time video,” IEEE Trans. Multimedia **11**, 1182–1193 (2009).
13. M. Hofmann, S. Schmidt, A. N. Rajagopalan, and G. Rigoll, “Combined face and gait recognition using alpha matte preprocessing,” in *Proceedings of International Conference on Biometrics (ICB)* (2012), pp. 390–395.
14. J. Jia, “Single image motion deblurring using transparency,” in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)* (2007), pp. 1–8.
15. X. Zhang and Y. Gao, “Face recognition across pose: a review,” Pattern Recogn. **42**, 2876–2896 (2009).
16. C. Ding and D. Tao, “A comprehensive survey on pose-invariant face recognition,” arXiv: 1502.04383 (2015).
17. M. Osadchy, D. Jacobs, and M. Lindenbaum, “Surface dependent representations for illumination insensitive image comparison,” IEEE Trans. Pattern Anal. Mach. Intell. **29**, 98–111 (2007).
18. S. Biswas, G. Aggarwal, and R. Chellappa, “Robust estimation of albedo for illumination-invariant matching and shape recovery,” IEEE Trans. Pattern Anal. Mach. Intell. **31**, 884–899 (2009).
19. K.-C. Lee, J. Ho, and D. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” IEEE Trans. Pattern Anal. Mach. Intell. **27**, 684–698 (2005).
20. H. Jia and A. Martinez, “Support vector machines in face recognition with occlusions,” in *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (2009), pp. 136–141.
21. X. Tan, S. Chen, Z.-H. Zhou, and J. Liu, “Face recognition under occlusions and variant expressions with partial similarity,” IEEE Trans. Inf. Forensics Secur. **4**, 217–230 (2009).
22. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell. **31**, 210–227 (2009).
23. Y. Liu, C. Liu, Y. Tang, H. Liu, S. Ouyang, and X. Li, “Robust block sparse discriminative classification framework,” J. Opt. Soc. Am. A **31**, 2806–2813 (2014).
24. V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, “Dictionary-based face recognition under variable lighting and pose,” IEEE Trans. Inf. Forensics Secur. **7**, 954–965 (2012).
25. V. M. Patel, Y.-C. Chen, R. Chellappa, and P. J. Phillips, “Dictionaries for image and video-based face recognition (invited),” J. Opt. Soc. Am. A **31**, 1090–1103 (2014).
26. H. T. Ho and R. Gopalan, “Model-driven domain adaptation on product manifolds for unconstrained face recognition,” Int. J. Comput. Vis. **109**, 110–125 (2014).
27. O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, “Non-uniform deblurring for shaken images,” Int. J. Comput. Vis. **98**, 168–186 (2012).
28. L. Xu, S. Zheng, and J. Jia, “Unnatural l0 sparse representation for natural image deblurring,” in *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (2013), pp. 1107–1114.
29. Z. Hu and M.-H. Yang, “Good regions to deblur,” in *Proceedings of European Conference on Computer Vision (ECCV)* (2012), pp. 59–72.
30. Z. Hu and M.-H. Yang, “Fast non-uniform deblurring using constrained camera pose subspace,” in *Proceedings of British Machine Vision Conference (BMVC)* (2012), pp. 1–11.
31. A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless, “Single image deblurring using motion density functions,” in *Proceedings of European Conference on Computer Vision (ECCV)* (2010), pp. 171–184.
32. X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (2012), pp. 2879–2886.
33. C. Vijay, C. Paramanand, A. N. Rajagopalan, and R. Chellappa, “Non-uniform deblurring in HDR image reconstruction,” IEEE Trans. Image Process. **22**, 3739–3750 (2013).
34. J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections* (Arizona State University, 2009).
35. A. Levin, D. Lischinski, and Y. Weiss, “A closed-form solution to natural image matting,” IEEE Trans. Pattern Anal. Mach. Intell. **30**, 228–242 (2008).
36. C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott, “A perceptually motivated online benchmark for image matting,” in *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (2009), pp. 1826–1833.
37. S. Dai and Y. Wu, “Removing partial blur in a single image,” in *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (2009), pp. 2544–2551.
38. V. Caglioti and A. Giusti, “On the apparent transparency of a motion blurred object,” Int. J. Comput. Vis. **86**, 243–255 (2010).
39. V. Blanz and T. Vetter, “Face recognition based on fitting a 3D morphable model,” IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1063–1074 (2003).
40. T. Sim, S. Baker, and M. Bsat, “The CMU pose, illumination, and expression database,” IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1615–1618 (2003).
41. A. Martnez and R. Benavente, “The AR face database,” Technical Report 24 (Computer Vision Center, 1998).
42. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: a database for studying face recognition in unconstrained environments,” Technical Report 07–49 (MIT, 2007).
43. R. Chellappa, J. Ni, and V. M. Patel, “Remote identification of faces: problems, prospects, and progress,” Pattern Recogn. Lett. **33**, 1849–1859 (2012).
44. N. Pinto, J. DiCarlo, and D. Cox, “How far can you get with a modern face recognition test set using only simple features?” in *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (2009), pp. 2591–2598.
45. S. R. Arashloo and J. Kittler, “Efficient processing of MRFs for unconstrained-pose face recognition,” in *Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (2013).
46. L. Wolf, T. Hassner, and Y. Taigman, “Effective unconstrained face recognition by combining multiple descriptors and learned background statistics,” IEEE Trans. Pattern Anal. Mach. Intell. **33**, 1978–1990 (2011).