Optimal Feature Learning and Discriminative Framework for Polarimetric Thermal to Visible Face Recognition

Benjamin S. Riggan[1,*] Nathaniel J. Short[1,2] Shuowen Hu[1]

[1]U.S. Army Research Laboratory, 2800 Powder Mill Rd., Adelphi, MD 20783
[2]Booz Allen Hamilton, 8283 Greensboro Dr., McLean, VA 22102
*Corresponding author: benjamin.s.riggan.ctr@mail.mil

Abstract

A face recognition system capable of day- and night-time operation is highly desirable for surveillance and reconnaissance. Polarimetric thermal imaging is ideal for such applications, as it acquires emitted radiation from skin tissue. However, polarimetric thermal facial imagery must be matched to visible face images for interoperability with existing biometric databases. This work proposes a novel framework for polarimetric thermal-to-visible face recognition, where polarimetric features are optimally combined to facilitate training of a discriminant classifier. We evaluate its performance on imagery collected under different expressions and at different ranges, and compare with recent deep perceptual mapping, coupled neural network, and partial least squares techniques for cross-spectrum face matching.

1. Introduction

For intelligence, surveillance and reconnaissance (ISR), face recognition systems capable of operating during the day and at night are highly desirable. Furthermore, accurate face recognition performance at extended standoff ranges (i.e., using low-resolution face imagery) is important. Over the past few decades, face recognition development has primarily focused on the visible spectrum, addressing challenges such as illumination, pose, and expression variations. However, visible spectrum face recognition is not suitable for covert nighttime ISR operations, due to the absence of natural illumination. In recent years, there has been growing interest in the infrared spectrum for face recognition. The infrared spectrum is divided into a reflection dominated region consisting of the near infrared (NIR) and shortwave infrared (SWIR) bands, and an emission dominated thermal region consisting of the mid-wave infrared (MWIR) and longwave infrared (LWIR) bands [1]. While facial signatures in the NIR and SWIR bands are more similar to the visible spectrum face imagery in existing biometric databases, NIR and SWIR face recognition systems typically require an active illuminator for nighttime imaging. The authors in [2-5] developed and assessed the performance of NIR and SWIR face recognition techniques. In contrast, thermal imaging in the MWIR and LWIR bands acquires naturally emitted radiation from facial skin tissue and is therefore ideal for both diurnal and nocturnal operation. However, due to the difference in phenomenology between thermal imaging and visible imaging, matching thermal face images to visible spectrum face imagery in existing biometric databases (i.e., thermal-to-visible face recognition) is highly challenging, achieving limited success. To enhance thermal face recognition, the polarization state information of thermal emissions can be exploited through an emerging technology called polarimetric thermal imaging [6-8].

The primary objective of this work is to propose a novel framework for polarimetric thermal-to-visible face recognition, where polarimetric features are optimally combined to facilitate training of a discriminant classifier. We evaluate its performance on a dataset containing varying expressions and collected at three different ranges. We compare the proposed technique with the deep perceptual mapping technique [9], a coupled neural network [10], and a partial least squares approach [11]. We found that the proposed framework significantly improves performance over these individual methods, and score level fusion methods. Furthermore, our framework not only proves to benefit polarimetric thermal-to-visible face recognition, but also significantly improves performance for conventional thermal-to-visible face recognition.

2. Background

In this section three different algorithms (partial least squares, deep perceptual mapping, and couple neural networks), which have individually demonstrated some success in cross-spectrum face recognition, are briefly discussed. More details regarding the generalization and extension of these approaches are provided in Section 3.

2.1. Partial Least Squares

Partial least squares (PLS), also known as projection to latent structures [12], was developed by Herman Wold and extended by Svante Wold. Having its first applications in the social sciences [13], PLS has been proven to be robust

to multi-collinearity [14], where the descriptor variables are highly correlated, which frequently occurs in computer vision applications. Recently, PLS has been applied successfully to several computer vision applications [11, 15-18]. Hu et al. [11] applied partial least squares regression for conventional thermal-to-visible face recognition, using a one-vs-all model building framework. PLS regression was used to build a model for each subject, using the feature vectors extracted from the visible images of a given subject as positive samples, and the feature vectors extracted from all the visible images of the remaining subjects in the gallery as negative samples. Furthermore, thermal cross-examples from a small pool of training subjects were used to augment the negative samples in learning an implicit mapping to improve the discriminative capabilities of the learned models for cross-spectrum face recognition [11].

2.2. Deep Perceptual Mapping

The DPM method in [9], like PLS, poses the problem of thermal-to-visible face recognition as a regression problem. However, unlike PLS, DPM uses a multilayer neural network to perform regression directly on the features (e.g., scale invariant feature transform (SIFT) or histogram of oriented gradients (HOG)) rather than using positive and negative labels for classification. Thus, DPM provides a simple mechanism to learn a relatively deep, nonlinear transformation that maps features from a visible image to the approximate features in the corresponding thermal image. Note that a DPM can be, alternatively, trained to predict visible HOG or SIFT features given the corresponding thermal features.

In [9], a DPM is trained to predict thermal SIFT features given the corresponding SIFT features from the visible domain. A gallery is constructed from the predicted thermal representations of the visible gallery images, and matching is performed by computing the cosine similarity between the feature representation from a thermal probe image and the predicted thermal feature representations from the gallery. Sarfraz et al. [9] demonstrated that extracting dense SIFT features followed by principal components analysis (PCA) dimensionality reduction was effective for thermal-to-visible face recognition on the UND X1 database [19-21].

2.3. Coupled Neural Network

An alternative approach for thermal-to-visible face recognition is a coupled neural network (CpNN). CpNNs, like the coupled auto-encoder method in [10], learn how to extract common latent features between visible and thermal imagery. A CpNN trains two encoders—one for the visible domain and one for the thermal domain—to produce similar features for corresponding inputs. The primary difference between CpNNs and DPMs is that CpNNs learn how to extract similar features, rather than extracting features from each domain and learning a mapping explicitly between these features.

In Section 3.1, DPMs and CpNNs are considered to be two approaches for learning correlated features. These methods are generalized and extended to use the polarimetric images, instead of conventional thermal images.

3. Methodology

3.1. Learning Feature Representations via Direct and Indirect Regression

Two methods that are used for directly or indirectly learning correlated feature representations are: DPMs and CpNNs. The former method directly learns a mapping from a visible features (HOG or SIFT) to a thermal or polarimetric features, or vice versa. The latter approach learns two mappings—visible-to-latent representation and thermal-to-latent representation—that extract common or correlated features. In either case, the objective is to learn one (or more) general transformations that produce better correlated feature representations between the visible and thermal/polarimetric domains. CpNNs may be considered a form of *indirect* regression, whereas DPMs are a *direct* regression model.

3.1.1 Deep Perceptual Mapping

In general, direct regression on raw pixel intensities (even using local patches) of thermal and visible images is a difficult task due to the significant differences between the visible and IR domains. This challenge is dramatically reduced through feature extraction and dimensionality reduction using PCA (e.g., approach of [9]).

First, HOG or SIFT features extracted from visible and IR domains are already better correlated than local patches from each domain. Additionally, these features are somewhat robust to intensity variations and image translations. Secondly, PCA reduces the size of the input to the DPM model (i.e., neural network), which further reduces the number trainable parameters. Thus, the direct regression problem is more tractable.

In this work, not only do we use a DPM model for conventional thermal image representations, we extend the DPM model to operate using polarimetric imagery (Figure 1). The polarimetric image representation consists of three components, referred to as Stokes images, $S_0, S_1,$ and S_2. The fourth Stokes image, S_3, is ignored here as it is rarely measureable and taken to be zero in practical applications [8]. S_0 captures the intensity of the thermal emissions and can be considered to be equivalent to the conventional thermal image, while S_1 and S_2 capture polarization-state information [6-8].

The DPM model may be extended to the polarimetric imagery by simply modifying the network structure. Assuming that the goal is to predict the visible features given only the polarimetric features as inputs, the network structure is modified as:

- Input: concatenate SIFT/HOG features from each of the Stokes images: $S_0, S_1,$ and S_2,
- Hidden layers: Similar number of layers and number units per layer as [8],
- Output: SIFT/HOG features from the visible (V) domain.

A DPM model is defined as a multilayer neural network with k hidden layers and a single output layer, i.e., $n = k + 1$ layers. The output from any layer ℓ may be compactly represented by:

$$\boldsymbol{f}^{\ell}(\cdot) = \begin{cases} \sigma\left(\boldsymbol{W}^{\ell}\boldsymbol{f}^{\ell-1} + \boldsymbol{b}^{\ell}\right) & 1 \le \ell < n \\ \boldsymbol{W}^{\ell}\boldsymbol{f}^{\ell-1} & \ell = n \end{cases}, \quad (1)$$

where $\boldsymbol{W}^{\ell}$ is the weight matrix, $\boldsymbol{b}^{\ell}$ is the bias vector, $\boldsymbol{f}^{\ell-1}$ is the output from the previous layer and the argument (denoted by "$\cdot$") in Eq. 1 (note that $\boldsymbol{f}^{0} = \boldsymbol{x}$, where $\boldsymbol{x}$ denotes the input), and $\sigma(\cdot)$ is the element-wise, nonlinear activation function (e.g., sigmoid, hyperbolic tangent, or ReLU). Using this notation, a sequence of functional compositions: $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{f}^{n} \circ \cdots \circ \boldsymbol{f}^{2} \circ \boldsymbol{f}^{1}(\boldsymbol{x})$ can be used to describe the entire model. Here, we use a similar network structure as [9]: two hidden layers with 200 hyperbolic tangent units each.

We extract dense SIFT features at two different scales (Gaussian smoothed with standard deviation of 0.6 and 1.0) using overlapping patches of 20×20 pixels with a stride of 8 pixels from $V, S_0, S_1,$ and S_2. Prior to feeding these features to the DPM, PCA is used to further reduce the dimensionality of each local descriptor from 128 features to 64 features. Afterwards, each feature vector is augmented with the patch locations in image coordinates to provide additional spatial dependence between input and output features. These 66-dimensional feature vectors are provided to train the DPM model, which has 119,600 trainable parameters.

Next, we describe how the modified DPM model is trained. Given the corresponding features from $V,$ and $S_0, S_1,$ and S_2, we minimize the following objective function:

$$J_{DPM}(\boldsymbol{W}^{n}, \ldots, \boldsymbol{W}^{1}, \boldsymbol{b}^{n}, \ldots, \boldsymbol{b}^{1}) = \sum_{i=1}^{M} \left\| \boldsymbol{v}_i - \widehat{\boldsymbol{v}}_i \right\|^2, \quad (2)$$

where $\boldsymbol{v}_i$ denotes the known visible features descriptor (described previously), $\widehat{\boldsymbol{v}}_i = \boldsymbol{F}(\boldsymbol{t}_i)$ denotes the estimated visible features from the corresponding thermal/polarimetric features, and M is the number of training samples. Here, $\boldsymbol{t}_i = \boldsymbol{s}_{0i}$ or $\boldsymbol{t}_i = [\boldsymbol{s}_{0i}^T \quad \boldsymbol{s}_{1i}^T \quad \boldsymbol{s}_{2i}^T]^T$ for the conventional thermal or polarimateric cases, respectively, where $\boldsymbol{s}_{ji}$ denotes the local feature extracted from S_j for $j \in \{0,1,2\}$. We minimize Eq. 2 using mini-batch gradient descent.

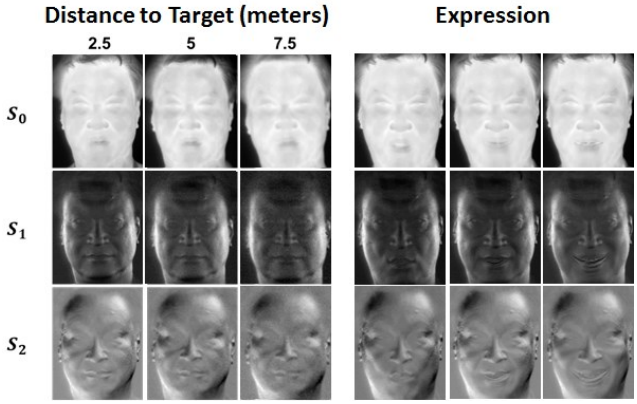

Figure 1: Example imagery acquired using polarimetric camera, illustrating loss of detail at extended ranges and variation in expressions contained within the dataset tested.

Using the trained DPM model, we estimate local visible features from local polarimetric features. The local features of an image are concatenated and L_2 normalized. Finally, we compare the normalized probe and gallery descriptors using the cosine similarity, or $\boldsymbol{v}^T\widehat{\boldsymbol{v}}$, since descriptors are normalized.

3.1.2 Coupled Neural Network

We use a CpNN model to perform a type of indirect regression, which is similar to [10]. Unlike [10], which jointly learns two mappings, we assume that one mapping is fixed. Nonetheless, the problem is still considered to be a type of indirect regression because the regression occurs in the latent space, instead of the input space as with DPM. The primary reason for fixing one of the mappings is to ensure that both DPM and CpNN are performing a one-way regression (e.g., thermal-to-visible).

A CpNN is given by two encoder networks (with k hidden layers), which are denoted by $\boldsymbol{F}_V(\boldsymbol{x}) = \boldsymbol{f}_V^{k} \circ \cdots \circ \boldsymbol{f}_V^{2} \circ \boldsymbol{f}_V^{1}(\boldsymbol{x})$ and $\boldsymbol{F}_S(\boldsymbol{y}) = \boldsymbol{f}_S^{k} \circ \cdots \circ \boldsymbol{f}_S^{2} \circ \boldsymbol{f}_S^{1}(\boldsymbol{y})$. These encoders represent multilayer neural networks, where the output from layer ℓ is defined as $\boldsymbol{f}_{\cdot}^{\ell} = \sigma(\boldsymbol{W}_{\cdot}^{\ell} f^{\ell-1} + \boldsymbol{b}_{\cdot}^{\ell})$. Here, the subscript "$\cdot$" is used to denote either V (visible) or S (polarimetric/thermal). A key difference that differentiates CpNN from DPM is that the output represents the common latent features, not a reconstruction. Also, unlike DPM, where PCA is performed on SIFT features, CpNN uses the uncompressed features as input.

Only the uncompressed features are fed to the CpNN because the encoders already reduce the dimensionality. Although, a non-linear encoder is not equivalent to PCA (only linear auto-encoders with a smaller hidden layer are equivalent [22]), there remains a loss of information when using either PCA or an encoder. Thus, reducing the

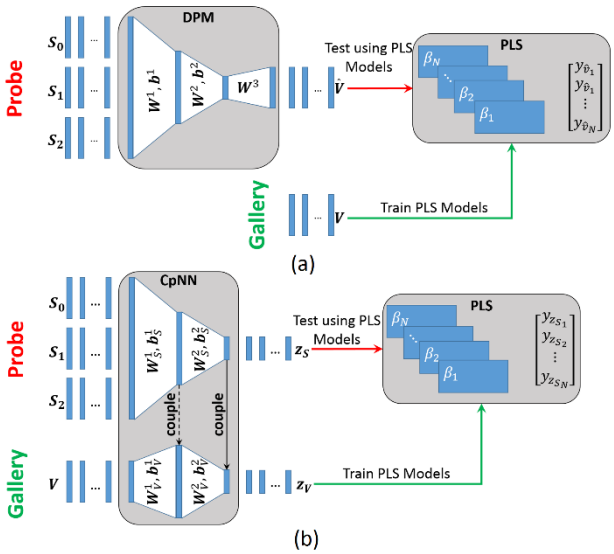

Figure 2: (a) The combined DPM and PLS framework feeds the probe inputs through the trained DPM model prior to classifying with PLS. (b) The combined CpNN and PLS framework feeds the probe and gallery inputs through the trained CpNN model prior to classifying with PLS.

dimensionality using PCA is redundant and unnecessary in this case.

Therefore, given the 128-dimensional dense SIFT features from V, S_0, S_1,and S_2, we can train a CpNN. First, the visible encoder, $\boldsymbol{F}_V$, is trained so that the input (visible features) may be optimally reconstructed from the encoder output, much like a stacked auto-encoder [23, 24]. After this encoder is trained, the parameters $\boldsymbol{W}_V^{\ell}, \boldsymbol{b}_V^{\ell}$ for $\ell = 1 \dots k$ are fixed during the coupling procedure. In order to couple the latent features, the thermal encoder, $\boldsymbol{F}_S$, is trained to minimize the following objective:

$$J_{CpNN}(\boldsymbol{W}_S^k, \dots, \boldsymbol{W}_S^1, \boldsymbol{b}_S^k, \dots, \boldsymbol{b}_S^1) = \sum_i^M \|\boldsymbol{z}_{Vi} - \boldsymbol{z}_{Si}\|^2, \quad (3)$$

where $\boldsymbol{z}_{Vi} = \boldsymbol{F}_V(\boldsymbol{v}_i)$ and $\boldsymbol{z}_{Si} = \boldsymbol{F}_S(\boldsymbol{t}_i)$. This objective (Eq. 3), which seeks a solution where $\boldsymbol{z}_{Vi} \approx \boldsymbol{z}_{Si}$, is also minimized using mini-batch gradient descent optimization.

In this work, we used a CpNN composed of two hidden layers with 92 and 64 units respectively. This results in only 41,372 trainable parameters in practice, since in an operational setting one may assume that a visible encoder (i.e., feature extractor) is already trained.

After training, these encoders are used to extract the common latent features from the visible gallery images and the probe thermal/polarimetric images. After extracting local features, we normalize image descriptors (Section 3.1.1). Then, using each of the normalized probe and gallery image descriptors, we compute the match scores using the cosine similarity, or $\boldsymbol{z}_V^T \boldsymbol{z}_S$ since both descriptors are normalized.

3.2. Discriminative Regression (PLS)

In PLS regression, let $\boldsymbol{X}_{m \times d}$ be a descriptor variable matrix from m samples with feature vector dimensionality d, and $\boldsymbol{y}_{m \times 1}$ be the corresponding univariate response variable (i.e., label vector). An entry in $\boldsymbol{y}$ is +1 if the corresponding sample in $\boldsymbol{X}$ is a positive sample in the one-vs-all model building framework, otherwise the entry's value is -1 for a negative sample. PLS regression finds a set of latent vectors from $\boldsymbol{X}$ that best predict $\boldsymbol{y}$, as defined in the following equations:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{X}_{res}, \quad (4)$$
$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{q}^T + \boldsymbol{y}_{res}. \quad (5)$$

In Eqs. 4 and 5, $\boldsymbol{T}$ and $\boldsymbol{U}$ are score matrices containing the extracted latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$, $i = 1 \dots p$. $\boldsymbol{P}$ and $\boldsymbol{q}$ are referred to as the loadings, and $\boldsymbol{X}_{res}$ and $\boldsymbol{y}_{res}$ are the residuals. To compute the latent vectors, the following optimization problem:

$$\max \operatorname{cov}(\boldsymbol{t}_i, \boldsymbol{u}_i)^2 = \max_{\boldsymbol{w}_i} \operatorname{cov}(\boldsymbol{X}\boldsymbol{w}_i, \boldsymbol{y})^2 \ \forall i \ s.t. \ |\boldsymbol{w}_i| = 1, \quad (6)$$

is solved iteratively, generating a weight matrix $\boldsymbol{W}$ containing a set of weight vectors $\boldsymbol{w}_i$.
Then, $\boldsymbol{W}$, $\boldsymbol{P}$, $\boldsymbol{T}$, and $\boldsymbol{y}$ can be used to compute a regression vector $\boldsymbol{\beta} = \boldsymbol{W}(\boldsymbol{P}^T\boldsymbol{W})^{-1}\boldsymbol{T}^T\boldsymbol{y}$. Given a testing feature vector $\boldsymbol{f}$ (i.e., feature vector from a thermal probe image), a similarly score y_f for the given subject's model is computed as:

$$y_f = \bar{y} + \boldsymbol{\beta}^T \boldsymbol{f}. \quad (7)$$

Given N subjects in the gallery database represented by their PLS regression models, a similarity score between the probe feature vector and every subject's model is computed. This set of similarity scores can then be used for face identification or face verification.

Here, we fuse complementary features from S_0, S_1, and S_2 by computing the average across the local feature sets derived from the Stokes images, producing a feature set that is strongly correlated with the visible counterpart. While this approach has shown good performance for discriminative classification [6, 7], in the following sections we discuss two alternatives.

3.3. Score level fusion

To further improve the performance of the recognition system, we consolidate the output of the discriminative regression and feature regression approaches into a single score by performing match score level fusion. The match score indicates a measure of similarity between two feature sets, given by the respective classifiers. In our case, the thermal (or polarimetric thermal) and visible feature sets. In this paper, we perform a simple sum score fusion, where the match score from PLS is combined with the match score of

DPM (or CpNN) through simple addition of the normalized scores. The motivation behind this approach is that the strengths of one technique can be used to supplement another's weaknesses. The combined score is then used to represent the similarity of the probe and gallery sample.

3.4. Proposed Recognition Framework

Here, we combine the workflow of the complementary discriminative regression, PLS, and feature regression, DPM and CpNN, approaches into a sequential framework. We train DPM and CpNN as described in Section 3.1 and learn the best combination of features from the Stokes images (i.e., the feature fusion module) and train PLS to discriminate between match and non-match classes (i.e., the classification module) using the encoded feature sets.

Both of the proposed frameworks, namely PLS ∘ DPM and PLS ∘ CpNN, are depicted in Figure 2. For the PLS ∘ DPM framework, the DPM model is trained to estimate the visible features from the given polarimetric/thermal features. Then, PLS is trained as discussed in Section 3.2, except that cross-examples are the estimated visible image descriptors from DPM. Therefore, substituting the visible estimates, $\hat{\boldsymbol{v}}$, for $\boldsymbol{f}$ into Eq. 7 yields

$$y_{\hat{v}} = \bar{y} + \boldsymbol{\beta}^T \hat{\boldsymbol{v}}, \qquad (8)$$

which denotes the similarity score.

Also, for the PLS ∘ CpNN framework, the CpNN model is trained to extract similar feature from both visible and polarimetric/thermal inputs. Then, PLS is trained using encoded visible image descriptors from CpNN as either positive or negative examples, and with the cross-examples given by the encoded polarimetric/thermal descriptors from CpNN. Therefore, substituting the encoded polarimetric/thermal descriptors, $\boldsymbol{z}_S$, for $\boldsymbol{f}$ into Eq. 7 yields

$$y_{z_s} = \bar{y} + \boldsymbol{\beta}^T \boldsymbol{z}_S, \qquad (9).$$

4. Experiments & Results

To quantitatively assess the performance improvement provided by polarimetric imaging for cross-spectrum face recognition, we use a dataset of simultaneously acquired visible and polarimetric thermal facial imagery of 50 subjects, an extension of the 20 subject dataset used in [6]. We select a subset of this data to analyze performance consisting of the following conditions: baseline (range 1), expression, range 2 and range 3 (Figure 1). The range data corresponds to facial imagery resolutions of 87, 44, and 31 pixels between the eyes at distances of 2.5, 5, and 7.5 meters, respectively.

All images are registered as in [25] and preprocessed as in [11]. In this work, we use a closely cropped 143x132 region of the aligned face, which provides a degree of invariance to changes in hair style and pose (as opposed to the loose crop used by Klare et al. [25]). Following typical evaluation protocols, we split our dataset into equally sized training and testing sets, containing 25 subjects each. Each subject has four samples from each modality, visible, thermal, and polarimetric thermal. The training set is used to learn mappings from thermal and polarimetric-thermal feature space to corresponding visible feature space using feature regression, as described in Section 3.1, and for selecting counter examples and negative examples to train models for PLS. In this work, we use all baseline (range 1) visible samples and five baseline thermal and polarimetric-thermal features from the training set to serve as negative- and counter-examples for training the discriminative classifier. For range 2 and 3 data, we use the five thermal and polarimetric-thermal counter-examples from the training set of subject imagery at the respective ranges. Corresponding visible and thermal/polarimetric thermal imagery from the training subjects at range 1 is used to perform the feature learning using DPM and CpNN.

Finally, each scenario is tested 100 times, where training and testing groups are created by randomly sampling, without replacement, from the entire set of subjects. To directly compare performance, the same training/test groups are used for PLS, DPM, and CpNN for each test. Performance is reported as the average of the 100 tests, where four probe samples (12 in the case of expression data) from the testing set are matched against each gallery model. Gallery models contain information from all four visible, and only-visible, images corresponding to each of the 25 subjects in the testing dataset. The results of testing are summarized in the rank-based Cumulative Match Characteristic (CMC) curves shown in Figure 3. Figure 3a shows the overall performance of each individual method in all scenarios for both polarimetric-thermal (polar) and conventional thermal probes (therm). While DPM and CpNN methods use SIFT features as input, we use HOG features for PLS as suggested in [10], Figure 3b shows improved results in overall performance achieved by performing score level fusion of the feature regression and discriminative regression approaches. Also in Figure 3b, we show the results of the developed framework for polarimetric feature fusion and classification using the approach described in Section 3.4. In this experiment, all methods use SIFT features as input.

These curves further show a significant improvement in cross-spectrum face recognition performance when incorporating polarization-state information into the feature set for classification. Figure 3a shows that we are able to provide a 17% improvement and 9–11% improvement overall to rank-1 identification rate when using discriminative regression or feature regression, respectively, on a diverse dataset. Figure 3b reveals a slight decrease in performance when comparing the best performing individual methods (Figure 3a) with the corresponding score level fusion methods. This is best

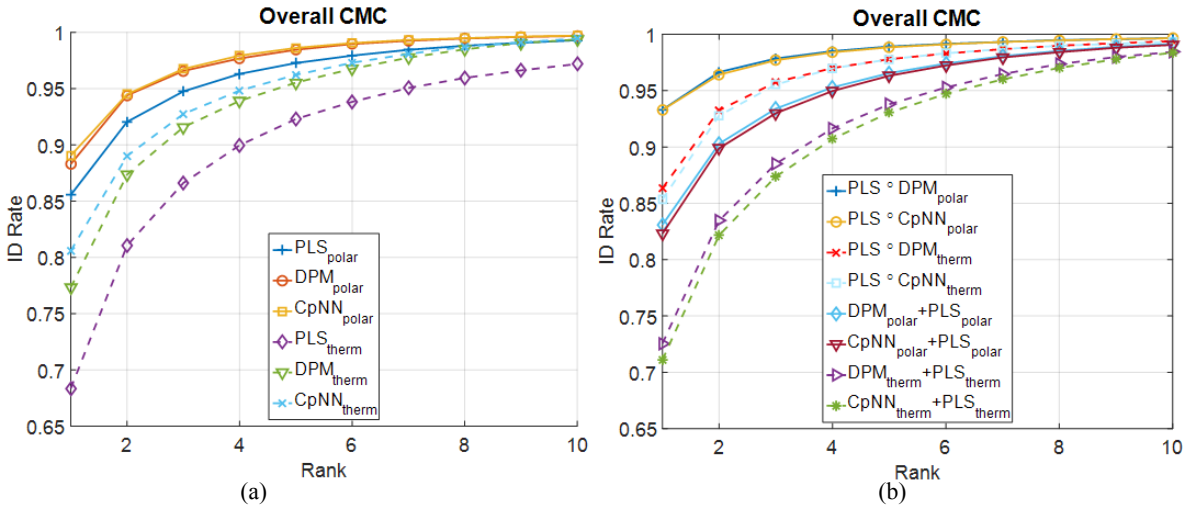

Figure 3: (Left) Overall performance from testing PLS, DPM, and CpNN using polarimetric (polar) samples and thermal (therm) samples as probes. (Right) Overall performance fusing unsupervised and supervised approaches. Plots show average CMC curve from 100 trials.

explained by the fact that the score level fusion is most beneficial when the normalized scores are complementary. In this case, since PLS performs the worsts out of all the methods, its poor performance brings down performance of DPM or CpNN. However, the proposed framework (Section 3.4) is shown to overcome the limitations of PLS by providing the best overall recognition performance for both conventional thermal and polarimetric thermal cases.

In Table 1, rank-1 identification rates achieved by individual algorithms outlined in the paper are reported for each evaluation scenario. It can be seen that the feature regression methods outperform the discriminative classifier across all evaluations: baseline (range 1), expression, and ranges 2 and 3. When comparing the best performing individual method: PLS, DPM, or CpNN, against the proposed framework, improvements of 4.9%, 6.4%, and 2.9% for the thermal expression, range 2, and range 3 conditions, respectively, are achieved. Also, the proposed framework achieves an improvement of 4.4%, 2.6%, and 4.3% for the polarimetric expression, range 2, and range 3 cases, respectively.

Table 1: Breakdown of performance results from testing polarimetric and thermal imagery acquired under various conditions. Rank 1 identification rates are reported as the average of the 100 trials.

Scenario	Rank 1 Identification Rate				
	Probe	**PLS**	**DPM**	**CpNN**	**Proposed Framework**
Overall	***Polar***	0.855	0.883	0.891	**0.933**
	Therm	0.683	0.773	0.806	**0.863**
Baseline	***Polar***	0.953	0.943	0.943	**0.988**
	Therm	0.818	0.889	0.906	**0.952**
Expression	***Polar***	0.903	0.903	0.918	**0.962**
	Therm	0.717	0.798	0.836	**0.885**
Range 2	***Polar***	0.845	0.908	0.900	**0.934**
	Therm	0.635	0.766	0.797	**0.861**
Range 3	***Polar***	0.625	0.736	0.747	**0.790**
	Therm	0.496	0.590	0.625	**0.654**

The table also indicates that we are able to achieve a significant improvement in extended range cross-spectrum face recognition performance by using polarimetric features in addition to total intensity thermal. We achieve an improvement in performance of 7.7%, 7.3%, and 13.6% for the expression, range 2, and range 3 cases, respectively, when incorporating polarization-state information into the probe feature set for cross-spectrum matching.

5. Conclusions

The aim of this work was to develop and evaluate a polarimetric thermal-to-visible face recognition framework that exploits methods for learning robust feature representations and for discriminative classification. Individually, the feature learning components, such as DPM or CpNN, were optimized to extract common information from the two domains. The performances of DPM and CpNN were shown to exceed that of PLS alone in both the conventional thermal-to-visible and polarimetric thermal cases. The combinations DPM+PLS and CpNN+PLS appeared to yield worst recognition performance due to the poor performance of PLS.

In this work, we presented CpNN and DPM as two methods for feature regression. The results revealed that in almost all cases CpNN performed better than DPM. Moreover, the number of trainable parameters for CpNN is fewer than DPM in an operational scenario. Fewer parameters may reduce the potential for over fitting.

The proposed framework that exploits the complementary nature of feature regression with discriminative classification was extensively evaluated on a diverse dataset that includes both visible and polarimetric thermal imagery, which also incorporates various ranges and facial expressions. This framework achieved the top performance on every evaluation set: baseline, expression, range 2, and range 3. Moreover, this approach was able to

overcome some of the disadvantages of the individual methods. Also, our results agree with the findings from [6, 7] which demonstrated a boost in recognition performance when using polarimetric thermal imaging.

6. Acknowledgements

This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-12-2-0019. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, and M.A. Abidi, "Recent advances in visual and infrared face recognition – a review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103-135, 2005.

[2] B. Klare and A.K. Jain, "Heterogeneous face recognition: matching NIR to visible light images," *in Proc. Int'l. Conf. on Pattern Recognition*, pp. 1513-1516, 2010.

[3] F. Juefei-Xu, D. K. Pal, and M. Savvides, "NIR-VIS heterogeneous face recognition vis cross-spectral joint dictionary learning and reconstruction," *in Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 141–150, 2015.

[4] T. Bourlai, N. Kalka, A. Ross, B. Cukic, and L. Hornak, "Cross-spectral face verification in the short wave infrared (SWIR) band," *in Proc. Int'l. Conf. on Pattern Recognition*, pp. 1343-1347, 2010.

[5] F. Nicolo and N.A. Schmid, "Long range cross-spectral face recognition: matching SWIR against visible light images," *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 6, 2012.

[6] N. Short, S. Hu, P. Gurram, et al., "Improving cross-modal face recognition using polarimetric imaging", *Optics Letters*, vol. 40(6), pp. 882-885.

[7] N. Short, S. Hu, P. Gurram, et al., "Exploiting polarization-state information for cross-spectrum face recognition", Proc. *IEEE Biometrics: Theory, Applications, Systems*, Sept. 2015.

[8] Gurton K, Yuffa A, Videen G, "Enhanced facial recognition for thermal imagery using polarimetric imaging", *Optics. Letters*, vol. 39, no. 13, pp. 3857-3859, 2014.

[9] M. S. Sarfraz, R. Stiefelhagen, "Deep Perceptual Mapping for Thermal to Visible Face Recognition", in Proc. British Machine Vision Conference, Sept. 2015.

[10] B. S. Riggan, C. Reale, N. M. Nasrabadi, "Coupled Auto-Associative Neural Networks for Heterogeneous Face Recognition", *IEEE Access*, vol. 3, pp. 1620–1632, 2015.

[11] S. Hu, J. Choi, A.L. Chan, W.R. Schwartz, "Thermal-to-visible face recognition using partial least squares", *Journal of the Optical Society of America A*, vol. 32(3), pp. 431-442, 2015.

[12] S, Wold, M. Sjöström, L. Eriksson. "PLS-regression: a basic toll of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.

[13] I. Helland, "Partial least squares regression," *in Encyclopedia of Statistical Sciences* (Wiley, 2006), pp. 5957–5962.

[14] H. Wold, "Estimation of principal components and related models by iterative least squares," in Multivariate Analysis (Academic, 1966).

[15] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," *in Proc. Int'l Conf. on Computer Vision*, pp. 24–31, 2009.

[16] A. Kembhavi, D. Harwood, and L. S. Davis, "Vehicle detection using partial least squares," *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 1250–1265, 2011.

[17] A. Sharma and D. Jacobs, "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch," *in Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 593–600, 2011.

[18] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face identification using large feature sets," *IEEE Trans. Image Process.*, 21, pp. 2245–2255, 2012.

[19] P. J. Flynn, K.W. Bowyer, and P.J. Phillips, "Assessment of time dependency in face recognition: An initial study," *Audio and Video-Based Biometric Person Authentication*, pp. 44-51, 2003.

[20] X. Chen, P.J. Flynn, and K.W. Bowyer, "Visible-light and Infrared Face Recognition," *in Proc. ACM Workshop on Multimodal User Authentication*, pp. 48-55, 2003.

[21] X. Chen, P.J. Flynn, and K.W. Bowyer, "IR and visible light face recognition," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 332-358, 2005.

[22] P. Baldi and K. Hornik, "Neural netwoks and principal component analysis: learning from examples without local minima," *Neural Networks*, vol. 2, no. 1 pp. 53-58, 1989.

[23] H. Larochelle, D. Erhan, and P. Vincent, "Deep learning using robust interdependent codes," *in Proc. Int'l Conf. on Artificial Intelligence and Statistics*, pp. 312-319, 2009.

[24] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing features with denoising autoencoders," *in Proc. Int'l Conf. on Machine Learning*, pp. 1096-1103, 2008.

[25] B.F. Klare and A.K. Jain, "Heterogeneous face recognition using kernel prototype similarity," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1410-1422, 2013.