Probabilistic Deformation Models for Challenging Periocular Image Verification

Jonathon M. Smereka, *Student Member, IEEE,* Vishnu Naresh Boddeti, *Member, IEEE,* B.V.K. Vijaya Kumar, *Fellow, IEEE*

***Abstract*—The periocular region as a biometric trait has recently gained considerable traction, especially under challenging scenarios where reliable iris information is not available for human authentication. In this paper, we consider the problem of one-to-one ($1:1$) matching of highly non-ideal periocular images captured in-the-wild under unconstrained imaging conditions. Such images exhibit considerable appearance variations including non-uniform illumination variations, motion and defocus blur, off-axis gaze, and non-stationary pattern deformations. To address these challenges, we propose Periocular Probabilistic Deformation Models (PPDMs) which, 1) reduce the image matching problem to matching local image regions, and 2) approximate the periocular distortions by local patch level spatial translations whose relationships are modeled by a Gaussian Markov Random Field (GMRF). Given a periocular image pair, we determine the distortion-tolerant similarity metric by regularizing local match scores by the maximum a-posteriori probability (MAP) estimate of the relative local deformations between them. Unlike existing global periocular image matching techniques, by accounting for local image deformations in the periocular matching process PPDM exhibits greater tolerance to pattern variations. We demonstrate the effectiveness of our model via extensive evaluation on a large number of "in-the-wild" periocular images. We find that PPDMs outperform many benchmark $1:1$ image matching techniques (improving verification rates at 0.1% false accept rate by ~30% over previous work and ~40% when compared to the best baseline) in challenging scenarios leading to state-of-the-art verification performance on multiple real-world periocular datasets.**

I. Introduction

MATCHING images in the presence of image distortion is an important task in many applications including biometrics [1], optical character recognition (OCR) [2], image retrieval [3], and automatic target recognition [4]. In biometric recognition, iris recognition is very popular as it exhibits high recognition rates in controlled settings. However, in order to achieve the high levels of accuracy desired, a high quality, non-occluded iris image is required. In addition, any failure in segmentation of the iris from the image (caused by specular reflections, eyelashes, eye lids, non-frontal gaze, etc.) can significantly degrade system performance. One way to try to alleviate these issues is to use the entire periocular region.

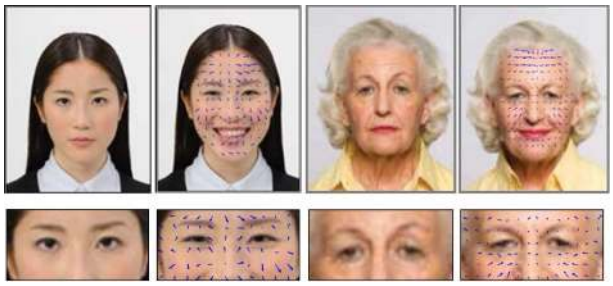


Fig. 1. Example passport photos with flow vectors overlaid to illustrate the deformation over the full face (top) and periocular regions (bottom) due to changing expression. Note that the different motion vectors in different regions of the face and bi-ocular images are of different lengths and orientations with the periocular region exhibiting less deformations compared to the other parts of the face.

Although a universally accepted definition of the periocular region does not seem to exist, we define it as the area of the face that includes the eyelids, eyelashes, eyebrow, and the skin surrounding the eye. Periocular recognition has been shown to have an increased resistance to the effects of aging [5], to be helpful in recognition performance when fused with iris images [6]–[8] and to be useful when fused with face images [9], [10]. There is also evidence of the periocular recognition outperforming iris and face recognition in challenging environments [11]–[14]. Some work [5] has focused on the bi-ocular region which includes both eyes within a single image, however this work and many others use the approach of splitting the left region from the right region to perform matching separately on each of the two periocular regions.

The periocular region does not undergo as much deformation as the rest of the facial region (see Fig. 1) under ideal imaging conditions. However, under challenging imaging conditions (such as subjects with off-axis gaze, partial occlusion, cross-sensor matching,

Manuscript received October 7, 2014; revised January 19, 2015; accepted May 12, 2015. Date of publication xxxxxxx xx, 2015; date of current version xxxxx xx, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was xxxxxxxxxxxx xxxxxxxxxxxx.

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

The authors are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213. (e-mail: jsmereka@cmu.edu, naresh@cmu.edu, kumar@ece.cmu.edu)

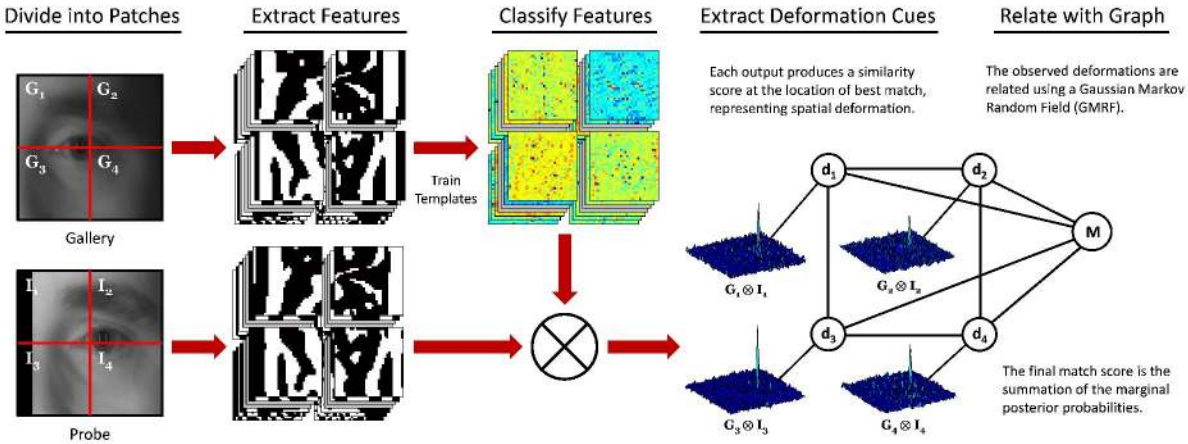
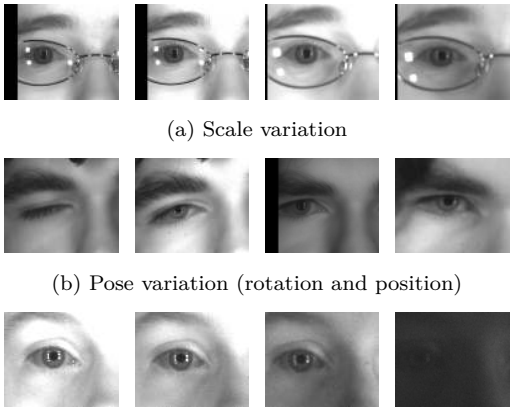


Fig. 2. System overview for determining a match. The probe, **I**, and gallery, **G**, images are each divided into non-overlapping patches where the corresponding probe and gallery patches are compared via template matching. The outputs from template matching are then used as inputs into a Gaussian Markov Random Field (GMRF) model that is trained to capture the relationship between the deformations of the image patches for authentic matches. The final match score (M) is the sum of the marginal posterior probabilities.



(a) Scale variation

(b) Pose variation (rotation and position)

(c) Illumination variation

Fig. 3. These periocular regions are extracted from near infrared (NIR) video sequences collected from moving subjects in an unconstrained environment exhibiting variations in illumination, scale, rotation, and position.

etc.), the periocular region exhibits significant pattern deformation due to sensor noise, movement (pose and scale change, Fig. 3a and 3b), illumination variations (Fig. 3c), eye gaze variations, and occlusions. The image matching challenge is noticeably exacerbated when only a single image is available for the gallery template, as is the case in several NIST biometric competitions [15]–[17], where a $1:N$ search is performed equivalently as N one-to-one $(1:1)$ comparisons. These evaluations are designed to mimic real world scenarios where a database sample (such as a passport or visa or driver's license image) is compared against an image of that person captured in the field. Shown in Fig. 1 is an example of two passport photos with flow vectors overlaid to demonstrate variability (due to expression difference) between a single gallery and a single probe image. To properly match these images and others in $1:1$ matching schemes, it is necessary to model and take into account the distortion between similar patterns.

Variability or the distortions in the patterns due to factors such as illumination, occlusion, facial expression changes (such as seen in Fig. 1), and/or possibly other non-stationary deformations can be accounted for partly by using robust pattern descriptors. Examples of robust features include LBP [18], SIFT [19], GIST [20], HOG [21], and others. However, in the presence of large variability and/or poor image quality, the robustness provided by just the pattern representations is usually insufficient to achieve desirable error rates as shown in [22] and [23]. A survey of techniques [24] indicates that most of the methods use some combination of HOG, SIFT, and LBP features for periocular matching.

Methods such as the Image Distortion Model (IDM) [25] and SIFT-Flow [26] try to directly model the pattern distortion process that occurs between the probe and gallery while determining the degree of match. Specifically, these methods aim to determine the individual pixel shifts required to best fit one image to another during matching, consequently measuring deformation at a fine level. While modeling image deformations at a fine level can result in improved image matching performance under low-noise or low-distortion conditions, it can also lead to poor robustness against more noisy observations as demonstrated in [27], [22], and [28] as well as later in this paper. In addition, these methods are often computationally prohibitive as will be evident from our comparison of computation times.

In this work, we seek to estimate a similarity (or match score) between images with relative distortions rather than estimate the pixel level motion between the images. Towards this end we estimate the deformation between images at a patch/region level, as opposed to pixel level. Using patches instead of pixels allows the model to account for large image variations while also capturing local distortions. Further, as the patches contain several pixels, the noise in the individual pixel distortions averages out to result in an improved estimate of the local distortions. The trade-off is that the model can no longer focus on granular details from pixel level alignment between the gallery and probe images.

We approximate the deformation between the patches/regions by a local translation between corresponding patches in the image pair being compared (we assume that a global alignment of the two images is

(a) FOCS (top) and UBIRIS v2.0 (bottom) (b) BDCP intraclass (c) UBIPr (top) and FRGC (bottom)

Fig. 4. Examples of periocular images from different match scenarios: (a) Unconstrained Subjects: near infrared (NIR) and visible light (VL) images demonstrating the challenging nature of the FOCS (top) and UBIRIS v2.0 (bottom) datasets. (b) Cross Sensor: the BDCP dataset matches users captured with two separate sensors, CFAIRS (top) and LG4000 (bottom). (c) Cross Dataset: the model is trained on one dataset (UBIPr, example images shown in the top row) and is tested on a different dataset (FRGC, example images shown in the bottom row). Matching is performed using a only a single probe and gallery sample, also referred to as one-to-one (1 : 1).

first carried out, e.g., using a face detector to put boxes around faces in images). This is completed by building a classifier per patch in which the output of each independent patch comparison produces a similarity and location of best match (representing a spatial translation). To capture the prior on valid patch translations, we employ a Gaussian Markov Random Field (GMRF) [29] due to their ability to compactly represent pairwise relationships between neighbors, as described in Section VI. An overview of this approach can be seen in Fig. 2.

Our work builds upon Probabilistic Deformation Models (PDMs) proposed by Thornton et al. [27] for matching iris images in the presence of distortions, demonstrating excellent performance given multiple well segmented iris images for training. However as we show in Section V and Section VIII-D, as the available training data decreases, in the context of periocular images, this model is not expressive enough to accurately match the probe to the gallery image. In this paper we introduce, 'Periocular PDM' (PPDM), to adapt PDM to challenging match scenarios while improving its algorithmic efficacy. Our key contributions are:

- We introduce algorithmic improvements to the original PDM model to improve deformation tolerance in 1 : 1 matching schemes. More specifically,
  - We explicitly model the spatial variability in the discriminative power of the patches composing the biometric signature through the likelihood distribution of our model.
  - We decouple the individual deformation components and estimate parameters as a linear regression problem to allow for a simpler, and more efficient method to better capture authentic deformations while reducing the time and memory complexity of model parameter estimation.
- A comprehensive comparison to several other methods on multiple periocular datasets demonstrating the need for deformation estimation in 1 : 1 matching schemes as well as the efficacy of PPDM. We achieve state-of-the-art periocular recognition results under the challenges of unconstrained subjects, illumination, occlusion, facial expression changes, and cross sensor matching.

Furthermore, we evaluate PPDM across multiple different challenging periocular matching scenarios:

1) **Unconstrained subjects**: The Face and Ocular Challenge Series (FOCS) [30] and University of Beira Interior iris (UBIRIS v2.0) [31] datasets display the presence of challenges shown in Fig. 3 and Fig. 4a.
2) **Cross sensor**: The Biometrics Exploitation Science and Technology (BEST) Development Challenge Problem (BDCP) [32] dataset compares images of users captured from two separate sensors (seen in Fig. 4b), the LG4000 camera [33] (gray scale images of just the eye) and the Honeywell Combined Face and Iris Recognition System (CFAIRS) [34] camera (RGB images capturing a region that varies between the full periocular region and just the eye).
3) **Separate dataset bias**: Each model is trained using frontal view images from the University of Beira Interior Periocular (UBIPr) [35] dataset and tested on a subset of images from the Face Recognition Grand Challenge (FRGC) [36].

II. Related Work

Matching biometric images with relative distortions is a challenging problem. Pupil dilation from uncontrolled lighting has been shown to yield higher dissimilarities between authentic iris image pairs [27], [37], [38]. Distortion of the fingerprint from the elasticity of the skin or the pressure often leads to difficulties in matching [39], [40]. Facial expressions introduce non-stationary image movements that notably affect matching performance [41], [42]. Even unnatural factors such as make-up and plastic surgery have been shown to degrade matching in both facial and periocular recognition systems [10], [43]. Additionally, large stand-off distances, moving subjects, aging, and pose can cause the performance of recognition algorithms to degrade considerably. Accounting for such deformations is not trivial, leading to several different approaches.

Kakadiaris et al. [44] as well as Cai and Da [45] extract the facial distortion when performing 3D face recognition by registering each face against a 3D annotated reference model. Then during authentication, the deformations are compared using a distance metric.

In an approach related to deformation extraction, Chen et al. [39] performs fingerprint matching by directly estimating the average deformation from minu-

tiae locations and orientations and then uses the deformation model to pre-distort the gallery prior to matching with a probe image. Schuckers et al. [46] developed an angular deformation model to compensate for off-angle iris samples by learning homographies at specific gaze deviations (0°, 15°, and 30°) and transforming the off-angle iris into a frontal-view image for recognition.

In another approach, Bhatt et al. [9] implicitly compensates for non-stationary variations within the face by generating features at multiple levels of granularity, and then using a genetic algorithm to determine the best way to weight and combine the features for matching. Juefei-Xu [47] uses kernel class dependence feature analysis (KCFA) with LBP descriptors encoded from discrete transform coefficients for robustness when performing bi-ocular recognition. Unfortunately, such methods require class label information for the training dataset along with a large number of training images per class. Our proposed method on the other hand only requires similarity and dis-similarity labels per pair of images and works with only one training image.

Outside the area of biometrics, estimating local deformations for similar images is not a new area of research. Many [2], [25] attempt to measure deformation at a fine level, though others [22], [28] note the benefit of using a coarse approximation to matching.

The method proposed in this paper, is most similar to [9], [28], measures deformation at a region level. However, instead of trying to find appropriate features, optimizing for a defined shape, or extracting the distortion from each sample, the patch distortion model in this paper is used to weight the patch similarities by the likelihood of the respective estimated deformations between the patches. Similar to [28], the spatial information from each region approximates local observed deformations, but as opposed to appending the global position to the feature vectors, a GMRF learns relevant region level xy translations between similar images.

Our approach for 1 : 1 matching of deformed periocular images is based on our previous work [11], [48] where in the former work we use PDM for periocular recognition to investigate the effectiveness of the ocular region as a biometric trait compared to iris recognition when dealing with challenging ocular images. While in the later work we investigate periocular image segmentation by applying PDM and m-SIFT to the FOCS and UBIPr datasets. In this work we expand on PDM, introducing PPDM as novel approach to improving system performance by handling poor deformation estimates frequently encountered in the presence of challenges from real biometric acquisition environments. We demonstrate the efficacy of PPDM by achieving state-of-the-art periocular recognition results under a separate dataset bias, cross sensor matching, and a comprehensive comparison to several other matching methods in addition to challenges presented from unconstrained subjects and therefore represents a large improvement in the state-of-the-art.

III. Overview

In this section we provide a broad overview of PPDM for 1 : 1 matching of deformable patterns, see Fig. 2 for a pictorial illustration. A given probe image, $\mathbf{I}$, and gallery image, $\mathbf{G}$, are divided into non-overlapping rectangular patches and the corresponding probe and gallery patches are matched via template matching. The resulting outputs are then used as inputs into the GMRF model, trained to capture the relationship between the deformations of the image patches for authentic matches, which estimates the deformations between the patches. These deformation estimates are used to weight the similarities between the corresponding patches to compute a match score between the gallery and the probe. Applying this method to challenging recognition scenarios presents additional difficulties to account for poor deformation cues from a lack of training data to build gallery templates, occlusions, pose change, etc. We tailor the PPDM model to these types of scenarios through the likelihood and prior distributions. The likelihood distribution is built to explicitly model the variability of the deformation cues across the periocular region. While the prior distribution is constructed via efficient parameter estimation with linear regression to better capture authentic deformations.

The local template matching scheme employed is described in Section IV and Section V details our deformation model. Details of the parameter estimation process for the model are provided in Section VI followed by model inference and similarity score computation in Section VII. Experiments throughout the paper are performed over the datasets described in Section VIII-A. Finally, full numerical experiments for periocular recognition are described in Section VIII with results of PPDM (Section VIII-D) compared against baselines (Section VIII-C).

IV. Deformation and Similarity Cues

We employ template matching to obtain the patch similarity and deformation cues that serve as inputs to the graphical model which determines a match score between $\mathbf{G}$ and $\mathbf{I}$ by regularizing the individual patch similarity score by their respective deformation estimates. The template matcher serves to provide **both** the *deformation cues*, $x - y$ translation of each patch, and the *similarity cues*, match score for each relative shift between the corresponding patches. Therefore the template matcher requires good localization performance. Although models like Support Vector Machines and Random Forests are widely used for discriminative template matching, they are neither well suited for the 1 : 1 template scenario (only one training example) nor are they explicitly designed for localization and hence are unable to provide high localization accuracy. Correlation filters (CFs) are another class of classifiers which are generally designed for high localization performance and can be designed even with just a single training

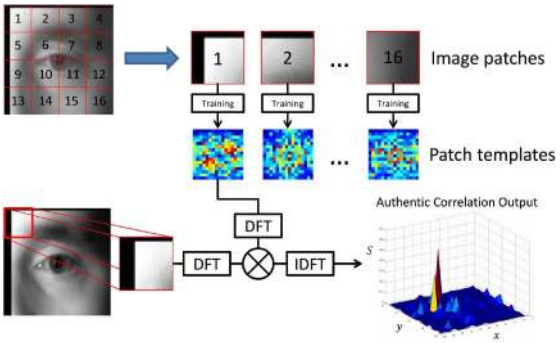


Fig. 5. Overview for determining patch similarity. A template is trained for each patch from gallery image and cross-correlated with the corresponding patch from the probe. An authentic correlation output produces a sharp peak at the location of best match.

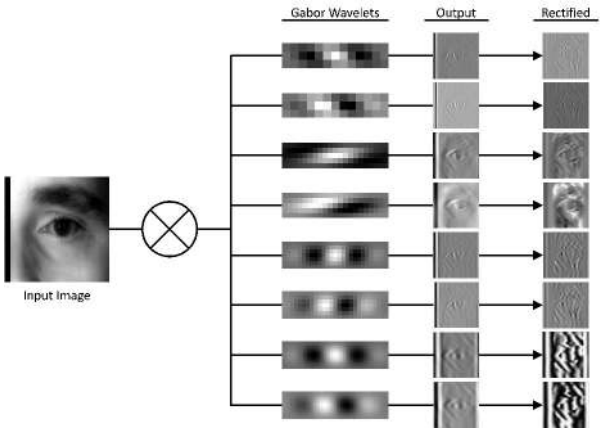
image and are hence better suited to provide both the deformation and the similarity cues.

CFs are well explained in previous publications [11], [49], [50] and hence we provide only a brief summary. A CF is a spatial-frequency array (equivalently, a template in the image domain) that is specifically designed from a set of training patterns that are representative of a particular class. As displayed in Fig. 5, the trained template is compared to a probe image by obtaining the cross-correlation as a function of the relative shift between the gallery and the probe. The idea behind CF design is to control the shape of the cross-correlation output between the training image and the filter by minimizing the mean square error (MSE) between the cross-correlation output and the ideal desired correlation output for an authentic (or impostor) input image. By explicitly controlling the shape of the entire correlation output, producing a sharp peak at the center of the correlation output for a centered authentic probe pattern and no such peak for an impostor, CFs provide more precise deformation estimates.

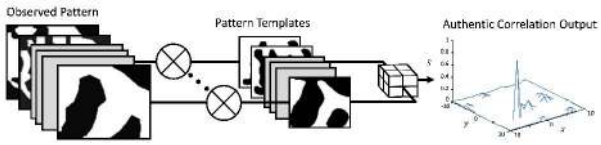
Considering that each patch region only provides limited discriminative information to the entire image, multiple feature sets are extracted to build a more robust template, i.e., each pixel or a group of pixels is represented by a vector of features. As displayed in Fig. 6a, we construct a set of feature channels by applying a rectified linear unit (RLU: $f(x) = \max(0, x)$)[1] to the output derived from a bank of Gabor filters [51]. We then design one CF per feature channel (K in total) such that the correlation output is a sum of individual outputs as depicted in Fig. 6b. Contrary to building K independent CFs, we design all K CFs *jointly* to ensure the output satisfies our design criteria. The joint CF design is posed as the following optimization problem:

$$\min_{\mathbf{h}_1,\ldots,\mathbf{h}_K} \frac{1}{r}\sum_{i=1}^{r}\left\|\sum_{k=1}^{K}\mathbf{z}_k^i \otimes \mathbf{h}_k - \mathbf{g}_i\right\|_2^2 + \lambda\sum_{k=1}^{K}\|\mathbf{h}_k\|_2^2 \quad s.t. \sum_{k=1}^{K}\mathbf{h}_k^T\mathbf{z}_k^i = u_i \quad i = 1, 2, \cdots, r \tag{1}$$

[1]The RLU provides additional discrimination between the resulting Gabor filter outputs to ensure that $\hat{\mathbf{D}}$ is full rank.



(a) Features are computed by applying a rectified linear unit (ReLU) to the output derived from a set of Gabor wavelets [51] (constituting the separate 'feature channels' when designing the CFs ($r = 1$ and $K = 8$).



(b) The outputs of individual feature channels are aggregated to produce the final correlation output which would have a sharp peak at the target location.

Fig. 6. Overview of (a) feature extraction and (b) correlation filter application.

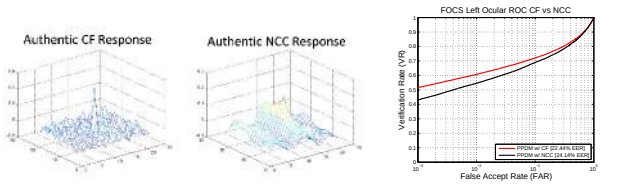
where $\otimes$ denotes the correlation operation, $\mathbf{z}_k^i$ is the $k-$th feature dimension of image $\mathbf{x}_i$ (of r images), $\mathbf{h}_k$ is the $k-$th correlation filter (corresponding to the $k-$th feature) and $\mathbf{g}_i$ is the desired correlation output for the $i-$th image and u_i is the desired ideal inner-product output while λ is the regularization parameter. The terms in Eq. 1 aim to minimize the MSE between the actual and desired correlation output for each training image while constraining the filter to produce a large value (i.e., $u_i = 1$) given an authentic correlation and small value (i.e., $u_i = 0$) otherwise.

The following closed form expressions for the CF, $\hat{\mathbf{H}} = \left[\hat{\mathbf{h}}_1^T, \ldots, \hat{\mathbf{h}}_K^T\right]^T$ (where $\hat{\mathbf{H}} \in \mathbb{C}^{KM\times 1}$, K feature sets and M is the dimensionality of each feature set), can be derived by posing the optimization problem in the frequency domain provided $\mathbf{g}_i$ is a delta function (see [50] for more details):

$$\hat{\mathbf{H}} = \hat{\mathbf{T}}^{-1}\hat{\mathbf{Z}}\left(\hat{\mathbf{Z}}^{\dagger}\hat{\mathbf{T}}^{-1}\hat{\mathbf{Z}}\right)^{-1}\mathbf{u} \tag{2}$$

where the hat indicates the 2D Discrete Fourier Transform (DFT) and $\hat{\mathbf{Z}} = \left[\hat{\mathbf{z}}^1, \hat{\mathbf{z}}^2, \ldots, \hat{\mathbf{z}}^r\right]$ is the training matrix composed of concatenated vectors $\hat{\mathbf{z}}_k^i$ from the r training images of K feature sets in the frequency domain (i.e., $\hat{\mathbf{z}}^i = \left[(\hat{\mathbf{z}}_1^i)^T, \ldots, (\hat{\mathbf{z}}_K^i)^T\right]^T$ where $\hat{\mathbf{z}}_k^i \in \mathbb{C}^M$ and $\hat{\mathbf{Z}} \in \mathbb{C}^{KM\times r}$), and $\hat{\mathbf{T}} = \lambda I + (1-\lambda)\hat{\mathbf{D}}$ where I is the identity matrix to provide regularization to noise and $\hat{\mathbf{D}}$ is the average energy of the extracted features.

$$\hat{\mathbf{D}} = \begin{bmatrix} \frac{1}{r}\sum_i \hat{\mathbf{z}}_1^{(i)*}\hat{\mathbf{z}}_1^{(i)} & \cdots & \frac{1}{r}\sum_i \hat{\mathbf{z}}_1^{(i)*}\hat{\mathbf{z}}_K^{(i)} \\ \vdots & \ddots & \vdots \\ \frac{1}{r}\sum_i \hat{\mathbf{z}}_K^{(i)*}\hat{\mathbf{z}}_1^{(i)} & \cdots & \frac{1}{r}\sum_i \hat{\mathbf{z}}_K^{(i)*}\hat{\mathbf{z}}_K^{(i)} \end{bmatrix} \tag{3}$$



(a) (b)

Fig. 7. (a) Example CF and normalized cross-correlation (NCC) gallery template responses from the same authentic probe image patch. (b) ROCs from PPDM with NCC (43.02% VR at 0.1% FAR) versus PPDM with a CF (51.59% VR at 0.1% FAR) on the FOCS dataset (left ocular).

Since the patch templates need to be learned for every match pair, it is desirable to efficiently train the templates. Solving for the template in Eq. 2 requires the inverse of a rather large but sparse matrix $\hat{\mathbf{T}} \in \mathbb{C}^{KM \times KM}$ with a special block structure where each block is a diagonal matrix. By leveraging this structure we recursively invert this matrix blockwise using the schur complement which can be computed efficiently since the blocks are diagonal matrices.

Each output $\mathbf{C}_i \in \mathbb{R}^{a \times b}$ (where $a \times b$ is the size of a single patch from the probe image, M total pixels) ideally exhibits a sharp peak at the location of the best match thereby providing an estimate of the relative shift (x, y) between the probe and gallery patches. The lateral shift from the center of the correlation output represents the local deformation, $\mathbf{d}$ (location of the peak from center), while the peak height is defined as the similarity score, $S(\mathbf{I}, \mathbf{G}; \mathbf{d})$ (similarity of corresponding patches of $\mathbf{I}$ and $\mathbf{G}$ at $\mathbf{d}$), as illustrated in Fig. 8.

Since the role of the template matcher in our framework is to provide both similarity and deformation cues (modeled as spatial translations, see Section V for details), it is important that the template is designed for accurate pattern localization in addition to pattern similarity. We demonstrate the effectiveness of our CF based template matching scheme, which is specifically optimized for pattern localization, by comparing it to a simple normalized cross-correlation (NCC), which is not specifically optimized for pattern localization between the probe and gallery features. Fig. 7a shows the response of a gallery CF template and a gallery NCC template to the same authentic probe image patch. The CF template output exhibits a distinct sharp peak at the location of the best match, while the NCC output is much less sharp. Since the template not only provides the similarity cues but the deformation cues as well, sharper peaks lead to more robust matching as seen by the receiver operating characteristic (ROC) curves in Fig. 7b.

V. Deformation Model

The deformation that each pixel in the probe image can undergo relative to the gallery image is often related to the deformations of the other neighboring pixels in the probe. Rather, the groups of pixels that move similarly are correlated. In this paper, we approximate this relative deformation between corresponding rectangular

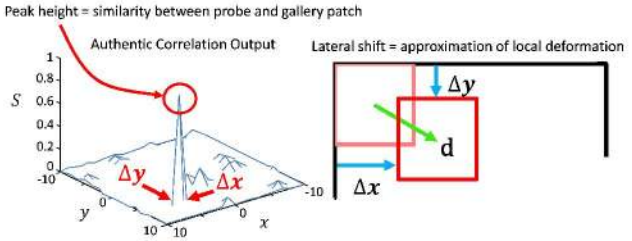


Fig. 8. Synthetic example output from comparing an authentic probe image patch with the corresponding gallery image patch. The peak height represents the similarity, and the lateral shift (Δx and Δy) represents the deformation.

probe and gallery patches by a local spatial translation. The estimated deformations for authentic match pairs are expected to be more consistent compared to those from impostor match pairs. This expected consistency in the spatial translations across the different patches for authentic match pairs is captured as a prior using a Gaussian Markov Random Field (GMRF) model which is learned from training data. Given a pair of images we regularize the observed relative deformations of the corresponding patches by a prior which captures the space of valid deformations for authentic match pairs.

A. MAP Estimation

Both the $\mathbf{I}$ and $\mathbf{G}$ are divided into N non-overlapping rectangular patches, and for the $i-$th patch pair the relative spatial translation is denoted by $\{(\Delta x_i, \Delta y_i)\}$ as illustrated in Fig. 8. We define the deformation for the image pair as a vector $\mathbf{d} = [\triangle x_1, \Delta y_1, \cdots, \triangle x_N, \Delta y_N]^T$.

We compute the underlying latent deformation using maximum-a-posteriori (MAP) estimation. For an authentic match pair $\mathbf{I}$ and $\mathbf{G}$ we estimate the most likely deformation vector between the two images:

$$\begin{aligned}\hat{\mathbf{d}} &= \underset{\mathbf{d}}{\operatorname{argmax}}\, P(\mathbf{d}|\mathbf{G}, \mathbf{I}) \\ &= \underset{\mathbf{d}}{\operatorname{argmax}}\, P(\mathbf{I}|\mathbf{G}, \mathbf{d})\, P(\mathbf{d}|\mathbf{G}) \\ &= \underset{\mathbf{d}}{\operatorname{argmax}}\, P(\mathbf{I}|\mathbf{G}, \mathbf{d})\, P(\mathbf{d}) \\ &\approx \underset{\mathbf{d}}{\operatorname{argmax}} \left\{ P(S(\mathbf{I}, \mathbf{G}; \mathbf{d})) \cdot \exp\left(-\frac{1}{2}\mathbf{d}^T \boldsymbol{\Sigma}_{\mathbf{d}}^{-1} \mathbf{d}\right) \right\} \end{aligned} \quad (4)$$

where the likelihood $P(\mathbf{I}|\mathbf{G}, \mathbf{d})$ is modeled by a distribution over the patch similarity scores $P(S(\mathbf{I}, \mathbf{G}; \mathbf{d}))$ and the prior probability distribution, $P(\mathbf{d})$, is modeled as a GMRF, with mean $\mathbf{0}$ (after accounting for global relative shift between the probe and gallery) and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{d}} \in \mathbb{R}^{2N \times 2N}$.

B. Likelihood Model

In [27] the likelihood $P(\mathbf{I}|\mathbf{G}, \mathbf{d})$ is modeled as,

$$P(\mathbf{I}|\mathbf{G}, \mathbf{d}) = \prod_{i=1}^{N} P(S_i(\mathbf{I}, \mathbf{G}; \mathbf{d}_i)) \quad (5)$$

where the likelihood is approximated by the similarity score $S(\cdot)$ between the probe and gallery for computational tractability. The similarity score, $S(\mathbf{I}, \mathbf{G}; \mathbf{d})$, is obtained from template matching across all spatial translations as shown in Fig. 8 between $\mathbf{I}$ and $\mathbf{G}$ for

a relative distortion parameter $\mathbf{d}$, and is modeled by a single normal distribution:

$$P(\mathbf{I}|\mathbf{G},\mathbf{d}) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{d}_i, \mu, \sigma) \tag{6}$$

While modeling the similarity using the same distribution across the entire image is sufficient when multiple samples are available for training, it does not perform well in 1 : 1 matching scenarios. More specifically, this method does not perform well across the periocular region where the amount of discriminative texture for each patch varies significantly. Therefore to allow for the spatial variability in the distribution of the similarity scores, we relax the likelihood model and allow the distribution of the similarity scores to vary across the image, a process we refer to as 'Spatially varying Likelihood' (SVL), i.e.,

$$P(\mathbf{I}|\mathbf{G},\mathbf{d}) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{d}_i, \mu_i, \sigma_i) \tag{7}$$

As described later in Section VII, the likelihood is used to modify the correlation outputs from each comparison. Incorporating a distribution per patch allows for the model to account for the discriminative capability of each individual region.

C. Deformation Prior

With each deformation variable representing an image region, the rectangular patch segmentation we employ in this paper will correspond to a rectangular graph structure. We model the prior $P(\mathbf{d})$ by a GMRF:

$$P(\mathbf{d}) = \frac{\prod_{(i,j)\in\varepsilon} \Psi_{\mathbf{d}}(\mathbf{d}_i, \mathbf{d}_j)}{\sum_{\mathbf{d}} \prod_{(i,j)\in\varepsilon} \Psi_{\mathbf{d}}(\mathbf{d}_i, \mathbf{d}_j)} = \frac{1}{Z_{\mathbf{d}}} \exp\left(-\frac{1}{2}\mathbf{d}^T \boldsymbol{\Sigma}_{\mathbf{d}}^{-1} \mathbf{d}\right) \tag{8}$$

where $\Psi_{\mathbf{d}}(\mathbf{d}_i, \mathbf{d}_j)$ is the potential function relating the deformations between neighboring nodes, ε denotes the set of all the edges in the graph, and $Z_{\mathbf{d}}$ is the partition function. For GMRFs [29], [52], the potential functions can be parametrically defined as follows while retaining the structure of the precision matrix $\boldsymbol{\Sigma}_{\mathbf{d}}^{-1}$ (where $\boldsymbol{\Sigma}_{\mathbf{d}_{ij}}^{-1} = 0$ for non-neighboring $\mathbf{d}_i$ and $\mathbf{d}_j$):

$$\Psi_{\mathbf{d}}(\mathbf{d}_i, \mathbf{d}_j) = \exp\left[-\frac{1}{2}\left(\alpha \|\mathbf{d}_i\|^2 + \alpha \|\mathbf{d}_j\|^2 + \beta \|\mathbf{d}_i - \mathbf{d}_j\|^2\right)\right] \tag{9}$$

where the α parameter captures the individual relative node deformations and the β parameter captures the consistency (smoothness) of the deformations between neighboring patches. These parameters are encoded into the sparse precision matrix, setting $-\beta$ on the off diagonals (of connecting nodes) and $\kappa_j(\alpha+\beta)$ on the diagonal (κ_j is the number of neighbors for node j).

As the distribution of the $i-$th deformation field vector is Gaussian, the deformation sub-vectors $\mathbf{x}$ and $\mathbf{y}$ are also Gaussian. Defining separate independent potentials for the $\mathbf{x}$ and $\mathbf{y}$ deformation components:

$$\Psi_{\mathbf{dx}}(\mathbf{dx}_i, \mathbf{dx}_j) = \exp\left[-\frac{1}{2}\left(\alpha_x \|\mathbf{dx}_i\|^2 + \alpha_x \|\mathbf{dx}_j\|^2 + \beta_x \|\mathbf{dx}_i - \mathbf{dx}_j\|^2\right)\right] \tag{10}$$

where the potential function for the $\mathbf{y}$ component can be defined similarly.

The resulting prior from decoupling the $\mathbf{x}$ and $\mathbf{y}$ components is:

$$\begin{aligned} P(\mathbf{d}) &= \frac{\prod_{(i,j)\in\varepsilon} \Psi_{\mathbf{dx}}(\mathbf{dx}_i, \mathbf{dx}_j)\Psi_{\mathbf{dy}}(\mathbf{dy}_i, \mathbf{dy}_j)}{\sum_{\mathbf{dx}}\sum_{\mathbf{dy}} \prod_{(i,j)\in\varepsilon} \Psi_{\mathbf{dx}}(\mathbf{dx}_i, \mathbf{dx}_j)\Psi_{\mathbf{dy}}(\mathbf{dy}_i, \mathbf{dy}_j)} \\ &= \frac{1}{Z_{\mathbf{d}}} \exp\left(-\frac{1}{2}\left(\mathbf{dx}^T \boldsymbol{\Sigma}_{\mathbf{dx}}^{-1} \mathbf{dx} + \mathbf{dy}^T \boldsymbol{\Sigma}_{\mathbf{dy}}^{-1} \mathbf{dy}\right)\right) \end{aligned} \tag{11}$$

The $\mathbf{x}$ and $\mathbf{y}$ components of the deformation vector, $\mathbf{d}$, are separable allowing for the potential defining the coupled model, $\Psi_{\mathbf{d}}$, to be decomposed as the product of the potentials for each component.

Just as we vary the likelihood spatially we can also vary α and β spatially. This allows for training individual parameters for each node and edge weight to characterize the graph. In Eq. 12 and Eq. 13 we define potentials that separate each individual α_i and β_{ij} for coupled and decoupled models, respectively:

$$\begin{aligned} \Psi_{\mathbf{d}_{ij}}(\mathbf{d}_i, \mathbf{d}_j) = \exp\Big[& \\ & -\frac{1}{2}\left(\alpha_i \|\mathbf{d}_i\|^2 + \alpha_j \|\mathbf{d}_j\|^2 + \beta_{ij} \|\mathbf{d}_i - \mathbf{d}_j\|^2\right)\Big] \end{aligned} \tag{12}$$

$$\begin{aligned} \Psi_{\mathbf{dx}_{ij}}(\mathbf{dx}_i, \mathbf{dx}_j) = \exp\Big[& -\frac{1}{2}\Big(\alpha_{x|i} \|\mathbf{dx}_i\|^2 \\ & + \alpha_{x|j} \|\mathbf{dx}_j\|^2 + \beta_{x|ij} \|\mathbf{dx}_i - \mathbf{dx}_j\|^2\Big)\Big] \end{aligned} \tag{13}$$

VI. Parameter Estimation

Our proposed model as described in Section V requires the following set of parameters: $\{(\alpha_1, \cdots, \alpha_N), (\beta_1, \cdots, \beta_{|\varepsilon|})\}$ and $\{(\mu_1, \cdots, \mu_N), (\sigma_1, \cdots, \sigma_N)\}$, where $|\varepsilon|$ is the cardinality of the set of all edges in the graph.

A. Likelihood Parameters

The parameters $\{(\mu_1, \cdots, \mu_N), (\sigma_1, \cdots, \sigma_N)\}$ model the similarity scores obtained from matching the gallery patch template to the corresponding probe patch. We learn the parameters via maximum likelihood (ML) estimation from training data, i.e., we estimate the empirical mean and variance of the normal distributions for every patch from the set of similarity values obtained from a training set of authentic match pairs.

B. Prior Parameters

To learn the deformation parameters we need to know the ground truth deformations between the corresponding gallery and probe patches. However, observing the ground truth deformations is not possible and manually labeling the deformations is too difficult. Therefore, [52] treats the ground truth deformations as latent variables and estimates the model (prior) parameters via a computationally expensive and iterative expectation maximization (EM) based approach. While guaranteed to converge to some local maxima, the EM procedure depends highly on initial starting points, often requiring repeating the procedure multiple times to obtain parameter values that fit well to the data. Instead we found that simply treating the deformations estimated from the likelihood (correlation outputs) as the

Algorithm 1 GMRF: Parameter Learning

Require: Empirical Covariance: $\hat{\mathbf{S}}$
1: Initialize: $\mathbf{W} = \hat{\mathbf{S}}$
2: **while** Repeat for $k = 1, 2, \ldots, N, 1, 2, \ldots, N, \ldots$ until convergence **do**
3: Partition $\mathbf{W} = \begin{bmatrix} \mathbf{W_{11}} & \mathbf{w}_{12} \\ \mathbf{w}_{12}^T & w_{22} \end{bmatrix}$ {where $\mathbf{w}_{12} = \mathbf{W}_{[1:N-1,k]}$ and $w_{22} = W_{kk}$}
4: Constrain: $\mathbf{W}_{11}^* = \mathbf{W}_{11}$ and $\mathbf{s}_{12}^* = \mathbf{s}_{12}$ s.t. $\mathbf{W}_{11}^*(i,j) = 0$ and $\mathbf{s}_{12}^*(i,j) = 0 \;\; \forall (i,j) \notin \varepsilon$
5: Solve $\hat{\rho}^* = (\mathbf{W}_{11}^*)^{-1}\mathbf{s}_{12}^*$
6: Update: $\mathbf{w}_{12} = \mathbf{W}_{11}^*\hat{\rho}^*$
7: **end while**
8: Invert result $\mathbf{Z} = \mathbf{W}^{-1}$
9: Extract parameter values from $\mathbf{Z}$

ground truth deformations and learning the deformation parameters $\{(\alpha_1, \cdots, \alpha_N), (\beta_1, \cdots, \beta_{|\varepsilon|})\}$ via linear regression can help us learn deformation parameters which provide significant performance gains over the EM based approach presented in [52]. The result is a much simpler model for better generalization while being orders ($\geq 100\times$) of magnitude faster to learn since inference, whose computational complexity grows linearly with the number of training instances, is no longer needed to be performed during training.

Recall that for GMRFs the specific structure of the precision matrix, $\mathbf{\Sigma_d}^{-1}$, carries the organization of the model as zero terms for nodes not connected by an edge in the graph ($\mathbf{\Sigma}_{\mathbf{d}_{ij}}^{-1} = 0$ for $(i,j) \notin \varepsilon$, [29]). Methods like Graphical LASSO [53] use this relationship to estimate a meaningful (sparse) graph structure by maximizing the log-likelihood of the precision matrix. In our problem, we use a pre-determined grid structure and manipulate the Graphical LASSO problem to impose this structure as the specific sparsity constraints to solving the ML estimation problem. We first estimate the covariance matrix to derive the precision matrix and then derive the model parameters from the result. Given the empirical covariance matrix, $\hat{\mathbf{S}}$, we estimate the precision matrix, Δ, by solving:

$$\begin{aligned} \max_{\Delta} \quad & log|\Delta| - trace(\hat{\mathbf{S}}\Delta) \\ s.t. \quad & \delta_{ij} = 0 \quad \forall (i,j) \notin \varepsilon \end{aligned} \tag{14}$$

where $log|\Delta| - trace(\hat{\mathbf{S}}\Delta)$ represents the log-likelihood of the data and $\delta_{ij} = 0$ are the constraints enforcing the graph (connectivity) structure in the precision matrix. Decoupling the parameters, we first solve for the $\mathbf{x}$ component, $\hat{\mathbf{S}} \triangleq \hat{\mathbf{\Sigma}}_{\mathbf{dx}}$, and then solve for the $\mathbf{y}$ component, $\hat{\mathbf{S}} \triangleq \hat{\mathbf{\Sigma}}_{\mathbf{dy}}$, to derive the final graph parameters, where by Eq. 11, $\alpha_i = \alpha_{x|i} + \alpha_{y|i}$ and $\beta_{ij} = \beta_{x|ij} + \beta_{y|ij}$.

By the use of Alg. 1 the resulting sparse matrix $\mathbf{Z}$ has non-zero terms for neighboring nodes consistent with the structure of the graph. Thus, extracting the individual α and β parameter values draws directly from the precision matrix structure, where $\beta_{ij} = \mathbf{Z}_{ij}$ and $\alpha_i = \frac{\mathbf{Z}_{ii} - \sum_{k \in \pi} \beta_k}{\kappa_i}$ (π is the set of neighbors for node i, and $\kappa_i = |\pi|$ is the cardinality) for $i = 1, \ldots, N$.

The deformations exhibited by an authentic match pair are expected to be more coherent than the deformations exhibited by an impostor match pair and a single α and β may not be sufficient to capture the coherency in the deformations, specifically in challenging environments. Nonetheless we provide a simple method for extracting a single α_x and β_x model parameter set from the estimated precision matrix $\mathbf{Z}$ via a regression formulation (can also be used for the $\mathbf{y}$ component):

$$\min_{\Phi} \quad \|\mathbf{b} - \mathbf{A}\Phi\|_2^2 + \gamma\|\Phi\|_2^2 \tag{15}$$

where $\mathbf{b} = \begin{bmatrix} \mathbf{Z}_{11} \\ \vdots \\ \mathbf{Z}_{NN} \\ \mathbf{Z}_{\varepsilon(1)} \\ \vdots \\ \mathbf{Z}_{\varepsilon(M)} \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} \kappa_1 & \kappa_1 \\ \vdots & \vdots \\ \kappa_N & \kappa_N \\ 0 & -1 \\ \vdots & \vdots \\ 0 & -1 \end{bmatrix}$, $\Phi = \begin{bmatrix} \alpha_x \\ \beta_x \end{bmatrix}$, γ is a regularization parameter to prevent over-fitting, and M is the cardinality of ε (here $\mathbf{Z}_{11}$ represents the value located at the first row and first column of the matrix $\mathbf{Z}$, while $\mathbf{Z}_{\varepsilon(1)}$ denotes value that represents the first (i,j) pair in the set of neighboring nodes and $\mathbf{Z}_{\varepsilon(M)}$ denotes the last).

VII. Inference

Given the observations, i.e., correlation planes from each patch comparison and the learned GMRF, we estimate the deformations between the patches via inference on the GMRF. We employ Iterated Conditional Modes (ICM) over other methods due to its speed advantage. While ICM has become less popular due to its sensitivity to the initial estimate [54], it has been found to be good for our purposes when initialized with the correlation outputs, $\mathbf{C}_k$ $(k = 1, 2, ..N)$, considering the correlation filter based templates provide good deformation estimates. Since we observed empirically that ICM converges in fewer than four iterations, we perform a maximum of four iterations of ICM. The result is a set of outputs representing the marginal probability distribution under the assumption that this is an authentic match pair. The final match score is computed as the sum of the individual patch scores,

$$M = \sum_{k=1}^{N} \langle \mathbf{S}_k, \varphi_k \rangle \tag{16}$$

where each patch score is the vectorized inner products of the likelihood, $\mathbf{S}_k \in \mathbb{R}^{a \times b}$ (where $a \times b$ is the dimensionality of a single patch from the probe image), with the marginal posterior distribution of deformation given the observed evidence from correlation, $\varphi_k \in \mathbb{R}^{a \times b}$. The likelihood is determined by modifying $\mathbf{C}_k$:

$$\mathbf{S}_k = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\mathbf{C}_k - \mu_k}{\sigma_k}\right)\right) \tag{17}$$

where $\operatorname{erf}(\cdot)$ is the Gaussian error function, and the mean (μ_k) and variance (σ_k) are learned from training.

VIII. Experimental Results

We evaluate Probabilistic Deformation Models (PDMs) against many other 1 : 1 image matching techniques for matching deformed patterns via extensive and large scale experiments on periocular recognition.

Fig. 9. An example of UBIRIS v2.0 intraclass images spanning five distances (columns: 4 to 8 meters) and two imaging sessions (rows: S1 and S2).

A. Datasets

We now briefly describe the datasets introduced in Section I that are used to evaluate the efficacy of PPDM in comparison to other image matching algorithms.

- **Face and Ocular Challenge Series (FOCS) [30]:** The periocular data in the FOCS dataset (Fig. 3 and 4a) consists of still images of a single eye region at a resolution of 600×750 pixels with smaller iris regions, specular reflections, and out of focus blurring. The database has 9581 images (4792 left, 4789 right) of 136 subjects with between 2 and 236 images per subject. These regions were extracted from near infrared (NIR) video sequences collected from moving subjects[2] in an unconstrained environment exhibiting drastic variations in sensor noise, illumination, gaze, and occlusion. We pre-process the FOCS images by a simple histogram normalization to compensate for harsh illumination after centering on the pupil.
- **BEST Development Challenge Problem (BDCP) [32]:** As stated in Section I, the BDCP dataset[3] is composed of images captured from two different sensors, the LG4000 [33] and CFAIRS [34] cameras. Images from the BDCP set, as shown in Fig. 4b, appear more suitable for iris recognition with images at resolutions of 480×640 pixels and 600×750 pixels for the LG4000 and CFAIRS sensors respectively. The dataset consists of a total of 4314 images, 2577 images (1288 left, 1289 right) from the CFAIRS sensor of 82 subjects, and 1737 images (869 left, 868 right) from the LG4000 sensor of 99 subjects (subjects are shared between the two sensors for a total 99 subjects). The CFAIRS set has between 6~60 samples/subject (31 samples/subject on average) while the LG4000 set has between 5~24 samples/subject (17 samples/subject on average). We pre-process the BDCP images by cropping the images at the eye corners to account for the scale variation between the LG4000 and CFAIRS sensors (this is the *only* dataset in which scale is normalized during matching as the objective is to evaluate cross sensor performance).
- **Face Recognition Grand Challenge (FRGC) [36]:** This dataset is composed of periocular regions manually segmented[4] from a subset of images from the FRGC Database (Fig. 4c). This set consists of 2272 images (1136 left, 1136 right) corresponding to a total of 568 unique subjects (2 samples per eye per subject) with a resolution of 241×226 pixels. The first image of every subject is the gallery entry, while the second is the probe. This experimental setup was used in [1] (EER: 1.59% left, 1.93% right) and [57] (EER: 6.96% from averaging the left and right periocular results). Note [8], [58] use a different subset of images and hence the results are not comparable.
- **University of Beira Interior Iris (UBIRIS) v2.0 [31]:** This dataset is built for the purpose of simulating a less constrained acquisition of visible wavelength iris images and there has been some work [8], [12], [59], [60] which evaluates the feasibility of its use for periocular recognition. In total, 11102 images are acquired in unconstrained conditions (see Fig. 9) from 261 subjects over two sessions and five distances (4 to 8 meters). Unfortunately many of the UBIRIS images do not contain eyebrows or other parts of the periocular region. Provided the many efforts [7], [12], [48], [57]–[59] which demonstrate the discriminative power of including the eyebrows and surrounding skin texture when performing periocular recognition with visible light images, we only evaluate images from distances of 6m, 7m, and 8m using the experimental setup from [12] (however we don't match individual distances, e.g., 6m vs 6m, 7m vs 7m, and 8m vs 8m, rather use 'Experiment 3', comparing images at all distances 6m-8m vs images at all distances 6m-8m). Note that [8], [59], [60] each use separate and unique subsets[5] collected from all distances (4m-8m) and hence results are not comparable.
- **University of Beira Interior Periocular (UBIPr) [35]:** Designed as a dataset for periocular recognition experiments in uncontrolled acquisition environments, the images (Fig. 4c) are captured with a visible light sensor and vary in pose, gaze, stand-off distance, and illumination. Composed of 10252 (5126 left, 5126 right) periocular images over 344 subjects, the pose variation in the images include frontal and both side views and the stand-off distance varies from 8 meters to 4 meters with varying resolution: 501×401 pixels for 8m, to 1001×801 pixels for 4m. Unlike UBIRIS data, the images from all stand-off distances in the UBIPr set contain the full periocular region (including eyebrows and surrounding cheek information), thus performance is evaluated over the entire dataset.

The FOCS set can be considered as a benchmark for a low quality, high distortion dataset while the

[2]Frames are extracted from Iris on the Move [55] videos of the Multiple Biometric Grand Challenge (MBGC) Portal Track V2 dataset [56].

[3]The BDCP data served as a cross sensor challenge set for performers in the ocular track of the Intelligence Advanced Research Projects Activity (IARPA) Biometrics Exploitation Science and Technology (BEST) program.

[4]Coordinates were provided by the authors of [57].

[5] [59] uses 2340 images (separated by match pair rather than by image), [60] uses 480 images (400 gallery, 80 probe), and [8] uses 1000 images (96 for training and 904 for evaluation).

BDCP set exhibits less distortion than FOCS (no user movement or gaze variation) over different sensors in a relatively benign environment. The UBIPr and UBIRIS v2.0 datasets supplement the FOCS data by offering higher quality unconstrained images at varying stand-off distances, and the FRGC set has little to no visible distortion with good image quality.

B. Baseline Algorithms

We now briefly describe the baseline algorithms that we compare to PPDM on the periocular datasets. We also provide comprehensive timing metrics in Table VII for each algorithm presented.

- **LBP-GIST (GLBP)** [12]: We implemented the fusion (computed as a weighted sum of the chi-squared distance) of LBP and GIST features to capture both texture and shape.
- **LBP-HOG-SIFT (LHS)** [24], [35], [57]: We implemented the fusion (using logistic regression) of local binary patterns (LBP), histogram of oriented gradients (HOG),[6] and dense SIFT (DSIFT) features (using the VLFeat library [61]). For the LBP and HOG feature sets, we use an 8 pixel cell size and determine match scores using city block (LBP) and euclidean distances (HOG and DSIFT).
- **Modified SIFT (m-SIFT)** [1]: Matching is performed using a modified version of the standard SIFT features. The keypoints to match are spatially constrained to lie in the same region for the gallery and the probe image.
- **Full Correlation (F-CF):** Each image is directly compared using the technique described in Section IV, however, each image is not divided into patches, rather, a template is built using the entire image after feature extraction using Gabor filters (see Fig. 6a).
- **SIFT-Flow (SFlow)** [26]: The alignment between the probe and gallery periocular images is estimated by matching DSIFT features. The pixel-wise motion field is estimated via energy minimization similar to optical flow. While SIFT-Flow can provide accurate pixel-wise motion estimation between the probe and gallery it is computationally expensive as is evident from Table VII.

C. Evaluation

Each method is evaluated in a $1:1$ image-to-image matching scenario on each periocular dataset. The verification rates (VRs), computed as 1 - the False Reject Rate (FRR) at 0.001 False Acceptance Rate (FAR), of the best performing experiments are shown in Table I and Table II with the corresponding receiver operating characteristic (ROC) curves in Fig. 10 (equal error rates are also displayed in the legend of each ROC).

In each of the tests presented using the FOCS, BDCP, and UBIPr datasets we use 5-fold cross validation reporting overall VRs at 0.001 FAR (from the scores obtained by concatenating the associated folds) as a measure of system performance. Within each fold, each image is independently compared against another (left vs left and right vs right) to generate score matrices for the left and right periocular matches from each region selection (excluding self-comparisons). Thus, when constructing the correlation filters for determining the deformation cues between a probe and gallery, only a single image is used for building the filter with no other authentic or impostor images included (we set $\lambda = 10^{-5}$ for all experiments that use a correlation filter).

Each of the baseline algorithms also use the same 5-fold cross validation paradigm for a fair comparison over varying periocular image sizes: 64×64, 128×128, 256×256, and 384×384 pixels, concatenating the match scores from the same test sets used in the PDM experiments. By running the experiments over varying image resolutions (and reporting the best results) we remove any potential performance bias based on the dimensionality of the image. Our results indicate that images up to size 384×384 pixels have sufficient detail to capture the underlying periocular features as better performing baselines obtained best results at a lower resolution, e.g., m-SIFT and SIFT-Flow usually had best results at 256×256 pixel resolutions.

The FRGC experiments are performed such that the first image sample for each subject is the gallery image and the remaining images are the probe images to create a 568×568 score matrix. For FRGC experiments using PDM and LHS (regression weights for score fusion) methods, training is performed with *only* frontal view images[7] from UBIPr.

Finally experiments on the UBIRIS v2.0 dataset only evaluate images from distances of 6m, 7m, and 8m, designating the first two images per distance as the gallery set and the remaining per distance as the probe set. This match scenario is also performed by [12] (however we don't match individual distances, e.g., 6m vs 6m, etc., rather use 'Experiment 3', comparing images at all distances 6m-8m vs images at all distances 6m-8m) who report a rank-1 identification accuracy of ~17% (obtained from the CMC curve) after fusing the results from both periocular regions.

D. Results

Table I summarizes the best result for each of the baselines we compare PDM based algorithms (PDM [27], PDM with SVL, and PPDM) against for $1:1$ image matching. Of the five baselines, we observe that SIFT-Flow, which estimates pixel level pattern deformation, performs the best on FOCS, FRGC, UBIRIS, and UBIPr datasets, while m-SIFT performs best on BDCP, closely followed by full image correlation (F-CF). In general, as expected, we observe that algorithms which are tolerant to pattern distortions, like SIFT-Flow, perform better than algorithms which are

[6] Termed gradient orientation histograms (GOH) in [57].

[7] By using only the frontal views at all stand-off distances in the UBIPr dataset we collect 3418 (1709 left, 1709 right) periocular images (~5 samples/subject)

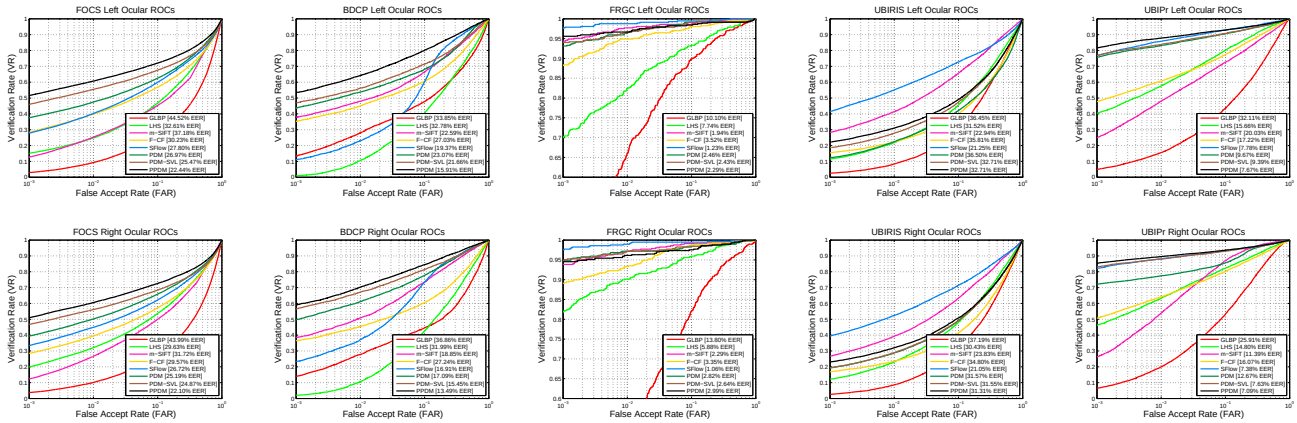


Fig. 10. Best baseline and PDM/PPDM ROC curves for left and right periocular regions on each dataset. Corresponding equal error rates (EER) are displayed in the legend for each method.

TABLE I
LEFT (L) AND RIGHT (R) PERIOCULAR RECOGNITION BASELINE VRS AT 0.001 FAR.

		FOCS	BDCP	FRGC	UBIRIS	UBIPr
GLBP	R	3.59%	13.91%	22.54%	2.39%	6.27%
	L	2.89%	13.25%	30.63%	2.45%	4.68%
LHS	R	19.83%	2.02%	81.69%	12.10%	46.22%
	L	15.10%	0.94%	69.89%	11.50%	40.39%
m-SIFT	R	12.21%	37.71%	93.66%	26.29%	20.11%
	L	12.67%	37.56%	94.37%	27.99%	24.26%
F-CF	R	28.43%	36.30%	89.08%	16.85%	50.70%
	L	28.25%	35.21%	88.20%	15.42%	47.67%
SFlow	R	33.39%	11.18%	97.54%	39.39%	82.08%
	L	27.73%	23.26%	97.54%	41.56%	76.08%

TABLE II
LEFT (L) AND RIGHT (R) PERIOCULAR RECOGNITION VRS AT 0.001 FAR FROM PDM/PPDM ON EACH DATASET.

		FOCS	BDCP	FRGC	UBIRIS	UBIPr
PDM	R	39.43%	49.64%	94.72%	19.27%	72.20%
	L	37.55%	43.53%	93.13%	12.24%	75.59%
PDM w/ SVL	R	46.75%	56.57%	94.89%	19.37%	83.03%
	L	45.95%	46.94%	94.01%	18.66%	76.82%
PPDM	R	51.00%	58.96%	94.54%	23.07%	85.28%
	L	51.59%	53.51%	95.60%	22.18%	81.49%

not explicitly designed to be tolerant to pattern distortions, such as GLBP based nearest neighbor classifiers.

Fig. 10 displays the ROCs which also supports our assessment in Section VIII-A about the difficulty of these datasets. The overall effectiveness of each baseline algorithm demonstrates that the data quality in the UBIPr and FRGC datasets allow for good recognition rates, while data that is prone to more distortion between the subject samples, such as varying the sensor and/or time of acquisition as described in the BDCP, FOCS, and UBIRIS sets, degrades the system performance.

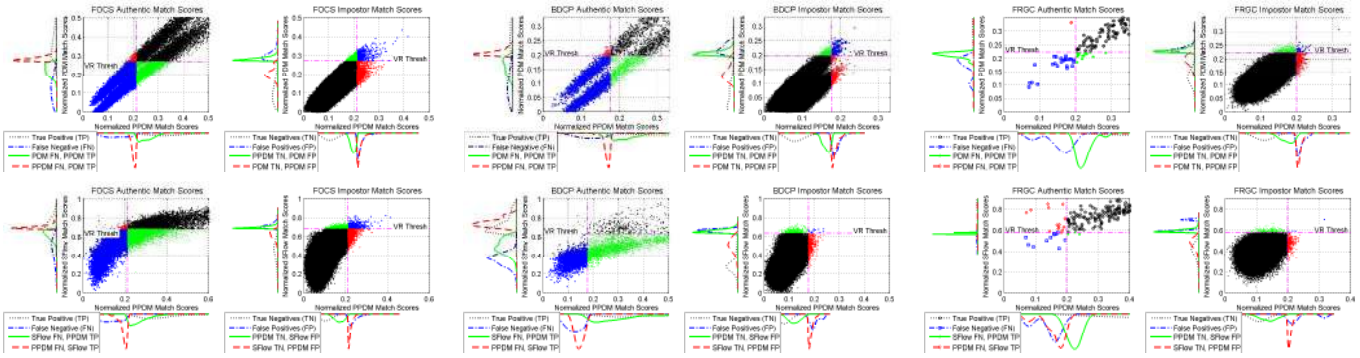
The best results for PDM and PPDM on ocular datasets occurs at 128×128 pixel images with a 6×6 patch configuration. Table II summarizes results from PDM and PPDM over each dataset. The experimental results show that the proposed PPDM design outperforms both the original PDM method as well as the baselines by a significant margin on the FOCS, BDCP, and UBIPr datasets.

SIFT-Flow performs best on the FRGC dataset with a VR of 97.54% for both the left and right eyes respectively with PPDM achieving a comparable VR of 95.60% and 94.54% for the left and right eye. On the UBIRIS dataset the VR improves from 12.24% left and 19.27% right with PDM to 22.18% left and 23.07% right when using PPDM, however underperforms against m-SIFT and SIFT-Flow (which are deliberately designed for *large* scale changes) with the best performance as 41.56% on the left and 39.39% on the right regions.

On the FOCS dataset the VR improves from baselines of 28.25% left (F-CF) and 33.39% right (SIFT-Flow) to 51.59% left and 51.00% right when using PPDM. The baseline for BDCP of 37.56% VR for the left eye and 37.71% VR for the right eye from m-SIFT, respectively, improves to 53.51% and 58.96% for the left and right eye respectively with PPDM. Finally, on the UBIPr dataset the VR improves from 76.08% left and 82.08% right with SIFT-Flow to 81.49% left and 85.28% right when using PPDM. This indicates that there is a consistent performance increase from PPDM over PDM (by ~30%) and all baselines (by ~40%) across the FOCS, BDCP, and UBIPr datasets (SIFT-Flow outperforms PPDM on the UBIRIS and FRGC datasets). The displayed results demonstrate the significance of the PPDM method on challenging data and match scenarios (1 : 1).

IX. DISCUSSION

In this section we analyze PPDM results and comment on some additional properties of the proposed algorithm. Fig. 11 displays scatter plots with density information (outside curves) comparing the authentic and impostor match scores from PPDM with PDM (top) and SIFT-Flow (bottom) for the FOCS, BDCP, and FRGC datasets (left periocular region). Each plot separates the cluster into four regions based on the match score threshold which determines the verification rate (VR) at 0.001 FAR. The regions with points colored **blue** and **black** indicate match scores for



(a) FOCS (b) BDCP (c) FRGC

Fig. 11. Scatter plots with density information compare the authentic and impostor match scores from PPDM with PDM (top) and SIFT-Flow (bottom) for the FOCS, BDCP, and FRGC datasets. Each plot separates the cluster into four regions based on the match score threshold which determines the verification rate (VR) at 0.001 FAR. The **blue** and **black** regions indicate match scores for which *both algorithms* perform the same classification, the **green** regions indicate *only* PPDM performing a correct classification, and **red** regions indicating a correct classification from *only* PDM or SIFT-Flow, respectively.

TABLE III
COMPARING ALGORITHM SPECIFIC FALSE NEGATIVES (FN) FROM AUTHENTIC MATCH SCORE CLUSTERS.

	PPDM FN vs PDM FN		PPDM FN vs SFlow FN	
	# of samples	change	# of samples	change
FOCS	472 vs 7752	16.42×	272 vs 12646	46.49×
BDCP	102 vs 1206	11.82×	12 vs 4694	391.17×
FRGC	1 vs 15	15.00×	14 vs 3	0.21×

TABLE IV
COMPARING ALGORITHM SPECIFIC FALSE POSITIVES (FP) FROM IMPOSTOR MATCH SCORE CLUSTERS.

	PPDM FP vs PDM FP		PPDM FP vs SFlow FP	
	# of samples	change	# of samples	change
FOCS	3174 vs 3162	1.00×	3794 vs 3780	1.00×
BDCP	562 vs 562	1.00×	888 vs 888	1.00×
FRGC	281 vs 280	1.00×	318 vs 318	1.00×

which *both algorithms* have the same classification result (**both correct** or **both incorrect**).[8] While the regions with points colored **green** indicate where *only* PPDM correctly classifies the probe image to the gallery with the remaining region (**red** points) indicating a correct classification from *only* PDM or SIFT-Flow, respectively. We compare the score distributions of PPDM, PDM, and SIFT-Flow to provide an indication of where competing methods fail and where PPDM succeeds. Specifically we compare authentic and impostor match score errors, i.e., false negatives (FN) and false positives (FP) displayed in the **red** and **green** regions. We chose to include SIFT-Flow since the method is the most effective baseline (outperforming PPDM on FRGC).

From Tables III and IV and the graphs in Fig. 11 we observe that PPDM is more effective at decreasing the number of false negatives (average change is >50×) than reducing the number of false positives (~1.00×). That is, in these instances we observe quantitatively that PPDM succeeds by being more effective at identifying an authentic match (reducing the false accept rate). This result is verified by computing the statistical significance of PPDM's improved performance when

[8]We focused the plots at the crossing point of both VR thresholds thus possibly cutting off part of the blue and black clusters.

TABLE V
COMPARING THE NUMBER OF CORRECT AUTHENTIC CLASSIFICATIONS FROM PPDM AND PDM-SVL WHEN PDM PRODUCES A FALSE NEGATIVE.

	only PPDM is correct	*only* PDM-SVL is correct	*only* PDM errors
FOCS	2844 (32.76%)	56 (0.65%)	4908 (56.53%)
BDCP	540 (32.03%)	56 (3.32%)	666 (39.50%)
FRGC	10 (62.50%)	0 (0.00%)	5 (31.25%)

# of samples (% of authentic matches not sharing a classification)

TABLE VI
COMPARING THE NUMBER OF CORRECT IMPOSTOR CLASSIFICATIONS FROM PPDM AND PDM-SVL WHEN PDM PRODUCES A FALSE POSITIVE.

	only PPDM is correct	*only* PDM-SVL is correct	*only* PDM errors
FOCS	1958 (27.77%)	222 (3.15%)	1204 (17.08%)
BDCP	308 (18.92%)	180 (11.06%)	254 (15.60%)
FRGC	169 (25.26%)	6 (0.90%)	111 (16.59%)

# of samples (% of impostor matches not sharing a classification)

compared to chance. By using the values displayed in the tables we tallied the total number of separate errors for each specific algorithm (e.g., for PPDM vs PDM on FOCS, PPDM has 3646 errors and PDM has 10914 errors) and performed a McNemar test [62] to calculate corresponding $p-$values. The resulting $p-$values indicate that improvement from PPDM on the FOCS and BDCP datasets is significant with $p < 0.0001$, while PPDM/SIFT-Flow improvements on FRGC are not significant with $p > 0.01$ (however, we already knew this given the small dataset size and similar verification rates). Fig. 12 displays some true positive (TP) and true negative (TN) examples where PPDM outperforms PDM by the widest margin, e.g., the difference between the PPDM TP scores and the VR threshold is combined with the difference between corresponding PDM match scores and sorted (to get the largest margin). From Fig. 12 it is clear that PPDM is improving on challenging matches within both FOCS and BDCP datasets.

When quantitatively comparing PPDM to SIFT-Flow within the FOCS and BDCP datasets there is a large number of shared errors (**blue** regions) that is not also reflected in the FRGC results (which suggests that

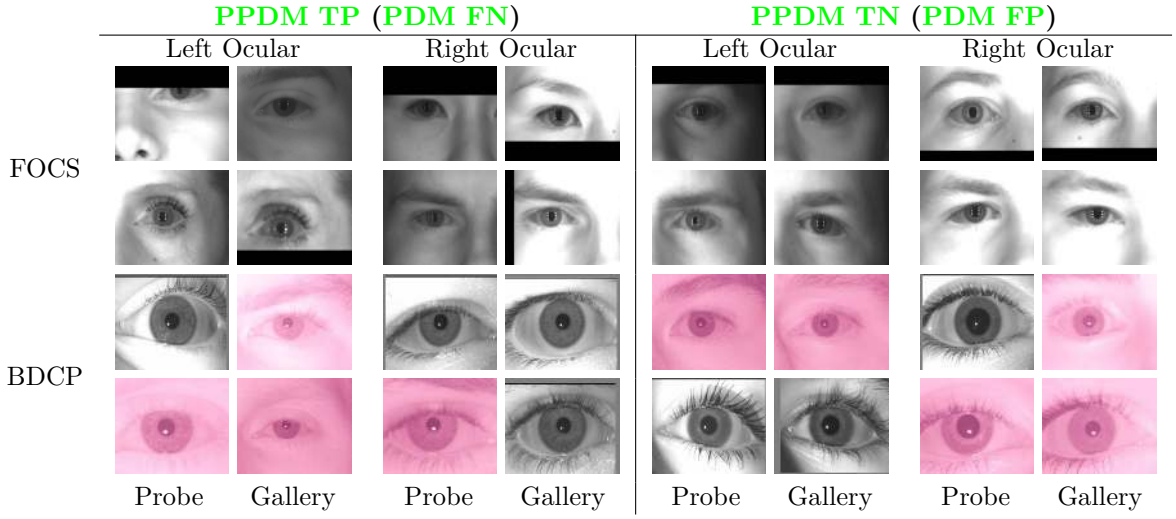


Fig. 12. True positive (TP) and true negative (TN) examples where PPDM outperforms PDM by the largest margin on the FOCS and BDCP datasets.

in this case the two algorithms may perform well if combined through score fusion). As a fine grained approach, SIFT-Flow has shown to perform well under low-noise/low-distortion conditions by reducing the number of false positives (FP), while PPDM being a coarse approximation has shown to be robust in harsh environments by reducing the number of false negatives (FN). However, as displayed in Table VII, SIFT-Flow is often computationally prohibitive, where PPDM (MATLAB implementation) is significantly more computationally efficient (~6× faster) than SIFT-Flow (MEX implementation). Other methods such as m-SIFT (MEX implementation [61]) may be noticeably faster, however sacrifice accuracy.

In addition to determining where PPDM succeeds over PDM, we also briefly examine when the proposed likelihood (SVL) or prior model is more responsible for PPDM's success. Similar to Tables III and IV, we focus on authentic and impostor match score errors, specifically targeting instances when PDM incorrectly classifies the probe image to the gallery but PDM-SVL and/or PPDM succeed (equivalent to the **green** regions in Fig. 11). Tables V and VI display the number of samples found in each resulting cluster and how the cluster size relates to the total number of matches outside overlapping classifications.[9] To put it into perspective, one can add columns 1 and 3 (*only* PPDM is correct + *only* PDM errors) to get the number of samples for PDM FN and PDM FP from Tables III and IV, respectively. For example, from the FOCS authentic match scores we see that, 2844 (*only* PPDM is correct) + 4908 (*only* PDM errors) = 7752 (PDM FN), thus, from the 7752 samples we can observe that 4908 occur from *both* PPDM and PDM-SVL, while the remaining 2844 come from *only* PPDM. This process can also be repeated for columns 2 and 3 (*only* PDM-SVL is correct + *only* PDM errors) to illustrate the breakdown of the PDM FN and PDM FP clusters when compared against just PDM-SVL.

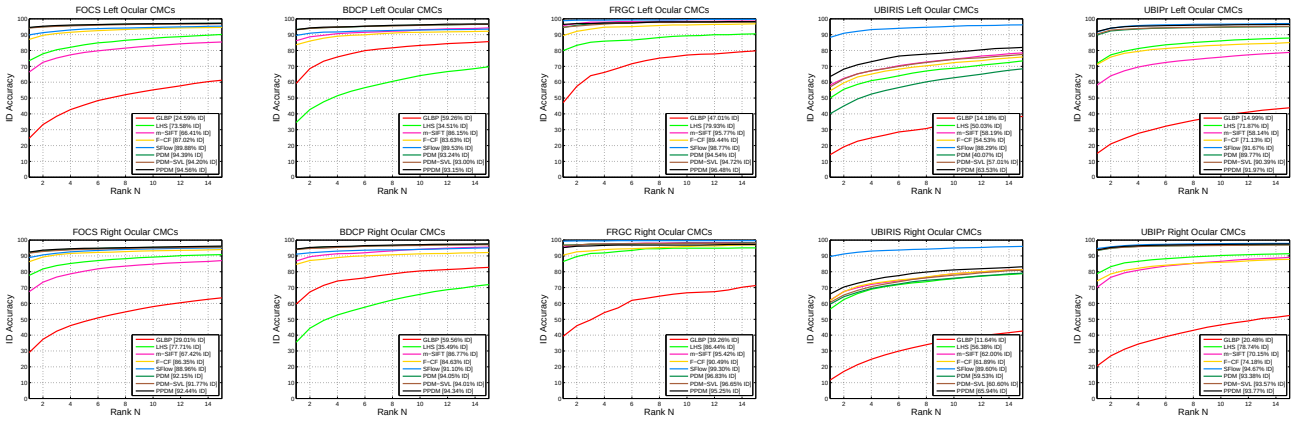
[9]Composed of the set of matches in which all models do not share the same result (equivalent to regions which are *not* colored **blue** or **black** in Fig. 11).

TABLE VII
COMPUTATION TIMES (IN MILLISECONDS) FROM THE BEST RESULTING IMAGE SIZE/MODEL ON EACH PERIOCULAR DATASET. EACH TIME PER COMPARISON IS RECORDED FROM MATCHING A RANDOM ASSORTMENT OF ~1000 AUTHENTIC AND ~9000 IMPOSTOR COMPARISONS (GENERATING A 100×100 SCORE MATRIX) AND AVERAGING THE RESULT. THESE TESTS WERE COMPLETED (SINGLE THREAD) ON A 64-BIT LAPTOP WITH A 2.67 GHz DUAL CORE INTEL I7-620M CPU AND 8 GB OF RAM.

	FOCS	BDCP	FRGC	UBIRIS	UBIPr
GLBP	0.443	0.067	0.067	0.021	0.025
LHS	4.131	3.351	4.060	3.340	104.471
m-SIFT	2.445	16.779	5.290	15.105	8.520
F-CF	22.391	52.843	19.257	50.993	176.419
SFlow	625.183	648.244	550.751	302.957	1404.772
PDM	93.331	94.441	98.310	85.830	84.435
PPDM	99.021	93.566	103.991	87.399	80.477

For all of the datasets we first observe that there are comparatively few match pairs where *only* PDM-SVL provides a correct classification (thus the proposed likelihood and prior models rarely conflict), suggesting that instances where both PPDM and PDM-SVL succeed (i.e., *only* PDM errors) can be attributed primarily to SVL. Hence, because the largest number of samples appear in such instances (*only* PDM errors) for authentic match scores, we can infer that most of PPDM's performance gains over PDM from reducing the number of FNs on FOCS and BDCP are due to SVL. We also observe that while there is little change in the total number of FPs between PPDM and PDM (seen in Table IV), the breakdown of the clusters shows that the proposed parameter estimation method is slightly more effective than SVL. However this result does not discount the PPDM prior model which provides a large increase to system performance over EM (and more than SVL on FRGC) while *greatly* reducing the computational expense by not requiring the use of inference.

Furthermore, while the focus of this paper has been



FOCS BDCP FRGC UBIRIS UBIPr

Fig. 13. Best baseline and PDM/PPDM CMC curves for left and right periocular regions on each dataset. Corresponding rank-1 identification (ID) rates are displayed in the legend for each method.

$1:1$ image matching, we also briefly investigate:

- $1:N$ **identification -** Fig. 13 displays cumulative match characteristic (CMC) curves for the best baselines and PDM/PPDM algorithms on each dataset. From the figures we observe the performance against the baselines is consistent with the results from $1:1$ recognition, achieving excellent accuracies on all datasets. However, unlike in verification scenarios, we see that PPDM is similar to PDM during identification scenarios (rather than greatly outperforming PDM). As previously discussed, PPDM improves on PDM by decreasing the number of false negatives rather than reducing the number of false positives. While this is beneficial in verification scenarios (less authentic pairs are classified as impostors), it has little benefit in increasing the rank-1 identification rate (which is better served by decreasing false positives, thereby having less impostors classified as authentic).
- **More than one gallery image per template -** Table VIII displays the resulting EERs, VRs, and rank-1 IDs when using an additional gallery image per subject (i.e., 2 authentic and no impostors) as well as the setup from [27] (3 authentic and 10 impostor images) on the FOCS dataset. The results show PDM outperforming PPDM when extra data is available for the gallery template (3 Auth, 10 Imp). As previously stated, the crux of the PDM algorithm is the strength of the deformation cues, i.e., the patch similarity strength and spatial shift. Thus, as the number of training images increases the number of variations in the correlation planes will decrease and the PPDM techniques will overfit to the training data.

Finally, based on the experimental results and above analysis we would like to pinpoint some strengths and limitations of PPDM. We observe that in general, PPDM is well suited for $1:1$ verification scenarios due to its ability to identify an authentic match pair (reducing the number of false negatives). However, while able to produce consistently good performance, it fails to achieve large gains in identification scenarios over PDM due to its limited ability to reduce the number of incorrectly classified impostor matches. The results from the FOCS dataset demonstrate that PPDM provides large improvements even in the presence of scale, rotation, and pose variations (even outperforming baselines that are designed to be scale and/or rotation tolerant such as LBP+HOG+SIFT, m-SIFT, and SIFT-Flow). Nevertheless, even though PPDM rank-1 ID rates on the UBIRIS data significantly outperform previous state-of-the-art in [12], the resulting EERs and VRs indicate that there exists areas for continued development when in the presence of *drastic* scale changes.

TABLE VIII
FOCS LEFT (L) AND RIGHT (R) PERIOCULAR RECOGNITION EERS, VRS AT 0.001 FAR, AND RANK-1 IDS USING ADDITIONAL TRAINING DATA.

		2 Auth, 0 Imp			3 Auth, 10 Imp		
		EER	VR	ID	EER	VR	ID
PDM	R	18.03%	47.10%	99.94%	15.47%	34.89%	99.22%
	L	20.47%	42.86%	99.22%	14.57%	46.97%	98.96%
PPDM	R	15.98%	52.52%	99.87%	16.94%	39.45%	99.24%
	L	17.59%	49.19%	99.38%	17.11%	38.70%	99.09%

X. CONCLUSIONS

Periocular recognition is an emerging biometric modality for human recognition under challenging scenarios. While there has been much work on periocular recognition, most assume that periocular patterns do not undergo local in-plane deformations. In this paper we show that periocular regions indeed undergo local in-plane deformations, especially under challenging imaging conditions, and accounting for these local distortions in the image matching process can result in significant improvements in the image matching performance leading to state-of-the-art periocular verification on multiple challenging periocular datasets.

We build upon the Probabilistic Deformation Models (PDM) approach originally proposed by Thornton et al. [27] to introduce the 'Periocular PDM' (PPDM) framework, achieving substantial performance and computational improvements. We derive the PPDM framework by designing the model to specifically consider local discriminative information under the intuition that

challenging environments increase intra-class variation. The proposed Spatially varying Likelihood (SVL) is a direct implementation of this concept by regulating the likelihood distribution to fit patches based on how they perform. The analysis presented in Section IX identifies SVL as responsible for the majority of PPDM's performance gains over PDM, illustrating the value of modeling local discriminative ability.

In a similar fashion, the proposed GMRF parameter estimation technique stems from the desire to tolerate less reliable deformation measurements. Learning is completed using the empirical covariance thereby removing the need to perform costly inference over the noisy deformation planes. The result being a simpler model for better generalization while also offering significant computational advantages.

We demonstrated the effectiveness of the PPDM approach as well as the limitations though rigorous experimentation and in-depth analysis. From the presented results we found that PPDM offers ~30% improvement over PDM and ~40% improvement over the best baseline in verification performance. In future work, we intend to explore the use of an alignment scheme (e.g., affine transformation) to address extreme scale and/or rotation variations, as well as the design of a discriminative model for deformation estimation to better separate authentic and impostor match pairs.

References

[1] A. Ross, R. Jillella, V. N. Boddeti, J. M. Smereka, B. V. K. Vijaya Kumar, P. Pauca, R. Barnard, and R. Plemmons, "Matching highly non-ideal ocular images: An information fusion approach," in *IAPR Int. Conf. on Biometrics*, 2012.

[2] D. Keysers, T. Deselaers, C. Gollan, and H. Ney, "Deformation models for image recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1422–1435, Aug. 2007.

[3] J.-W. Hsieh and W. Grimson, "Spatial template extraction for image retrieval by region matching," *IEEE Trans. on Image Processing*, vol. 12, no. 11, pp. 1404–1415, 2003.

[4] A. Rodriguez, V. N. Boddeti, B. V. K. Vijaya Kumar, and A. Mahalanobis, "Maximum margin correlation filter: A new approach for localization and classification," *IEEE Trans. on Image Processing*, vol. 22, no. 2, pp. 631–643, 2012.

[5] F. Juefei-Xu, K. Luu, M. Savvides, T. D. Bui, and C. Y. Suen, "Investigating age invariant face recognition based on periocular biometrics," in *Int. Joint Conf. on Biometrics*, Oct. 2011, pp. 1–7.

[6] C.-W. Tan and A. Kumar, "Human identification from at-a-distance images by simultaneously exploiting iris and periocular features," in *Int. Conf. on Pattern Recognition*, 2012, pp. 553–556.

[7] D. L. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross, "On the fusion of periocular and iris biometrics in non-ideal imagery," in *Int. Conf. on Pattern Recognition*, Aug. 2010, pp. 201–204.

[8] C.-W. Tan and A. Kumar, "Towards online iris and periocular recognition under relaxed imaging constraints," *IEEE Trans. on Image Processing*, vol. 22, pp. 3751–3765, 2013.

[9] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Recognizing surgically altered face images using multiobjective evolutionary algorithm," *IEEE Trans. on Information Forensics and Security*, vol. 8, no. 1, pp. 89–100, 2013.

[10] R. Jillela and A. Ross, "Mitigating effects of plastic surgery: Fusing face and ocular biometrics," in *IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, 2012, pp. 402–411.

[11] V. N. Boddeti, J. M. Smereka, and B. V. K. Vijaya Kumar, "A comparative evaluation of iris and ocular recognition methods on challenging ocular images," in *Int. Joint Conf. on Biometrics*, Oct. 2011.

[12] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," in *IEEE Int. Conf. on Biometrics: Theory Applications and Systems*, 2010, pp. 1–6.

[13] F. Juefei-Xu and M. Savvides, "Unconstrained periocular biometric acquisition and recognition using COTS PTZ camera for uncooperative and non-cooperative subjects," in *IEEE Workshop on Applications of Computer Vision*, 2012, pp. 201–208.

[14] P. E. Miller, J. R. Lyle, S. J. Pundlik, and D. L. Woodard, "Performance evaluation of local appearance based periocular recognition," in *IEEE Int. Conf. on Biometrics: Theory Applications and Systems*, 2010, pp. 1–6.

[15] G. W. Quinn and P. J. Grother, "Performance of face recognition algorithms on compressed images," National Institute of Standards and Technology (NIST), Tech. Rep. NISTIR 7830, Dec. 2011.

[16] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale results," National Institute of Standards and Technology (NIST), Tech. Rep., March 2007.

[17] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer, "An introduction to the good, the bad, & the ugly face recognition challenge problem," in *IEEE Int. Conf. on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 346–353.

[18] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[20] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[21] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Proc. of the 9th European Conf. on Computer Vision*. Springer-Verlag, 2006, pp. 428–441.

[22] B. Georgescu and P. Meer, "Point matching under large image deformations and illumination changes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 674–688, 2004.

[23] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[24] G. Santos and H. Proenca, "Periocular biometrics: An emerging technology for unconstrained scenarios," in *IEEE Workshop on Computational Intelligence in Biometrics and Identity Management*, April 2013, pp. 14–21.

[25] D. Keysers, J. Dahmen, T. Theiner, and H. Ney, "Experiments with an extended tangent distance," in *Int. Conf. on Pattern Recognition*, vol. 2, 2000, pp. 38–42.

[26] C. Liu, J. Yuen, and A. Torralba, "SIFT Flow: Dense correspondence across scenes and its applications," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 978–994, May 2011.

[27] J. Thornton, M. Savvides, and B. V. K. Vijaya Kumar, "A bayesian approach to deformed pattern matching of iris images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 596–606, April 2007.

[28] T. Deselaers and H. Ney, "Deformations, patches, and discriminative models for automatic annotation of medical radiographs," *Pattern Recognition Letters*, vol. 29, pp. 2003–2010, 2008.

[29] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC, 2005.

[30] NIST. Face and Ocular Challenge Series (FOCS). [Online]. Available: http://www.nist.gov/itl/iad/ig/focs.cfm

[31] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. Alexandre, "The UBIRISv2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, Aug 2010.
[32] S. Siena, V. N. Boddeti, and B. V. K. Vijaya Kumar, "Coupled marginal fisher analysis for low-resolution face recognition," in *Proc. of the 12th Int. Conf. on Computer Vision*, 2012, pp. 240–249.
[33] S. P. Fenker and K. W. Bowyer, "Experimental evidence of a template aging effect in iris biometrics," in *IEEE Workshop on Applications of Computer Vision*, 2011, pp. 232–239.
[34] C. Boehnen, D. Barstow, D. Patlolla, and C. Mann, "A multi-sample standoff multimodal biometric system," in *IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, Sept. 2012, pp. 127–134.
[35] C. N. Padole and H. Proenca, "Periocular recognition: Analysis of performance degradation factors," in *IAPR Int. Conf. on Biometrics*, 2012, pp. 439–445.
[36] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 947–954.
[37] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "The importance of small pupils: A study of how pupil dilation affects iris biometrics," in *IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–6.
[38] M. Fairhurst and M. Erbilek, "Analysis of physical ageing effects in iris biometrics," *IET Computer Vision*, vol. 5, no. 6, pp. 358–366, 2011.
[39] Y. Chen, S. Dass, A. Ross, and A. Jain, "Fingerprint deformation models using minutiae locations and orientations," in *Proc. of the 7th IEEE Workshops on Application of Computer Vision*, 2005, pp. 150–155.
[40] D. Kwon, I. D. Yun, and S.-U. Lee, "A robust warping method for fingerprint matching," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2007, pp. 1–6.
[41] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant representations of faces," *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 188–197, Jan. 2007.
[42] I. Kemelmacher-Shlizerman and R. Basri, "3D face reconstruction from a single image using a single reference face shape," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 394–405, 2011.
[43] A. Dantcheva, C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?" in *IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, 2012, pp. 391–398.
[44] I. A. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza, "Multimodal face recognition: Combination of geometry with physiological information," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 1022–1029.
[45] L. Cai and F. Da, "Nonrigid-deformation recovery for 3D face recognition using multiscale registration," *IEEE Computer Graphics and Applications*, vol. 32, no. 3, pp. 37–45, 2012.
[46] S. A. C. Schuckers, N. A. Schmid, A. Abhyankar, V. Dorairaj, C. K. Boyce, and L. A. Hornak, "On techniques for angle compensation in nonideal iris recognition," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 5, pp. 1176–1190, 2007.
[47] F. Juefei-Xu and M. Savvides, "Subspace-based discrete transform encoded local binary patterns representations for robust periocular matching on NIST's face recognition grand challenge," *IEEE Trans. on Image Processing*, vol. 23, no. 8, pp. 3490–3505, Aug. 2014.
[48] J. M. Smereka and B. V. K. Vijaya Kumar, "What is a 'good' periocular region for recognition?" in *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 117–124.
[49] B. V. K. Vijaya Kumar, A. Mahalanobis, and R. Juday, *Correlation Pattern Recognition*. New York: Cambridge University Press, 2005.
[50] V. N. Boddeti and B. V. K. Vijaya Kumar, "Extended depth of field iris recognition with correlation filters," in *IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–8.
[51] J. Thornton, M. Savvides, and B. V. K. Vijaya Kumar, "An evaluation of iris pattern representations," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept. 2007, pp. 1–6.
[52] J. Thornton, "Matching deformed and occluded iris patterns: A probabilistic model based on discriminative cues," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, USA, 2007.
[53] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical LASSO," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.
[54] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, June 2008.
[55] J. R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. J. LoIacono, S. Mangru, M. Tinker, T. M. Zappia, and W. Y. Zhao, "Iris on the move: Acquisition of images for iris recognition in less constrained environments," *Proc. of the IEEE*, vol. 94, no. 11, pp. 1936–1947, Nov 2006.
[56] NIST. Multiple Biometric Grand Challenge (MBGC). [Online]. Available: http://face.nist.gov/mbgc/
[57] U. Park, R. R. Jillela, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum," *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 1, pp. 96–106, 2011.
[58] D. L. Woodard, S. J. Pundlik, J. R. Lyle, and P. E. Miller, "Periocular region appearance cues for biometric identification," in *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, June 2010, pp. 162–169.
[59] H. Proenca, "Ocular biometrics by score-level fusion of disparate experts," *IEEE Trans. on Image Processing*, vol. 23, no. 12, pp. 5082–5093, Dec 2014.
[60] J. C. Moreno, V. B. S. Prasath, and H. Proença, "Robust periocular recognition by fusing local to holistic sparse representations," in *Proc. of the 6th Int. Conf. on Security of Information and Networks*, ser. SIN '13. New York, NY, USA: ACM, 2013, pp. 160–164.
[61] A. Vedaldi and B. Fulkerson. (2008) VLFeat: An open and portable library of computer vision algorithms. [Online]. Available: http://www.vlfeat.org/
[62] J. R. Beveridge, K. She, B. Draper, and G. H. Givens, "Parametric and nonparametric methods for the statistical evaluation of human ID algorithms," in *Proc. 3rd Workshop on the Empirical Evaluation of Computer Vision Systems*, 2001.

Jonathon M. Smereka received BS degrees in Computer Engineering, Electrical Engineering, and Engineering Mathematics from the University of Michigan at Dearborn in 2009. He is currently in the Electrical and Computer Engineering program at Carnegie Mellon University where he received the MS degree and is a candidate for the PhD degree. Before entering Carnegie Mellon he was employed as a Computer Engineer with the U.S. Army Tank-Automotive Research, Development and Engineering Center (TARDEC). His research interests are in Computer Vision, Pattern Recognition and Machine Learning. He was awarded the Carnegie Institute of Technology Dean's Tuition Fellowship in 2010 and received the Science, Mathematics, and Research for Transformation (SMART) Fellowship in 2011.

Vishnu Naresh Boddeti received a BTech degree in Electrical Engineering from the Indian Institute of Technology, Madras in 2007, and his MS and Ph.D. degree in Electrical and Computer Engineering program at Carnegie Mellon University. He is currently a member of the Research Staff in the CyLab Biometrics Center at Carnegie Mellon University. His research interests are in Computer Vision, Pattern Recognition and Machine Learning. He was awarded the Carnegie Institute of Technology Dean's Tuition Fellowship in 2007 and received the best paper award at the BTAS conference in 2013.

B.V.K. Vijaya Kumar received his B.Tech., and M.Tech., degrees in Electrical Engineering from Indian Institute of Technology, Kanpur and his Ph.D. in Electrical Engineering from Carnegie Mellon University, Pittsburgh. Since 1982, he has been a faculty member in the Department of Electrical and Computer Engineering (ECE) at Carnegie Mellon where he is now a U.A. & Helen Whitaker Professor and the Associate Dean for Graduate and Faculty Affairs for the College of Engineering. Professor Kumar's research interests include Computer Vision and Pattern Recognition Algorithms and Applications and Coding and Signal Processing for Data Storage Systems. His publications include a book entitled Correlation Pattern Recognition, twenty book chapters, 390 conference papers and 190 journal papers. He is also a co-inventor of 12 patents. He served as a Pattern Recognition Topical Editor for the Information Processing division of Applied Optics and as an Associate Editor for IEEE Transactions on Information Forensics and Security. Professor Kumar has served on many conference program committees and was co-chair of several conference program committees. Professor Kumar is a Fellow of IEEE, a Fellow of SPIE, a Fellow of OSA and a Fellow of the International Association of Pattern Recognition (IAPR). In 2003, Prof. Kumar received the Eta Kappa Nu award for Excellence in Teaching in the ECE Department at CMU and the Carnegie Institute of Technology's Dowd Fellowship for educational contributions and he was a co-recipient of the 2008 Outstanding Faculty Research Award in CMU's College of Engineering.