Cross-domain face recognition using coupled independent component analysis

Dr. Xiangyan Zeng, Fort Valley State University

Xiangyan Zeng received her Ph.D. in computer science from University of the Ryukyus, Japan. She is currently a professor of computer science at Fort Valley State University. Her research interests include image processing, pattern recognition and machine learning.

Dr. Masoud Naghedolfeizi, Fort Valley State University

Dr.Masoud Naghedolfeizi is a senior professor in the Department of Mathematics and Computer Science at Fort Valley State University. His academic background includes Ph.D. in Nuclear Engineering, M.S. in Materials Science , and a B.S. in Mechanical Engineering with minor in instrumentation and control systems

CHUNHUA DONG, Fort Valley State College

Cross-domain face recognition using coupled independent component analysis

Introduction

Face recognition has been the prominent biometric technique for identity authentication and widely used in the commercial, military, and government sectors. The research endeavors have been largely focused on the visible spectrum. However, the lack of illumination during nighttime makes visible imaging sensors less effective. Thermal imaging is a way to improve the visibility of objects in a dark environment. Thermal cameras detect heat given off by an object and convert temperature information into images and therefore are a highly practical imaging modality for nighttime operation.

As most databases only contain facial imagery in the visible spectrum, heterogeneous face recognition (HFR) is essential for matching a thermal probe image against visible images. The large modality gap between faces captured in different spectra makes HFR a challenging problem. Numerous studies [1-4] have been conducted on reducing the large domain gap to facilitate cross-spectral recognition. Klare et al. [5] developed a generic framework based on nonlinear kernel prototype similarities for cross-modal matching. Hu et al. [6] proposed a one-vs-all framework using binary classifiers based on partial least squares (PLS), where thermal counter examples were used to augment the negative sample set during training. Recently, the great progress in high-quality face synthesis has made "recognition via synthesis generation" possible. Zhang et al. [7] proposed a generative adversarial network (GAN) to synthesize polarimetric thermal images into visible domain. Di et al. [8] recently proposed a more stable GAN network called AP-GAN for synthesizing visible faces from polarimetric thermal counterparts via attribute preserved synthesis. Zhang et al. [9] proposed TV-GAN which used the base pix2pix model with incorporated identity loss in the generator to synthesize faces in the visible RGB domain from their counterparts in the thermal-infrared domain.

In the meantime, developments in coupled architecture have shown promising results for cross-spectral synthesis and recognition problems. Reale et al. [10] achieved thermal-to-visible face recognition by learning coupled dictionaries in the two domains, where the dictionaries were obtained through bi-level optimization using stochastic gradient descent. Iranmanesh et al. [11] proposed a coupled deep neural network architecture that was able to find global discriminative features in a nonlinear embedding space to relate the polarimetric thermal faces to their corresponding visible faces.

Motivated by the recent advances in face recognition algorithms using coupled architecture, we proposed a coupled independent component analysis (Coupled ICA) architecture for cross-spectral thermal-to-visible face synthesis and recognition. Independent component analysis (ICA) is a well-known statistical analysis tool that estimates the original source signal from observed mixture signals. Previous research has suggested there exists a connection between ICA and the human visual system [12-13]. ICA of natural image patches produces the sparse representation of the image patches and a set of high-pass filters resembling the receptive fields of simple cells in visual cortex. ICA was recently applied to thermal facial images and was able to capture temperature fluctuations from the movements of eyes, respiration, et al. [14]. The Coupled ICA approach assumes a pair of corresponding thermal and visible image patches are from the same sources observed by the imaging systems represented by two different ICA projection matrices. The source space is a common latent space between the visible and the thermal domains. Firstly, the Coupled ICA process is performed through jointly updating the projection matrices in the two domains over a set of training data. Then, the obtained ICA filters are used to transform images into a domain-independent latent space via patch-wise synthesis. The proposed Coupled ICA framework forces the filtered images of a pair of corresponding images in the two domains to be identical in the common latent space. Finally, classification is conducted using the multi-class sparse representation (SR) algorithm.

While many methods in the literature yield better results on polarimetric thermal images than thermal images, the proposed method is specifically effective for conventional thermal images. This indicates that the coupled ICA acquires the 2D point spread function information in the basis functions and cannot capture the 3D geometric information in the polarimetric images.

Experimental Methods/Materials/Project Approach

In this section, we will describe the details of the proposed heterogeneous face recognition system. First, a coupled ICA model is trained using a set of visible and corresponding thermal image patches. Then visible and thermal images are synthesized using the filters obtained from the coupled ICA model. Lastly, a multi-class sparse representation (SR) framework is used for recognizing a thermal face identity in a visible facial gallery.

Independent Component Analysis of Images

Olshausen and Field [12] applied a sparseness-maximization network to natural image patches to find efficient codes for natural images. Their research demonstrates that a network that learns sparse codes of natural scenes produces a set of basis functions that are spatially localized, oriented, and bandpass (i.e., wavelet-like). Each image patch can be represented as a linear combination of the 'basis' patches, such that the mixing coefficients are as sparse as possible. Subsequently, Bell and Sejnowski [13] applied independent component analysis to similar data, but the purpose is to seek mutually independent components. It has led to a local representation similar to that obtained through sparse coding. In addition, independent component analysis has yielded localized edge detectors. In this section, we briefly describe the general ICA model of image analysis.

Image that each image patch, represented by a vector $\mathbf{x} = (x_1, x_2, \cdots\ x_d)$, is a linear combination of d basis functions that are fixed for all image patches. The coefficients of the linear combination vary with image patches. Assume that the basis functions form the columns of a projection matrix $\mathbf{A}$, this perceptual process can be described by:

$$\mathbf{x} = \sum_{i=1}^{d} \mathbf{a}_i\, s_i = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{a}_i$ $(i = 1, \cdots, d)$ are the basis functions of the perceptual system, and s_i is the scalar coefficient of $\mathbf{a}_i$. Given a set of random image patches, the goal of the ICA image analysis is to find a matrix $\mathbf{W} = \mathbf{A}^{-1}$ that represents the inverse of the perceptual process, so that the components of the resulting vector:

$$\mathbf{u} = \mathbf{W}\mathbf{x} \tag{2}$$

are mutually independent. The rows of matrix $\mathbf{W}$ can be used as filters to transform images obtained by the same system to their corresponding independent sources.

Coupled Independent Component Analysis Model

Heterogeneous Face Recognition (HFR) is a challenging issue because of the large domain discrepancy. Much research has been devoted to transforming images into latent space to reduce domain discrepancy [15-17]. Motivated by the findings of the above-mentioned ICA research, we propose a coupled independent component analysis (Coupled ICA) model to obtain a set of basis functions in each relevant domain. The coupled ICA aims to represent a visible image patch and its corresponding thermal image patch as two different linear combinations of the same coefficients i.e. the photographed object. Our goal is to transform face images in the infrared spectrum and the visible light spectrum into a new latent space to reduce the domain discrepancy, as shown in Figure 1.

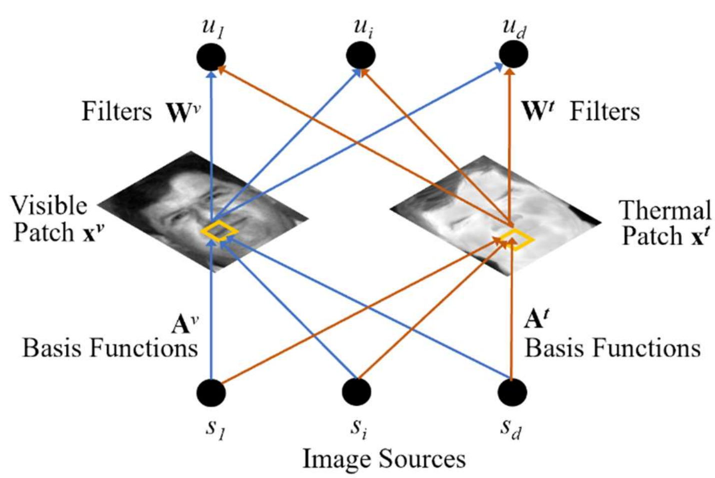

Figure 1. Illustration of Coupled ICA model for cross-domain linear image synthesis. Each patch is represented as a linear combination of basis functions, each associated with an element of an underlying vector of statistically independent "image sources". The image sources are recovered by a matrix of filters, more loosely "receptive fields". In Coupled ICA, it assumes the image sources from visible domain to be identical to those observed at the thermal domain.

Let $\mathbf{X}^v = [\mathbf{x}_1^v, \ldots, \mathbf{x}_n^v] \in R^{d \times n}$ be a a set of n d-dimentional visible image patches and $\mathbf{X}^t = [\mathbf{x}_1^t, \ldots, \mathbf{x}_n^t] \in R^{d \times n}$ be their corresponding thermal counterparts. A coupled ICA is performed on $\mathbf{X}^v$ and $\mathbf{X}^t$ as follows:

$$\mathbf{u} = \overline{\mathbf{W}}\overline{\mathbf{X}} = [\mathbf{W}^v, \mathbf{W}^t]\begin{bmatrix}\mathbf{X}^v \\ \mathbf{X}^t\end{bmatrix} \tag{3}$$

where $\mathbf{W}^v \in R^{k\times d}$ and $\mathbf{W}^t \in R^{k\times d}$ are the filter matrices for the two modalities. We use the FastICA algorithm to obtain the transform matrix $\overline{\mathbf{W}} = (\overline{\mathbf{w}}_1, \cdots \overline{\mathbf{w}}_i, \cdots \overline{\mathbf{w}}_d)^T$, where $\overline{\mathbf{w}}_i$ is a concatenated row vector of $\mathbf{W}^v$ and $\mathbf{W}^t$. A preprocessing is performed to whiten the image patch data. Then the FastICA learning rule iteratively finds a vector $\overline{\mathbf{w}}_i$ $(i = 1, \cdots d)$ so that the projection ${\overline{\mathbf{w}}_i}^T\overline{\mathbf{X}}$ maximizes the nongaussianity. The Coupled ICA learning process enforces that the optimization is in the concatenated domain space of $\mathbf{X}^v$ and $\mathbf{X}^t$, but not in each domain separately. The learned $\mathbf{W}^v$ and $\mathbf{W}^t$ are used to transform the visible and thermal image patches respectively to enhance the common features in these two domains.

Image Synthesis Based on the Coupled ICA Model

A set of visible image patches and their corresponding thermal patches are selected for the Coupled ICA training. The learned $\mathbf{W}^v$ is used to transform each visible face image in the gallery through patch-wise synthesis. For each pixel in an image, its surrounding $\sqrt{d} \times \sqrt{d}$ subwindow is transformed by the ICA filters. Because the subwindows are overlapping, d transformations are obtained for each pixel. The final synthesis result is the mean of the d transformations. A probe thermal face image $\mathbf{x}^t$ is similarly synthesized using $\mathbf{W}^t$ and matched against the transformed visible images in the gallery. An example of image synthesis based on the coupled ICA is shown in Figure 2.

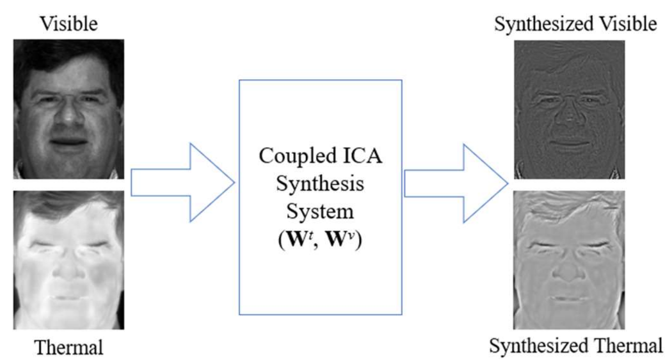

Figure 2. An example of image synthesis based on coupled ICA.

Face Recognition by Multi-class Sparse Representation

The proposed thermal-to-visible face recognition is performed on the synthesized thermal and visible images. Firstly, we employ a difference of Gaussians (DoG) filter and a nonlinear hyperbolic tangent sigmoid (tanh) function to enhance edge information and contrast. Secondly, the histogram of oriented gradients (HOG) features is extracted to reduce the gaps between the visible and thermal facial signatures. Finally, the classification is performed using the multi-class sparse representation (SR) discriminative model.

Preprocessing

To increase the correlation between the visible and thermal modalities, each face image is preprocessed. We apply a DoG band-pass filter, to emphasize the edges in addition to removing high and low frequency noise. Given an image I, the DoG filtering with two Gaussian kernels σ_0 and σ_1 is defined as follows:

$$I_{DoG}(x,y) = [G(x,y,\sigma_0) - G(x,y,\sigma_1)] * I(x,y)$$
$$G(x,y,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4}$$

where I_{DoG} is the DoG filtered image, $*$ is the convolution operator, and G is the Gaussian kernel. Furthermore, we apply a nonlinear hyperbolic tangent (tanh) sigmoid transformation on the DoG filtered image to enhance the edge information:

$$T\left(I_{DoG}(x,y)\right) = tah\left(\frac{I_{DoG}(x,y)}{\sigma_{DoG}}\right) \tag{5}$$

where σ_{DoG} is the standard deviation of the DoG filtered image I_{DoG}.

Histogram of Oriented Gradients Descriptor

The histogram of oriented gradients (HOG) is a feature descriptor that represents the local object appearance and shape information in an image. It comprises the distribution of intensity gradients and edge directions. HOG has been widely used for extracting facial features. We extract the HOG features of face images and feed them to the subsequent multi-class SR classification.

Sparse Representation for Face Recognition

In this study, we use the multi-class sparse representation (SR) model [18-20] for face recognition. In addition, we add thermal cross-examples of unrelated subjects to improve the discriminative capabilities of the cross-spectral face recognition system, as shown in Figure 3.

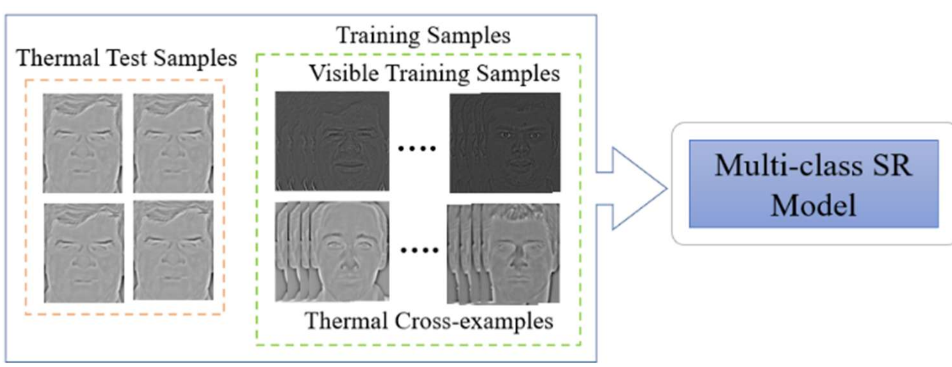

Figure 3. Multi-class sparse representation (SR) for visible-to-thermal face recognition

In the multi-class SR model, a test sample is represented by the linear combination of training samples, ideally only by training samples of the class associated with the test sample. Suppose *N* features are extracted from images and there are C known subject classes; each class has L visible training samples. Let $\mathbf{D}_i = [\mathbf{d}_{i,1}, \mathbf{d}_{i,2}, \dots, \mathbf{d}_{i,L}] \in R^{N \times L}$ be a matrix formed by the feature vectors of visible training samples of the i-th class and $\mathbf{D}_{cross} \in R^{N \times M_{cross}}$ be a matrix of M_{cross} thermal cross-examples, then $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_C, \mathbf{D}_{cross}] \in R^{N \times (LC + M_{cross})}$ is a dictionary composed of all visible training samples and thermal cross-examples. Given a thermal test sample $\mathbf{y}$, it can be represented by $\mathbf{y} = \mathbf{D}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in R^{LC+M_{cross}}$ is the representation coefficient vector. This linear equation system is underdetermined if $N < LC + M_{cross}$then its sparsest representation can be obtained by solving the following optimization problem:

$$\widehat{\boldsymbol{\alpha}} = argmin\|\boldsymbol{\alpha}\|_1 \;\; s.t. \;\; \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2 < \epsilon \tag{6}$$

where $\|*\|_1$ denotes the L1-norm, which counts the number of nonzero entries in a vector, and $\epsilon > 0$ is a constant. Since $\widehat{\boldsymbol{\alpha}}$ is achieved, using only the coefficients associated with the i-th class, the sparse coding error (or residual, SCE) is calculated by:

$$\gamma_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{D}_i \delta_i(\widehat{\boldsymbol{\alpha}})\|_1 \quad (i = 1,2, \dots, c) \tag{7}$$

where function $\delta_i : R^N \to R^N$ selects the coefficients associated with the i-th class. The SR classifier assigns the thermal test sample $\mathbf{y}$ to the class k if $\gamma_k(\mathbf{y}) = \min \gamma_i(\mathbf{y})$.

Results and Discussion

Database Description and Implementation Details

Experiments were conducted using two heterogeneous face datasets: the ARL multi-modal face dataset [21] and the Tufts face dataset [22], to evaluate the performance of the proposed thermal-to-visible face recognition approach. Each dataset was collected with different sensor systems and contains images acquired under different experimental conditions. The evaluation was repeated 100 times, and the average results were used to validate the robustness of the approach. While the numbers of training data and test data are the same as in referenced studies for fair comparison. the dataset was randomly partitioned into training and test data in each trial.

ARL Multi-modal Face Dataset

The ARL multi-modal face dataset comprises thermal face images, including conventional thermal and polarimetric thermal images, and their corresponding visible images related to 60 subjects [21]. Data was collected at three different distances: Range 1 (2.5 m), Range 2 (5 m), and Range 3 (7.5 m). An example of image pairs in the three ranges is shown in Figure 4. In addition, each subject has 4 baseline images with a neural expression and 12 images with different expressions in the mouth and the eyes. All images were normalized to 250 × 200 pixels.

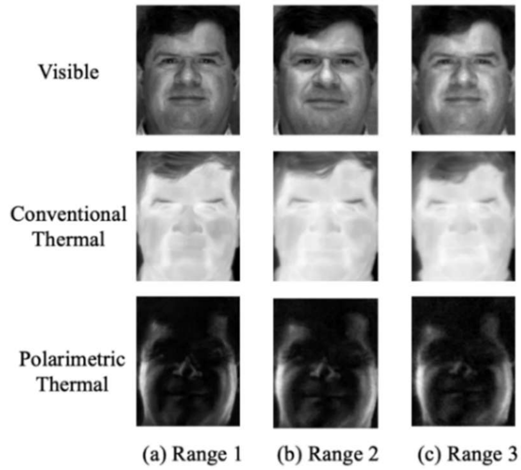

Figure 4. Example ARL dataset imagery from the baseline condition, showing visible, conventional thermal and polarimetric thermal images at three ranges

Tufts Face Dataset

Tufts face database is a publicly available face dataset collected from students, staff, faculty, and their family members at Tufts University [22]. It contains 7 image modalities: visible, near-infrared, thermal, computerized sketch, video, LYTRO and 3D images. In our study, we used paired thermal and RGB face image data. The dataset contains images of 128 × 128 pixels for 112 subjects including 74 females and 38 males, from 15 countries. Each subject has images with different expressions of neutral, smile, sleepy, and surprised, occlusion with sunglasses, and side views at various angles. In our study, we transformed all RGB images into gray scale images. Example images of a subject in this dataset are shown in Figure 5.

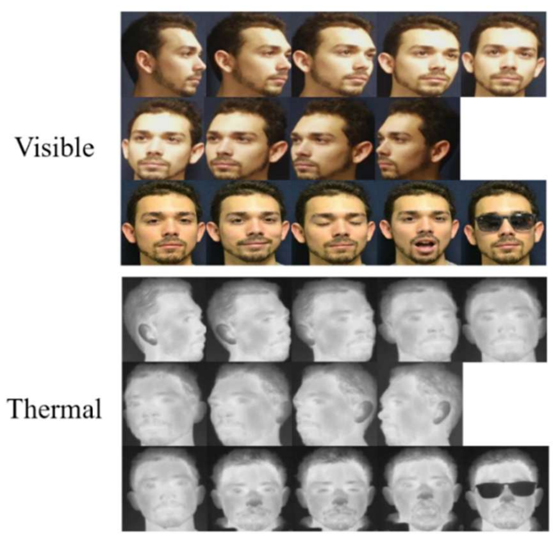

Figure 5. An example of a subject in Tufts dataset. It includes paired visible and thermal images with different views and expressions.

Qualitative Analysis

To evaluate the performance of our proposed method, we applied the method to each of the two datasets. Among the 60 subjects in the ARL dataset, 25 subjects were used to train the Coupled ICA model, and the remaining 35 subjects were used for testing the face recognition framework. The training patches for the Coupled ICA were randomly selected from the training images. The Range 1 baseline visible images of the 35 test subjects were used as positive and negative samples (dictionary atoms) for building the SR model. In addition, the Range 1 thermal images of the 25 training subjects were added as cross-examples to the SR model.

Before the HOG features were extracted and fed to the SR classifiers, each image was processed using the extended DoG to enhance edge information and contrast. The effectiveness of image synthesis and the extended DoG preprocessing is demonstrated in Figure 6. Figure 6(d) demonstrates the extended DoG of the synthesized images is able to capture meaningful information in both modalities.

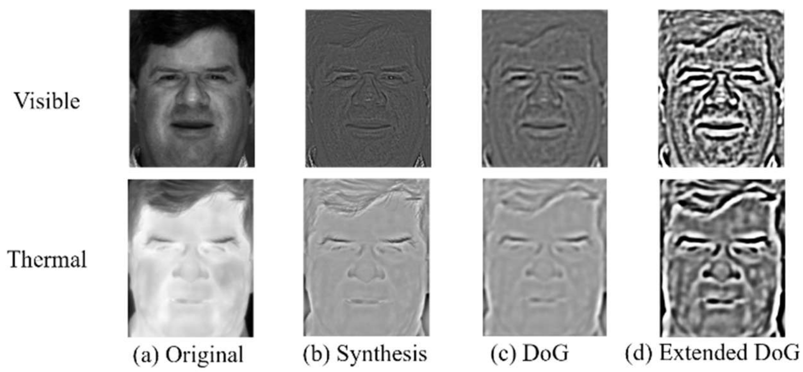

Figure 6. The extended DoG performance on visible spectrum and its corresponding thermal image.

To demonstrate the effectiveness of the proposed method, we compared the results with several state-of-the-art approaches. The quantitative results are summarized in Table 1. The performance is evaluated using the Rank-1 identification rates in five different scenarios: overall, Range 1 expressions, Range 1 baseline, Range 2 baseline, and Range 3 baseline. The overall accuracy is calculated from all thermal images of the 35 test subjects in the three ranges. Compared with partial least squares (PLS) [6], deep perceptual mapping technique (DPM) [23], and PLS-DPM [21], our proposed framework achieved a performance improvement of 42.09%, 19.83%, and 11.05%. In addition, we compared the performance of our method with that of the following five recent coupled deep learning methods: coupled neural network (CpNN) [24], PLS-CpNN [25], Generative Adversarial Network-based Visible Face Synthesis (GAN-VFS) [26], Coupled Deep Convolutional Neural Network (CpDCNN) [11], and Attribute-Guided Coupled Generative Adversarial Network (AGC-GAN) [27]. The results demonstrate that our method is also superior to the embedding deep learning techniques. The most significant advantages of our method are observed in Range 3 images, which is beneficial for practical applications where cameras are more than 5 meters away from subjects.

Table 1. Rank-1 Identification Rate for Cross-spectral Face Recognition on the ARL Dataset.

Scenario	PLS	DPM	CpNN	PLS-DPM	PLS-CpNN	CpDCNN	GAN-VFS	AGC-GAN	Coupled ICA
Overall	0.5305	0.7531	0.7872	0.8409	0.8452	0.8857	0.8561	0.8925	0.9514
Expressions	0.6276	0.7887	0.8213	0.8898	0.8907	0.9124	0.8934	0.9217	0.9619
Range 1	0.6211	0.8778	0.9102	0.9417	0.9388	0.9534	0.9412	0.9659	0.9650
Range 2	0.5197	0.7532	0.7904	0.8578	0.8586	0.8868	0.8701	0.9178	0.9643
Range 3	0.3448	0.5219	0.5588	0.5768	0.6014	0.7754	0.7559	0.8124	0.8878

The proposed method was also evaluated using the Tufts dataset. Compared to the previous dataset, this dataset is more challenging due to a large number of poses and expression variations. We compared the proposed method with the recent state-of-the-art methods that have used this dataset in their studies, including TV-GAN [16], CycleGAN [28], Pix2PixHD [29] and TR-GAN [30]. In order to perform a fair comparison, we adopted the same experiment parameters that 17 subjects were used for testing and the rest for training [30]. The training set contains 1294 thermal images and corresponding RGB images, while the test set contained 238 thermal and corresponding RGB images. The training set was used for generating the Coupled ICA model, and the testing set was used to build the multi-class SR model for evaluation. The thermal images in the training set were also used as cross-examples for training the multi-class SR classifiers. The results are summarized in Table 2. The proposed method yielded better performance than these deep learning models. In addition, deep learning approaches need much more data and computing resources for training. We used 1294 pairs of images for training in our method to match the same experiment data as the other methods. However, the Coupled ICA model could use much fewer training images to select the random patches. Overall, the recognition performance of the Tufts dataset is inferior to that of the ARL dataset, which may be attributed to the more diversified view angles and opaqueness of eyeglasses.

Table 2. Rank-1 Identification Rate for Cross-spectral Face Recognition on Tufts Dataset.

Scenario	TV-GAN	Pix2PixHD	CycleGAN	TR-GAN	Coupled ICA
Overall	0.5378	0.6174	0.7941	0.8865	0.8946

Face Verification Performance

We evaluate the face verification performance of proposed method and compare it with several recent works: GAN-VFS [7], Multi-AP-GAN [31], Multi-AP-GAN (GT) [31]. The cross-modal verification performance of different methods is evaluated using the Area Under the Curve (AUC) and Equal Error Rate (EER).

In face verification using the proposed method, the similarity score of a test sample and a class is calculated as the cosine similarity of the test sample and its reconstruction from the training samples of the class. Table 3 summarizes the AUC and EER of these state-of-the-art approaches and our proposed method. Face images in the Tufts dataset are more challenging due to a large number of pose and expression variations as well as a few numbers of images per variation. As a result, the performance of every method is slightly lower than what we observed in the ARL face dataset.

Table 3. Performance comparisons of the baseline methods and the proposed method.

Method	ARL Dataset		Tufts Dataset	
	AUC	EER	AUC	EER
GAN-VFS [18]	79.30%	27.34%	73.78% ± 0.46%	32.32% ± 0.53%
Multi-AP-GAN [45]	90.14% ± 2.17%	18.20% ± 2.65%	75.86% ± 0.88%	31.14% ± 0.74%
Multi-AP-GAN(GT) [45]	92.72% ± 2.03%	16.05% ± 2.15%	77.38% ± 0.98%	29.94% ± 0.79%
Coupled ICA	98.93% ± 0.03%	2.93% ± 0.52%	95.45% ± 1.85%	7.69% ± 2.57%

Ablation Study

In order to demonstrate the effectiveness of different modules in the proposed framework, we conduct the following ablation study using the ARL multi-modal face dataset.

(1) Thermal to visible recognition with HoG and multi-class SR classifier (Conventional-HOG).
(2) Thermal to visible recognition with Extend DoG, HOG and multi-class SR classifier (Conventional-EDoG-HOG).
(3) Thermal to visible recognition with Coupled ICA, HOG and multi-class SR classifier (CoupledICA-HOG).
(4) Thermal to visible recognition with Coupled ICA, Extend DoG, HOG and multi-class SR classifier (CoupledICA-EDoG-HOG).

Table 4 shows the Rank-1 identification rate for cross-spectral face recognition corresponding to each experimental module. This demonstrates the contribution of each component in our cross-spectral face recognition framework. The extended DoG improves the performance for Range 2 and 3 that have lower resolution but decreases the accuracy for the expression group. On the other hand, the Coupled ICA improves the accuracy in all the cases of different combinations of resolutions, expressions and feature descriptors, particularly Range 3. This study confirms that the Couped ICA improves spatial resolution and the overall performance of thermal-to-visible face recognition.

Table 4. The ablation study of the coupled ICA.

Scenario	Conventional-HOG	Conventional-EDoG-HOG	CoupledICA-HOG	CoupledICA-EDoG-HOG
Overall	0.8973	0.9062	0.9443	0.9514
Expressions	0.9345	0.8995	0.9594	0.9619
Range 1	0.9239	0.9435	0.9606	0.9650
Range 2	0.8986	0.9596	0.9603	0.9643
Range 3	0.7581	0.8352	0.8666	0.8878

Discussion

In addition to the Coupled ICA model, the SR classifiers and the cross-examples also contribute to the success of the framework. In addition, we investigate the potential application of the method in polarimetric thermal to visible face recognition. Finally, we conclude that face recognition research enhances the quality of education and student learning outcomes of the computer science program at Fort Valley State University.

Effect of Classifiers

In order to assess the effect of classifiers in the proposed Coupled ICA method, we compared sparse representation (SR), support vector machine (SVM) and partial least squares (PLS) in the multi-class framework. The same synthesized images and cross-examples are fed to the classifiers. Table 5 indicates that the combination of SR and the Coupled ICA provides the best recognition result on both the ARL and the Tufts datasets.

Table 5. Effect of SR, SVM and PLS classifiers.

Dataset	**SVM**	**PLS**	**SR**
ARL	0.0976	0.7431	0.9443
Tufts	0.0978	0.8890	0.9514

Effect of Cross-examples

Previous studies have demonstrated that cross-examples improve the performance of heterogeneous face detection [32] and recognition [16]. In this study, we investigate the performance of the Coupled ICA and SR model with and without thermal cross-examples. The overall accuracies of the two datasets in the experiments are listed in Table 6. The experiment results demonstrate that cross-examples improve the identification performance overall. The accuracy is improved by 7% for the ARL dataset and around 10% for the Tufts dataset. These results confirm that combining the Coupled ICA model with SR classification using cross-examples achieves the best performance for thermal-to-visible face recognition.

Table 6. Effect of cross-examples on Coupled ICA accuracy.

Scenario	**ARL Dataset**		**Tufts Dataset**	
	Without Cross	**With Cross**	**Without Cross**	**With Cross**
Overall	0.8856	0.9514	0.7976	0.9009

Application to Polarimetric Thermal Data

We also evaluate the proposed method on the polarimetric thermal modality of the ARL multi-modal face dataset. An interesting observation from our experiments is that the proposed method produced better results for conventional thermal images than polarimetric thermal images. Figure 7 shows the results for the conventional and polarimetric thermal-to-visible face recognition, which demonstrates the superiority of the thermal modality over the polarimetric modality. This is different from what has been discovered in the previous studies [4-7,16-18,31-32] that the thermal-to-visible face recognition benefits from additional geometrical and textural information in polarimetric images. That may be because the coupled ICA acquires the 2D point spread function information in the basis functions and cannot capture the 3D geometric information in the polarimetric images.

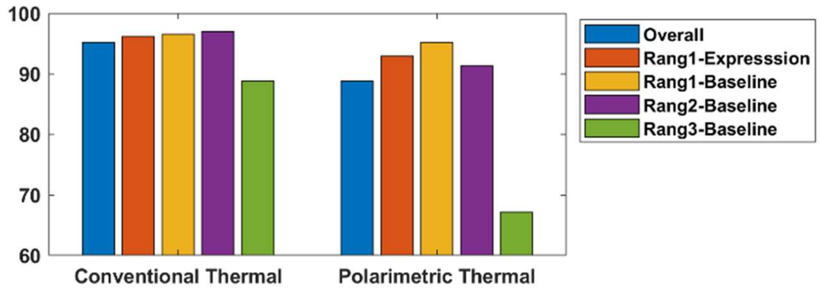

Figure 7. Results of conventional and polarimetric thermal-to-visible face recognition.

Educational Impact of Face Recognition Research

The cross-modal face recognition research created an educational component that computer science students conducted undergraduate research projects in artificial intelligence and machine learning using deep learning techniques. Their research findings have been presented at national and regional conferences and garnered positive feedback from the academic community [32-34]. The active engagement of undergraduate students in research offers significant benefits for both the students and the institution. Through these experiences, students develop strong critical thinking and problem-solving skills while gaining hands-on experience with research methods that prepare them for graduate study. At the same time, the computer science program has experienced a 35% increase in enrollment over the past five years, reflecting growing student interest and the positive impact of research-informed learning opportunities.

Acknowledgment

We thank Dr. Shuowen Hu and Dr. Matthew Thielke for providing access to the polarimetric face image dataset we used in our experiment.

References

[1] C. N. Fondje, S. Hu, N. J. Short and B. S. Riggan, "Cross-Domain Identification for Thermal-to-Visible Face Recognition," in *Proceedings of 2020 IEEE International Joint Conference on Biometrics (IJCB)*, Houston, TX, USA, Sep.2020, pp. 1-9.

[2] H. Zhang, B. S. Riggan, S. Hu, and N. J. Short, V. M. Patel, “Synthesis of high-quality visible faces from polarimetric thermal faces using generative adversarial networks,” *International Journal of Computer Vision*, vol. 127, pp. 845-862, 2019.

[3] D. Poster, S. Hu, N. Nasrabadi, and B. S. Riggan, “An examination of deep-learning based landmark detection methods on thermal face imagery,” in *Proceedings of 2019 IEEE/CVF*

Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, June 2019, pp. 980-987.

[4] M. S. Sarfraz and R. Stiefelhagen, "Deep perceptual mapping for cross-modal face recognition," *International Journal of Computer Vision*, vol. 122, pp. 426-438, Jul. 2016.

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

[6] S. Hu, J. Choi, A. L. Chan, and W. R. Schwartz, "Thermal-to-visible face recognition using partial least squares," *Journal of the Optical Society of America A*., vol. 32, no. 3, pp. 431-442, 2015.

[7] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu, "Generative adversarial network-based synthesis of visible faces from polarimetric thermal faces," in *Proceedings of 2017 IEEE International Joint Conference on Biometrics (IJCB),* Denver, CO, USA, Oct. 2017, pp. 100-107.

[8] X. Di, H. Zhang, and V. M. Patel, "Polarimetric thermal to visible face verification via attribute preserved synthesis," in *Proceedings of IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Redondo Beach, CA, USA, Oct. 2018, pp. 1-10.

[9] T. Zhang, A. Wiliem, S. Yang, and B. Lovell, "TV-GAN: generative adversarial network based thermal to visible face recognition," in *Proceedings of 2018 International Conference on Biometrics (ICB)*. Gold Coast, QLD, Australia, Feb. 2018, pp. 174-181.

[10] C. Reale, N. M. Nasrabadi, and R. Chellappa, "Coupled dictionaries for thermal to visible face recognition," in *Proceedings of 2014 IEEE International Conference on Image Processing (ICIP)*, Paris, France, Oct. 2014, pp. 328-332.

[11] S. M. Iranmanesh, A. Dabouei, H. Kazemi, and N. M. Nasrabadi, "Deep cross polarimetric thermal-to-visible face recognition," in *Proceedings of 2018 International Conference on Biometrics (ICB)*. Gold Coast, QLD, Australia, Feb. 2018, pp. 166-173.

[12] B. A Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Network*, vol. 7, no. 2, pp. 333-339, 1996.

[13] A. Bell and T. Sejnowski, "The independent components of natural scenes are edge filters," Vision Research, vol. 37, no. 23, pp. 3327-3338, Dec. 1997.

[14] R. Okamoto, S. Bando, and A. Nozawa, "Blind signal processing of facial thermal images based on independent component analysis," *IEEJ Transactions on Electronics, Information and Systems*, vol. 136, no. 8, pp. 1142-1148, Aug. 2016.

[15] K. Mallat, N. Damer, F. Boutros, A. Kuijper, and J. Dugelay, "Cross-spectrum thermal to visible face recognition based on cascaded image synthesis," in *Proceedings of 2019 International Conference on Biometrics (ICB)*, Crete, Greece, June 2019, PP. 1-8.

[16] B. S. Riggan, N. J. Short, and S. Hu, "Thermal to visible synthesis of face images using multiple regions," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, Mar. 2018, pp. 30-38.

[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1125-1134.

[18] Z. Huang, J. Wang, X. Wang, X. Song and M. Chen, "A study of sparse representation-based classification for biometric verification based on both handcrafted and deep learning features," *Complex & Intelligent Systems*, Sep. 2022, pp. 1583-1603.

[19] W. J. Shin, S. H. Lee, H. Min, H. Sohn, and Y. M. Ro, "Face verification using color sparse representation," in *Proceedings of 12th International Workshop on Digital-forensics and Watermarking*, Shanghai, China, Oct. 2013, pp. 290-299.

[20] Y. Chen, N. M. Nasrabadi, T. D. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*. vol. 8, no. 4, pp. 676-680, July 2011.

[21] S. Hu, N. J. Short, B. S. Riggan, C. Gordon, K. P. Gurton, M. Thielke, P. Gurram, A. L. Chan, "A polarimetric thermal database for face recognition research," in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA, June 2016, pp. 187-194.

[22] K. Panetta K. et al., "A comprehensive database for benchmarking imaging systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 509-520, Mar. 2020.

[23] M. S. Sarfraz and R. Stiefelhagen, "Deep perceptual mapping for cross-modal face recognition," *International Journal of Computer Vision*, vol. 122, pp. 426-438, Jul. 2016.

[24] B. S. Riggan, C. Reale, N. M. Nasrabadi, "Coupled auto-associative neural networks for heterogeneous face recognition," *IEEE Access*. vol. 3, pp. 1620-1632, Sep. 2015.

[25] B. S. Riggan, N. J. Short, and S. Hu, "Optimal feature learning and discriminative framework for polarimetric thermal to visible face recognition," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, Mar. 2016, pp. 1-7..

[26] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu, "Generative adversarial network-based synthesis of visible faces from polarimetric thermal faces," in *Proceedings of 2017 IEEE International Joint Conference on Biometrics (IJCB)*, Denver, CO, USA, Oct. 2017, pp. 100-107.

[27] S. M. Iranmanesh and N. M. Nasrabadi, "Attribute-guided deep polarimetric thermal-to-visible face recognition," in *Proceedings of 2019 International Conference on Biometrics (ICB)*. Crete, Greece, June 2019, pp. 1-8..

[28] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of IEEE international conference on computer vision (ICCV)*, Venice, Italy, Oct.2017, pp. 2223-2232.

[29] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA, Jun. 2018, pp. 8798-8807.

[30] L. Kezebou, V. Oludare, K. Panetta, and S. Agaian, "TR-GAN: thermal to RGB face synthesis with generative adversarial network for cross-modal face recognition," in *Proceedings of SPIE Defense + Commercial Sensing, Mobile Multimedia/Image Processing, Security, and Applications*. vol. 11399, Apr. 2020.

[31] X. Di, B. S. Riggan, S. Hu, N. J. Short, and V. M. Patel, "Multi-scale thermal to visible face

verification via attribute guided synthesis," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 2, pp. 266-280Apr. 2021.

[32] H. Tran, C. Dong, M. Naghedolfeizi, and X.-Y. Zeng, "Using cross-examples in Viola-Jones algorithm for thermal face detection," in *Proceedings of the ACM Southeast Conference*, April 2021.

[33] R. Steven, "Facial recognition through deep learning with Colab", presented at *ACM Mid-Southeast Conference*, Gatlinberg, USA, November 2025.

[34] L. Taylor-Wilson, "Assessing the Performance of Deep Learning Models for Skin Lesion Segmentation in Computer-Aided Diagnosis," presented at *Emerging Researchers National Conference in STEM*, Atlanta, USA, April 2025.