

# Gaussian Conditional Random Fields for Face Recognition

Jonathon M. Smereka and B.V.K. Vijaya Kumar  
Carnegie Mellon University  
5000 Forbes Ave, Pittsburgh, PA 15213

Andres Rodriguez  
Intel Corporation  
Hillsboro, OR 97124

## Abstract

We propose a Gaussian Conditional Random Field (GCRF) approach to modeling the non-stationary distortions that are introduced from changing facial expressions during acquisition. While previous work employed a Gaussian Markov Random Field (GMRF) to perform deformation tolerant matching of periocular images, we show that the approach is not well-suited for facial images, which can contain significantly larger and more complex deformations across the image. Like the GMRF, the GCRF tries to find the maximum scoring assignment between a match pair in the presence of non-stationary deformations. However, unlike the GMRF, the GCRF directly computes the posterior probability that the observed deformation is consistent with the distortions exhibited in other authentic match pairs. The difference is the inclusion of a derived mapping between an input comparison and output deformation score. We evaluate performance on the CMU Multi-PIE facial dataset across all sessions and expressions, finding that the GCRF is significantly more effective at capturing naturally occurring large deformations than the previous GMRF approach.

## 1. Introduction

Using facial features as a biometric modality is one of the most common forms of user identification. Considerable efforts have been made toward the advancement of face biometrics. This is partly due to the remarkable ability of humans to recognize faces [28], as well as the ease of acquisition (non-invasive and less controlled than iris or fingerprint) combined with an increasing availability of cameras/video for acquisition [26, 39, 40]. However, even in controlled acquisition, matching facial images with distortions is a challenging problem [11, 21, 18, 19]. Facial expressions introduce non-stationary image movements in both facial [4, 15] and ocular images [1] that notably affect matching performance.

Probabilistic Deformation Models (PDMs) are a technique proposed by Thornton et al. [34] to simultaneously estimate the non-linear deformation in iris images caused by the expansion and contraction of the pupil and match iris images using a Gaussian Markov Random Field (GMRF).

In controlled acquisition environments the PDM method works well and all images patches are treated equally. That is, the model assumptions make no explicit distinction in varying discrimination ability across the biometric image. As the method was originally intended for iris recognition, where the rich texture of the iris is segmented from the eye, there was no specific need to account for varying discrimination outside of removing artifacts from improper segmentation. In [29], the approach was adapted for periocular recognition, referred to as a Periocular PDM (PPDM), by extending the model to leverage the varying discrimination ability across the pair of images being matched. PPDM demonstrated that the original PDM method could be used across different biometric domains and demonstrated the benefits of deformation modeling of the ocular region in environments where iris recognition can fail.

In this work we further extend PDM to facial biometrics, which can contain significantly larger and more complex distortions (e.g., due to facial expressions) across the image. To accomplish this we derive a Gaussian Conditional Random Field (GCRF) to model these deformations. Like the PPDM GMRF, the GCRF finds the maximum scoring assignment between the probe and gallery in the presence of non-stationary deformations. Unlike the GMRF, the GCRF builds a log-linear model to directly obtain the posterior. Similar to [29] we develop a method for efficient training which *does not require Bayesian inference*.

Previous works [20, 7] have employed CRFs for face recognition, however to the authors' best knowledge, no one has studied the use of a GCRF for face recognition. What makes a Gaussian graphical model special is its ability to compactly represent the structure and parameterization of the model within the precision matrix [25]. Accordingly, most work [36, 35, 38, 37, 23, 30, 5] concerning GCRFs (or related models, e.g., Partial Gaussian Graphical Models [38]) comes from structure learning due to the compact representation of the graph within the precision matrix. A few [33, 22, 6] have investigated the use of a GCRF for prediction using a pre-defined graph structure. The first to do so was Tappen et al. [33], where the authors define a GCRF as a weighted set of linear convolution kernels to perform im-

age denoising. The model is then trained using a gradient-based approach to minimize the difference (or cost) between the predicted result and ground truth for a given set of parameters (the authors remark on the reduced computation compared to maximizing the likelihood).

Djuric et al. [6] and Radosavljevic et al. [22] however train their GCRF predictive models by finding parameters that maximize the conditional log-likelihood via gradient descent. When performing structure prediction, Wytock and Kolter [36] observe that gradient-descent methods for maximum likelihood estimation (MLE) are often slow to converge and may perform poorly due to ill-conditioned matrices. They instead use alternating direction method of multipliers (ADMM) and later in [35] derive a Newton coordinate descent approach. Similar to [35], we employ a Newton coordinate descent optimization to learn GCRF parameters, however we also take a note from Chiquet et al. [5] by including additional information to account for the structure among the predictors.

We demonstrate the efficacy of this approach on the CMU Multi-PIE (Pose, Illumination, and Expression) [12] facial dataset with verification rate (VR), equal error rate (EER), and rank-1 identification rate (ID) as measures of system performance. The use of the Multi-PIE face dataset is ideal for measuring performance in the presence of large distortions (e.g., users with a scream expression are matched against the same users with a neutral expression), and thus is an appropriate benchmark for measuring the model’s ability to capture naturally occurring extreme image deformations.

## 2. PPDM Gaussian Markov Random Field

The PPDM technique first divides a given probe image,  $\mathbf{I}$ , and gallery image,  $\mathbf{G}$ , into  $N$  rectangular patches. The corresponding probe and gallery patches are then matched via template matching to obtain a patch similarity,  $S(\cdot)$ , at the location of best match (local  $x - y$  spatial translation) which serve as inputs to the graphical model. For the  $i$ -th patch pair the relative spatial translation is denoted by  $\{(\Delta x_i, \Delta y_i)\}$ , and is used to define the deformation for the image pair as a vector  $\mathbf{d} = [\Delta x_1, \Delta y_1, \dots, \Delta x_N, \Delta y_N]^T$ . For an authentic match pair the underlying latent deformation,  $\hat{\mathbf{d}}$ , is computed using maximum-a-posteriori (MAP) estimation to determine the most likely deformation vector between the two images:

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

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 a global relative shift between the probe and gallery) and covariance matrix  $\Sigma_{\mathbf{d}} \in \mathbb{R}^{2N \times 2N}$ . For GMRFs [25] the graph connections are encoded by the structure of the precision matrix  $\Sigma_{\mathbf{d}}^{-1}$ , where the GMRF properties ensure that  $\Sigma_{\mathbf{d}_{i,j}}^{-1} = 0$  for non-neighboring  $\mathbf{d}_i$  and  $\mathbf{d}_j$ . More details on PPDM can be found within [29].

## 3. PPDM Gaussian Conditional Random Field

We now derive a Gaussian Conditional Random Field (GCRF) for the deformation model. The PPDM GMRF formulation operates under the assumption that authentic deformations follow a multivariate Gaussian distribution, and given favorable deformation estimates, can appropriately generate a model for valid patch translations. However, as the deformation estimates degrade due to noise, occlusion, etc., the ability of the GMRF to properly classify authentic deformation also degrades. Thus, to adapt to more challenging match scenarios, we want to learn a predictive model that determines the probability of authentic deformation given an input image comparison, without having to directly model the inputs themselves.

Since their introduction by Lafferty et al. [16] the CRFs have become a popular choice for modeling over a MRF as they do not require the representation of the dependencies between the input variables. This stems from the CRF modeling the conditional probability distribution, while a MRF models the joint probability distribution. Like the previously described PPDM GMRF, the GCRF tries to find the maximum scoring assignment given that the probe image matches the gallery template in the presence of non-stationary deformations (represented as local patch translations). However, unlike the GMRF, the GCRF builds a log-linear model to directly obtain the posterior,  $P(\mathbf{d}|\mathbf{G}, \mathbf{I}) = P(\delta|\Upsilon)$  (we use the variable  $\Upsilon$  to refer to the output from the matching between the probe,  $\mathbf{I}$ , and gallery,  $\mathbf{G}$ , images, and  $\delta$  as the deformation score). The difference is that in the GMRF we’re learning the statistics of the distribution to properly represent authentic deformation (approximated by spatial translations, i.e.,  $\Delta x_i$  and  $\Delta y_i$ ), while in the GCRF, we’re learning how to separate authentic and impostor deformations. In both cases we’re computing the latent deformation (how authentic/consistent the observed  $x$  and  $y$  values for each node are with previously seen deformations), but in the GCRF this value is a single score at each node, represented as  $\delta$ , while in the GMRF it’s a score per  $x$  and  $y$  at each node, represented as  $\mathbf{d}$ .

We assume a multivariate Gaussian distribution between our inputs,  $\Upsilon \in \mathbb{R}^{2N}$  (i.e.,  $\Delta x_i$  and  $\Delta y_i \forall i$ ), and outputs,  $\delta \in \mathbb{R}^N$  (the deformation score value for each node), with mean  $\mu = [\mu_{\Upsilon}^T, \mu_{\delta}^T]$ , covariance  $\Sigma$ , and precision matrix  $\Omega$ :

$$\Omega = \begin{bmatrix} \Sigma_{\Upsilon} & \Sigma_{\Upsilon\delta} \\ \Sigma_{\Upsilon\delta}^T & \Sigma_{\delta} \end{bmatrix}^{-1} = \begin{bmatrix} \Omega_{\Upsilon} & \Omega_{\Upsilon\delta} \\ \Omega_{\Upsilon\delta}^T & \Omega_{\delta} \end{bmatrix} \quad (2)$$

where  $\Omega_\delta \in \mathbb{R}^{N \times N}$  contains a structure similar to  $\Sigma_d^{-1}$  in the GMRF, directly encoding the conditional dependency relationships of  $\Upsilon$  and  $\delta$  as proven in [17] and [25]. The distribution is as follows:

$$P(\delta|\Upsilon) = \frac{1}{Z_{\delta\Upsilon}(\Upsilon)} \exp\left(-\frac{1}{2}\delta^T \Omega_\delta \delta - \Upsilon^T \Omega_{\Upsilon\delta} \delta\right) \quad (3)$$

with  $\Omega_\delta = (\Sigma_\delta - \Sigma_{\Upsilon\delta}^T \Sigma_\Upsilon^{-1} \Sigma_{\Upsilon\delta})^{-1}$  directly parameterizing the conditional independence relationships of the graph,<sup>1</sup> while the term  $\Omega_{\Upsilon\delta} = -\Sigma_\Upsilon^{-1} \Sigma_{\Upsilon\delta} \Omega_\delta \in \mathbb{R}^{2N \times N}$  maps the inputs (image comparison) to the outputs (probability of authentic deformation). Finally, the partition function  $Z_{\delta\Upsilon}(\Upsilon)$  ensures that the distribution over the latent deformations,  $\delta$ , integrates to one, making the result a proper probability distribution:

$$\begin{aligned} Z_{\delta\Upsilon}(\Upsilon) &= \int \exp\left(-\frac{1}{2}\delta^T \Omega_\delta \delta - \Upsilon^T \Omega_{\Upsilon\delta} \delta\right) \partial\delta \\ &= \frac{(2\pi)^{\frac{\lambda N}{2}}}{|\Omega_\delta|^{1/2}} \exp\left(\frac{1}{2}\Upsilon^T \Omega_{\Upsilon\delta} \Omega_\delta^{-1} \Omega_{\Upsilon\delta}^T \Upsilon\right) \quad (4) \end{aligned}$$

Sohn and Kim [30] as well as Yuan and Zhang [38] demonstrated that the GCRF can be equivalently expressed as a multivariate regression model in which:

$$\delta_i = \Upsilon_i \mathbf{B} + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \quad \text{for } i = 1, \dots, N \quad (5)$$

where  $\mathbf{B} = \Sigma_\Upsilon^{-1} \Sigma_{\Upsilon\delta} = -\Omega_{\Upsilon\delta} \Omega_\delta^{-1}$  is the matrix of regression coefficients and  $\epsilon_i$  is the vector for Gaussian distributed noise with covariance  $\mathbf{R} = \Omega_\delta^{-1}$ . This result demonstrates that the GCRF is effectively learning a direct mapping (shown here as  $\mathbf{B}$ ) between the input image comparisons,  $\Upsilon$ , and the output deformation scores,  $\delta$ . However we cannot estimate the regression coefficients,  $\mathbf{B}$ , using an ordinary least squares (OLS) solution due to the precondition that each input variable is independent (i.e.,  $\mathbf{B}$  and  $\epsilon$  are uncorrelated, formal proof in [14]).

### 3.1. Training

Similar to the PDDM GMRF in [29], we derive a method of parameter estimation which does not require Bayesian inference. Due to a static graph structure, the GMRF could be trained efficiently by manipulating a Graphical LASSO formulation [10] to impose the specific graph structure as the sparsity constraints when solving the maximum likelihood estimation (MLE) problem. Unfortunately, the GCRF is not only learning the parameter values needed to build the sparse precision matrix,  $\Omega_\delta$ , but also requires values for the dense matrix,  $\Omega_{\Upsilon\delta}$ , representing the correspondence between an input comparison and output deformation score.

Using Equation 3, the optimization problem for finding  $\Omega_\delta$  and  $\Omega_{\Upsilon\delta}$  that maximizes the likelihood of the deformations that occur in authentic match pairs is as follows (via minimizing the negative log-likelihood,  $\mathcal{L}(\Omega_\delta, \Omega_{\Upsilon\delta})$ ):

<sup>1</sup>We can easily return to the GMRF formulation if  $\Sigma_{\Upsilon\delta} \rightarrow 0$ .

$$\begin{aligned} \min_{\Omega_\delta, \Omega_{\Upsilon\delta}} & -\frac{1}{2} \ln(|\Omega_\delta|) + \frac{1}{2} \text{tr}(\Omega_\delta \hat{\mathbf{S}}_\delta) + \text{tr}(\Omega_{\Upsilon\delta} \hat{\mathbf{S}}_{\delta\Upsilon}) \\ & + \frac{1}{2} \text{tr}(\Omega_{\Upsilon\delta} \Omega_\delta^{-1} \Omega_{\Upsilon\delta}^T \hat{\mathbf{S}}_\Upsilon) \quad (6) \end{aligned}$$

where  $\hat{\mathbf{S}} = \begin{bmatrix} \hat{\mathbf{S}}_\Upsilon & \hat{\mathbf{S}}_{\Upsilon\delta} \\ \hat{\mathbf{S}}_{\delta\Upsilon} & \hat{\mathbf{S}}_\delta \end{bmatrix}$  is the empirical covariance matrix. While the problem is convex (formal proof in [38] - Proposition 1), it was shown by Wytock and Kolter [36] that the partition function term,  $\Omega_{\Upsilon\delta} \Omega_\delta^{-1} \Omega_{\Upsilon\delta}^T \hat{\mathbf{S}}_\Upsilon$ , poses significant challenges to optimization methods like gradient-descent (where the resulting  $\Omega_\delta$  may not be full rank). Thus, instead of a gradient-descent approach we employ a Newton coordinate descent optimization similar to [35] to learn GCRF parameters, while taking into account any strong correlations that may exist within the data (training deformation vectors).

Chiquet et al. [5] show that a GCRF LASSO estimation suffers from an inability to distinguish the non-zero coefficients of  $\Omega_{\Upsilon\delta}$  when strong correlations exist between the responses described by  $\Omega_\delta$ . Their regularization scheme instead uses a generalized Elastic-Net [41] formulation of Equation 6 to account for structure among the predictors during optimization. That is, while the  $\ell_1$  penalty imposed by LASSO encourages sparsity, this limit on included variables often results in a failure to include groups by only selecting one variable from the group and ignoring others. To remove the limitation of the number of selected variables and encourage groups, the Elastic-Net procedure adds an additional ‘smoothing’ quadratic penalty to the minimization problem. Consequently, when using the multivariate regression relationship in Equation 5, the resulting ‘smoothing’ operation stems from the quadratic  $\mathbf{B}^T \mathbf{L} \mathbf{B}$  in the same fashion as the partition function term in Equation 6, where  $\mathbf{L} \in \mathbb{R}^{2N \times 2N}$  is the Laplacian matrix (defined as the adjacency matrix<sup>2</sup> subtracted from the degree matrix<sup>3</sup>) encoding the structure of the predictors (x and y).

Since including component structural information is best applied when large pairwise correlations are present in the training data, we weight the Laplacian such that  $\hat{\mathbf{L}} = \zeta \mathbf{L}$  where  $\zeta$  is determined via cross-validation during training. Thus, by optimizing over  $\mathcal{L}(\Omega_\delta, \Omega_{\Upsilon\delta})$  from Equation 6 we can formulate the problem as:

$$\begin{aligned} \min_{\Omega_\delta, \Omega_{\Upsilon\delta}} & -\frac{1}{2} \ln(|\Omega_\delta|) + \frac{1}{2} \text{tr}(\Omega_\delta \hat{\mathbf{S}}_\delta) + \text{tr}(\Omega_{\Upsilon\delta} \hat{\mathbf{S}}_{\delta\Upsilon}) \\ & + \frac{1}{2} \text{tr}(\Omega_{\Upsilon\delta}^T (\hat{\mathbf{S}}_\Upsilon + \hat{\mathbf{L}}) \Omega_{\Upsilon\delta} \Omega_\delta^{-1}) \quad (7) \end{aligned}$$

which is jointly convex in  $(\Omega_{\Upsilon\delta}, \Omega_\delta)$  when the number of training samples  $\geq N$  and  $(\hat{\mathbf{S}}_\Upsilon + \hat{\mathbf{L}}) \succ 0$  (formal proof in [38] and [5]). Using Equation 7 we can use a Newton

<sup>2</sup>The adjacency matrix a symmetric matrix where the non-diagonal entries are zero if no edge exists between the corresponding nodes.

<sup>3</sup>The degree matrix is a diagonal matrix with the number of neighbors for each node along the diagonal.

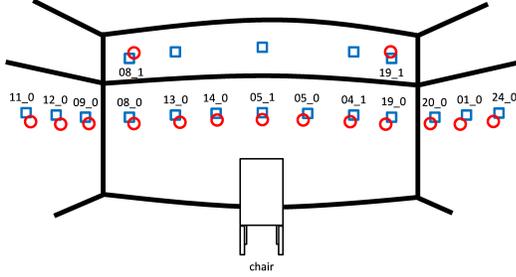


Figure 1. Illustration of the camera setup to acquire images of varying pose, expression, and lighting. Cameras (red circles), flashes (blue boxes) and their labels are shown surrounding the subject at  $15^\circ$  intervals. Illustration is based on that in [8].

coordinate descent approach similar to [35] (who in turn mirrors [13]) for determining  $\Omega_\delta$  and  $\Omega_{\Upsilon\delta}$ .

### 3.2. Inference

Just as in the case of the PPDM GMRF in [29], given the Bayesian model and observations (i.e., the outputs from template matching) our objective is to estimate the posterior probability distribution. For the GMRF, Iterated Conditional Modes (ICM) was used due to its considerable speed advantages despite its sensitivity to the initial estimate [31]. However, while the GCRF intuitively generalizes the GMRF (such that inference ‘hardness’<sup>4</sup> is not necessarily different for one or the other), we found that ICM was not as effective in determining proper posteriors. Thus we instead employ Gaussian Belief Propagation (GaBP) [3, 27], which uses the special properties of Gaussian distributions (e.g., the product of two Gaussian distributions is, up to a constant factor, also Gaussian) to simplify the traditional Belief Propagation (BP) algorithm. From each output from template matching,  $C_k$ , we extract the peak value,  $\hat{S}_k$ , and location,  $\Upsilon_k = [x_k, y_k]^T$ , and run GaBP inference until convergence or the maximum number of iterations is reached. The final match score,  $M$ , is then computed as the normalized inner product between the resulting marginal posterior distributions (or ‘beliefs’,  $\varphi$ ) from inference and output peak score values from template matching:

$$M = \frac{1}{N} \langle C, \varphi \rangle \quad (8)$$

## 4. Experiments

While PPDM in [29] was developed for deformation estimation over the ocular region, using a GCRF we extend the model to facial images. To evaluate the performance of the MRF and CRF architectures against large in-plane deformations (i.e., from facial expressions), we include experiments on the CMU Multi-PIE (Pose, Illumination, and Expression) [12] face image dataset.

The CMU Multi-PIE dataset contains full face images (640×480 pixels) obtained from 15 cameras surrounding

<sup>4</sup>It was shown in [24] that approximate inference in MRFs is NP-hard.

| Subset      | # Images | # Subjects | # Samples per Subject |
|-------------|----------|------------|-----------------------|
| Session 1   | 1494     | 249        | 6                     |
| Session 2   | 1827     | 203        | 9                     |
| Session 3   | 2070     | 230        | 9                     |
| Session 4   | 2151     | 239        | 9                     |
| Single Pose | 2514     | 337        | 2 - 11                |
| Total       | 7542     | 337        | 6 - 33                |

Table 1. CMU Multi-PIE face dataset information.

each of the 337 separate subjects. The design of the set is such that multiple images are captured from users over the aforementioned challenges. Of the 15 cameras (pose variation), 13 were spaced over  $15^\circ$  intervals at head height and the remaining 2 were located above the subject for a surveillance view (see Fig. 1). Over 6 months, images of the subjects were acquired in 4 separate sessions. During each of the sessions, subjects held a neutral expression as well as one or two additional expressions (for a total of 6 expressions across the dataset). Since we are specifically interested in performance over large in-plane deformations (i.e., expression), we only consider images from the first acquisition (no flash) over poses:  $-15^\circ$ ,  $0^\circ$ ,  $+15^\circ$ . In the experiments we evaluate the face images at each pose, collecting images over all expressions and sessions. A total of 7542 images are collected (2514 total images per pose, see Table 1 for statistics per session) with at least 264 users attending 2 or more sessions. Since the images contain a considerable amount of information outside of the face, we use annotations provided by [9] to crop each face as shown in Fig. 2. We then pre-process each image by the histogram normalization developed by Tan and Triggs [32] along with resizing to  $128 \times 128$  pixels for computational efficiency.

Each method is evaluated in a 1 : 1 image-to-image matching scenario using 5-fold cross validation. The verification rates (VRs), computed as  $1 - \text{the False Reject Rate (FRR)}$  at 0.001 False Acceptance Rate (FAR), equal error rates (EERs), and rank-1 identification rates (IDs) are computed from the concatenated scores from the associated folds, excluding self-comparisons. Numerical results are provided in Table 2 with Fig. 3 showing ROCs and CMCs.

## 5. Results

From the results in Table 2, we see that for each pose there is a distinct decrease in EER along with an increase in VR and rank-1 IDs. Table 3 shows where the GCRF is better than the GMRF. Specifically we compare authentic and impostor match score errors, i.e., false negatives (FN) and false positives (FP). Using the match score threshold which determines the EER, the Tables 3a and 3b first display the number of match pairs where *only* the GCRF produces an error compared with the number of match pairs where *only* the GMRF produces an error. For example; within Table

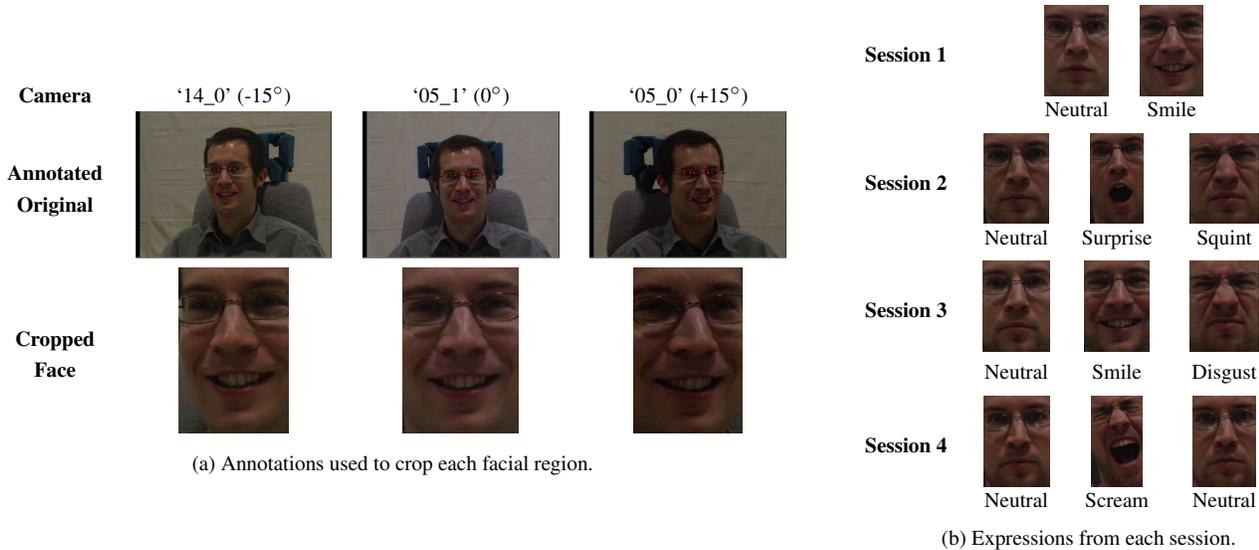


Figure 2. Sample images from the CMU Multi-PIE dataset. (a) We crop the facial regions using the annotations provided by [9], where the cropped image width and height values are determined from the available label locations with an additional 25 pixels extracted from both sides of the facial image and an additional 40 pixels extracted from the top of the facial image. (b) Examples of the cropped expressions from each session.

|      | GMRF   |        |        | GCRF          |               |               |
|------|--------|--------|--------|---------------|---------------|---------------|
|      | VR     | EER    | ID     | VR            | EER           | ID            |
| -15° | 51.15% | 19.84% | 73.64% | <b>56.44%</b> | <b>14.64%</b> | <b>75.86%</b> |
| 0°   | 49.09% | 21.33% | 72.28% | <b>56.49%</b> | <b>15.96%</b> | <b>74.80%</b> |
| +15° | 47.72% | 20.85% | 70.46% | <b>55.51%</b> | <b>16.00%</b> | <b>74.80%</b> |

Table 2. System performance using a GMRF and GCRF deformation model on each evaluated pose of the CMU Multi-PIE dataset.

3a, from the -15° pose, there are 148 match pairs where the GCRF produces a FN, but with the GMRF they are correctly classified as true positives (TPs), and 362 match pairs which are FNs when using the GMRF, but correctly are classified as TPs when using the GCRF. Then in the neighboring table we include the compared expressions (over all sessions) that primarily compose each score cluster. Since the results from each pose were largely similar we averaged the size of the set between the three poses for a more concise synopsis.

By using the values shown in the tables we tallied the total number of separate errors for each pose (e.g., for -15° the GCRF has 90996 errors and the GMRF has 155926 errors) and computed the statistical significance of the GCRF’s improved performance when compared to chance. Recall that while a  $p$ -value is not to be treated as a probability per-se, the value does indicate a level of confidence in the result. If the null hypothesis (no difference between the two algorithms) were true and the  $p$ -value indicates otherwise, then the data used to produce the  $p$ -value would be *highly irregular*. That is, the algorithm specific error clusters would need to be extremely abnormal (such as an error with the algorithm score assignment) to provide a  $p$ -value that demonstrates a likely difference between the two al-

gorithm’s performance, if in actuality there is none. The resulting  $p$ -values from a McNemar test [2] indicated that improvement from the GCRF on each pose is significant with  $p < 0.0001$ , demonstrating that the GCRF deformation model is effective at reducing the EER by making fewer errors on both authentic *and* impostor match comparisons.

Specifically, in authentic comparisons the GCRF is improving performance through properly recognizing a subject in the presence of large distortions (i.e., scream vs neutral and scream vs smile), while instances where the GMRF outperforms the GCRF are predominantly from comparisons with less deformation between the match pair (e.g., disgust vs neutral and smile vs neutral). Fig. 4 shows qualitatively what Table 3 shows quantitatively, displaying specific examples when the GCRF is outperforming the GMRF by the widest margin. From Fig. 4, it is clear that the GCRF is improving on challenging matches.

## 6. Conclusion

We derived a Gaussian Conditional Random Field (GCRF) for modeling the deformation that an authentic biometric sample may undergo between acquisitions. As a discriminative method, the GCRF trains a direct mapping between an input comparison and output score, and thus removes the need to model the dependencies between the output peak values from template matching. After evaluating performance on the CMU Multi-PIE facial dataset, we showed that the GCRF is significantly more effective than the previous Gaussian Markov Random Field (GMRF) approach in its ability to capture naturally occurring large in-plane deformations. In future work, we intend to explore the use of GCRFs with a non-static graph structure.

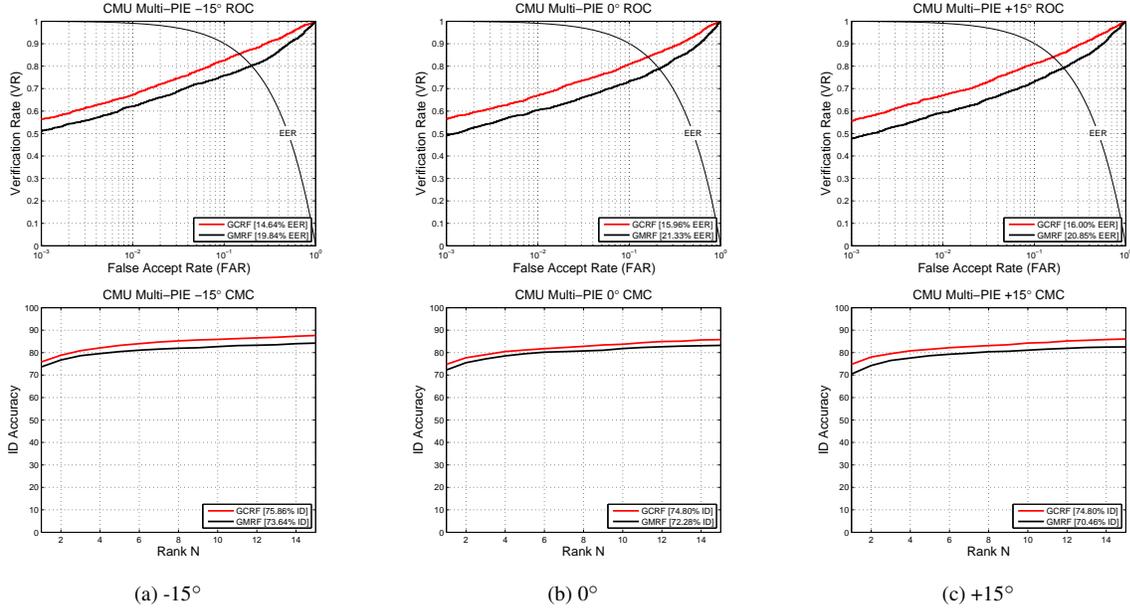


Figure 3. GMRF vs GCRF ROC (top) and CMC (bottom) curves for each evaluated pose of the CMU Multi-PIE dataset.

| GCRF FN vs GMRF FN |              |        | Score Cluster        | Compared Expressions | % of Set |
|--------------------|--------------|--------|----------------------|----------------------|----------|
|                    | # of Samples | Change |                      |                      |          |
| -15°               | 148 vs 362   | 2.45×  | GCRF TP<br>(GMRF FN) | Scream vs Neutral    | ~48%     |
|                    | 166 vs 386   | 2.33×  |                      | Surprise vs Neutral  | ~11%     |
|                    | 178 vs 372   | 2.09×  |                      | Scream vs Smile      | ~11%     |
| 0°                 | 148 vs 362   | 2.45×  | GMRF TP<br>(GCRF FN) | Disgust vs Neutral   | ~34%     |
|                    | 166 vs 386   | 2.33×  |                      | Smile vs Neutral     | ~20%     |
|                    | 178 vs 372   | 2.09×  |                      | Disgust vs Smile     | ~4%      |

(a) False negatives (FN) from authentic match score clusters.

| GCRF FP vs GMRF FP |                 |        | Score Cluster        | Compared Expressions | % of Set |
|--------------------|-----------------|--------|----------------------|----------------------|----------|
|                    | # of Samples    | Change |                      |                      |          |
| -15°               | 90848 vs 155564 | 1.71×  | GCRF TN<br>(GMRF FP) | Neutral vs Neutral   | ~41%     |
|                    | 98072 vs 165738 | 1.69×  |                      | Smile vs Neutral     | ~18%     |
|                    | 99288 vs 160340 | 1.61×  |                      | Smile vs Smile       | ~6%      |
| 0°                 | 90848 vs 155564 | 1.71×  | GMRF TN<br>(GCRF FP) | Neutral vs Neutral   | ~17%     |
|                    | 98072 vs 165738 | 1.69×  |                      | Smile vs Neutral     | ~14%     |
|                    | 99288 vs 160340 | 1.61×  |                      | Scream vs Neutral    | ~12%     |

(b) False positives (FP) from impostor match score clusters.

Table 3. Comparing GCRF and GMRF false negatives (FN) and false positives (FP) from authentic and impostor score clusters, respectively, along with the expressions of the images primarily composing each set.

## 7. Acknowledgments

We would like to thank the Air Force Research Laboratory Sensors Directorate and the Air Force Office of Scientific Research (AFOSR) for providing the main funding for this work through LRIR 14Y06COR. This research is also supported in part by an Amazon Web Services (AWS) Education Research Grant award.

## References

[1] E. Barroso, G. Santos, L. Cardoso, C. Padole, and H. Proença. Periocular recognition: How much facial ex-

pressions affect performance. *Pattern Analysis and Applications*, 18:1–14, June 2015.

[2] 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.

[3] D. Bickson. *Gaussian Belief Propagation: Theory and Application*. PhD thesis, The Hebrew University of Jerusalem, Jerusalem, Israel, 2009.

[4] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-invariant representations of faces. *IEEE Trans. on Image Processing*, 16(1):188–197, Jan. 2007.

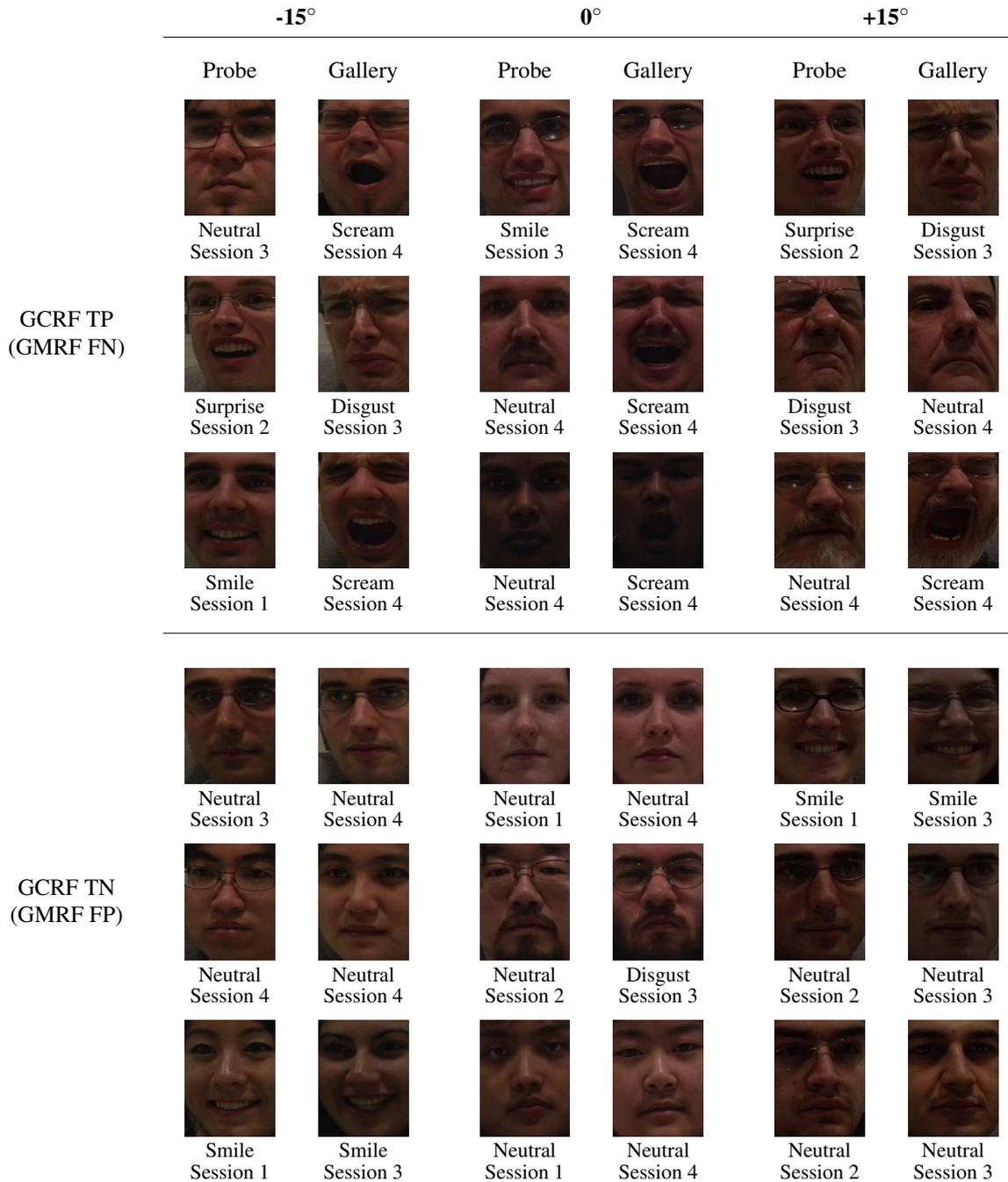


Figure 4. True positive (TP) and true negative (TN) examples where the GCRF outperforms the GMRF by the largest margin.

- [5] J. Chiquet, T. Mary-Huard, and S. Robin. Structured regularization for conditional gaussian graphical models. *CoRR*, abs/1403.6168, 2014.
- [6] N. Djuric, V. Radosavljevic, Z. Obradovic, and S. Vucetic. Gaussian conditional random fields for aggregation of operational aerosol retrievals. *IEEE Geoscience and Remote Sensing Letters*, 12(4):761–765, April 2015.
- [7] M. Du and R. Chellappa. Face association for videos using conditional random fields and max-margin markov networks. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PP(99), 2015.
- [8] A. Dutta, R. Veldhuis, and L. Spreuwens. A bayesian model for predicting face recognition performance using image quality. In *IEEE Int. Joint Conf. on Biometrics*, pages 1–8, 2014.
- [9] L. El Shafey, C. McCool, R. Wallace, and S. Marcel. A scalable formulation of probabilistic linear discriminant analysis: Applied to face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(7):1788–1794, July 2013.
- [10] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics*,

- 9(3):432–441, 2008.
- [11] T. Gass, L. Pishchulin, P. Dreuw, and H. Ney. Warp that smile on your face: Optimal and smooth deformations for face recognition. In *IEEE Int. Conf. on Automatic Face Gesture Recognition and Workshops*, pages 456–463, March 2011.
- [12] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *IEEE Int. Conf. on Automatic Face Gesture Recognition*, pages 1–8, 2008.
- [13] C.-J. Hsieh, M. A. Sustik, I. S. Dhillon, and P. Ravikumar. Sparse inverse covariance matrix estimation using quadratic approximation. In *Neural Information Processing Systems*, 2011.
- [14] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., 6th edition, 2007.
- [15] 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*, 33(2):394–405, 2011.
- [16] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Int. Conf. on Machine Learning*, pages 282–289, 2001.
- [17] S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.
- [18] S. Liao and A. C. S. Chung. A novel markov random field based deformable model for face recognition. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2675–2682, June 2010.
- [19] I. Masi, C. Ferrari, A. D. Bimbo, and G. Medioni. Pose independent face recognition by localizing local binary patterns via deformation components. In *Int. Conf. on Pattern Recognition*, pages 4477–4482, Aug. 2014.
- [20] G. Paul, K. Elie, M. Sylvain, O. Jean-Marc, and D. Paul. A conditional random field approach for face identification in broadcast news using overlaid text. In *IEEE Int. Conf. on Image Processing*, pages 318–322, Oct. 2014.
- [21] M. Pedersoli, R. Timofte, T. Tuytelaars, and L. V. Gool. Using a deformation field model for localizing faces and facial points under weak supervision. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 3694–3701, June 2014.
- [22] V. Radosavljevic, K. Ristovski, and Z. Obradovic. Gaussian conditional random fields for modeling patient’s response in acute inflammation treatment. In *Int. Conf. Machine Learning Workshop on Machine Learning for System Identification (ICML Workshop)*, 2013.
- [23] V. Radosavljevic, S. Vucetic, and Z. Obradovic. Neural gaussian conditional random fields. In T. Calders, F. Esposito, E. HÅ(ellermeier, and R. Meo, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 8725 of *Lecture Notes in Computer Science*, pages 614–629. Springer Berlin Heidelberg, 2014.
- [24] D. Roth. On the hardness of approximate reasoning. *Artificial Intelligence*, 82(1-2):273–302, 1996.
- [25] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC, 2005.
- [26] E. Sariyanidi, H. Gunes, and A. Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 37(6):1113–1133, 2015.
- [27] O. Shental, D. Bickson, P. H. Siegel, J. K. Wolf, and D. Dolev. Gaussian belief propagation for solving systems of linear equations: Theory and application. In *IEEE Int. Symp. on Information Theory*, July 2008.
- [28] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. of the IEEE*, 94(11):1948–1962, 2006.
- [29] J. M. Smereka, V. N. Boddeti, and B. V. K. Vijaya Kumar. Probabilistic deformation models for challenging periocular image verification. *IEEE Trans. on Information Forensics and Security*, 10(9):1875–1890, Sept. 2015.
- [30] K.-A. Sohn and S. Kim. Joint estimation of structured sparsity and output structure in multiple-output regression via inverse-covariance regularization. In *Proc. of the 15th Int. Conf. on Artificial Intelligence and Statistics*, volume 22, pages 1081–1089, 2012.
- [31] 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*, 30(6):1068–1080, June 2008.
- [32] X. Tan and B. Triggs. Enhanced local texture sets for face recognition under difficult lighting conditions. *IEEE Trans. on Image Processing*, 19(6):1635–1650, June 2010.
- [33] M. F. Tappen, C. Liu, E. H. Adelson, and W. T. Freeman. Learning gaussian conditional random fields for low-level vision. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2007.
- [34] 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*, 29(4):596–606, April 2007.
- [35] M. Wytock and J. Z. Kolter. Sparse gaussian conditional random fields: Algorithms, theory, and application to energy forecasting. In *Int. Conf. on Machine Learning*, volume 28 of *JMLR Proceedings*, pages 1265–1273, 2013.
- [36] M. Wytock and Z. Kolter. Sparse gaussian conditional random fields. In *Proc. of the NIPS Workshop on Log-Linear Models*, 2012.
- [37] E. Yang, P. K. Ravikumar, G. I. Allen, and Z. Liu. Conditional random fields via univariate exponential families. In *Advances in Neural Information Processing Systems 26*, pages 683–691. Curran Associates, Inc., 2013.
- [38] X.-T. Yuan and T. Zhang. Partial gaussian graphical model estimation. *IEEE Trans. on Information Theory*, 60(3):1673–1687, March 2014.
- [39] W. Y. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.
- [40] S. K. Zhou, R. Chellappa, and W.-Y. Zhao. *Unconstrained Face Recognition*. Springer-Verlag New York, Inc., 2005.
- [41] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.