Gender and Ethnicity Classification using Deep Learning in Heterogeneous Face Recognition

Neeru Narang and Thirimachos Bourlai
West Virginia University
MILab, LCSEE, 395 Evansdale Drive, Morgantown, WV 26506-6070, U.S.A.
nneeru@mix.wvu.edu, Thirimachos.Bourlai@mail.wvu.edu.

Abstract

Although automated classification of soft biometric traits in terms of gender, ethnicity and age is a well-studied problem with a history of more than three decades, it is still far from being considered a solved problem for the case of difficult exposure conditions, such as during night-time, in environments with unconstrained lighting, or at large distances from the camera. In this paper, we investigate the advantages and limitations of the automated classification of soft biometric traits in terms of gender and ethnicity in Near-Infrared (NIR) long-range, night-time face images. The impact of soft biometric traits in terms of gender and ethnicity is explored for the purpose of improving cross-spectral face recognition (FR) performance. The main contributions are, (i) a dual database collected in NIR band at night time and at four different distances of 30, 60, 90 and 120 meters is used, (ii) a deep convolution neural network to perform the classification in terms of gender and ethnicity is proposed, (iii) a set of experiments is performed indicating that, the usage of soft biometric traits to perform face matching, resulted in a significantly improved rank-1 identification rate when compared to the original biometric system (scenario dependent).

1. Introduction

Soft biometric traits are physical, behavioral and human characteristics such as gender, ethnicity, height, weight and skin color [9]. Some soft biometric traits such as height, weight and age, change over time [8]. However, traits such as gender and ethnicity are permanent and stable. These traits have been regularly used by the biometrics research community in various applications [10, 11]. Park et al. [11], used soft biometric information with in an existing biometric system to improve the overall performance of face recognition. The authors demonstrated that, soft traits can provide more valuable information in cases where a face image is occluded or captured at challenging viewing angles.

Predicting these soft biometric traits can be done reasonably well by human observers. However, when having to deal with large datasets this process needs to be performed fast, automatically and efficiently. The capabilities of existing biometric systems, when performing the gender and ethnicity classification in controlled conditions, e.g. indoors, outdoors, during day time, at short ranges etc., can result in operationally acceptable classification rates. However, automated classification, when working at night time environments, under un-controlled conditions and when the subject's face is captured at long standoff distances, is a very challenging task, especially when, at the same time, the face images are captured using different spectral band imaging sensors.

There are heterogeneous FR systems designed to automatically classify the gender and ethnicity class [4], for a database collected under controlled environments, at short standoff distances. The authors proposed the gender and ethnicity classification algorithm based on encoding gradient information on Gabor-transformed images. To perform the estimation of the gender class, authors used AR database, is composed of visible band face images, including frontal view face, with variable facial expressions and illumination conditions collected at the Computer Vision Center (CVC) in U.A.B. To perform the estimation of ethnicity class, they selected the Morph and CAS-PEAL databases. The Morph database contains metadata in the form of the gender, ethnicity, age, weight and height. Chen et al. [3], proposed an automated method for the gender classification. They selected a face database collected in controlled conditions in the thermal band and CBSR database collected in NIR band in a short range distance $(0.5m - 1.2m)$. To perform the classification, features were extracted based on LBP, PCA methods and further these features were used for the classification using a SVM, an LDA-based and a random forest classifier. Lagree et al. [6], proposed a method to predict the gender and ethnicity class from the features of iris texture. Based on the results, the

978-1-5090-1869-7/16/$31.00 ©2016 IEEE

Figure 1. Database: Raw Images (Top) and Normalized Images (Bottom).

authors reported that the prediction of gender is more difficult than the prediction of ethnicity.

Mery et al. [10], reported that the usage of soft biometric traits can increase the performance of biometric systems. The authors proposed a new approach called adaptive sparse representation of random patches. The proposed patch based approach is used to recognize facial attributes such as gender, race, beard, and disguise. Finally, the authors made a comparison with other available methods (SVM-RBF, Adaboost) and reported that their system outperformed Adaboost and SVM-RBF based methods. They used AR, UND and FRGC 2.0 databases, that cover different band, including the visible and thermal bands. Levi et al. [7], proposed a deep learning based method for the classification of gender and age for the wild database collected in the visible band. The authors reported the classification results for the age and gender using the Adience benchmark, which consists of images collected from smart-phone devices. Based on the experimental results, they concluded that Convolutional Neural Network can be used for the age and gender classification.

This work is an effort to solve a more complicated problem as we are dealing with a database captured at night time environments and variable long standoff distances, starting from 30 meter and up to 120 meters, in 30 meters intervals. We propose a deep learning based, scenario-dependent, and band-adaptable (it is working well for both visible and NIR face images) algorithmic approach for the classification of soft biometric traits in terms of gender and ethnicity. A set of face recognition experiments are performed, indicating that the usage of soft biometric traits achieved a significantly better performance (i.e. 45% improvement for the Caucasian class and 26% for the Asian class for the Cross-spectral scenario at 60m distance) than the performance achieved by the primary FR system.

2. Methodology

In this section, we outline the challenging database selected to perform the experiments, a deep learning based method for the classification of soft biometric traits in terms of gender and ethnicity and, finally, a set of experiments performed in order to find the significant impact of usage of soft biometric traits in face recognition.

2.1. Database

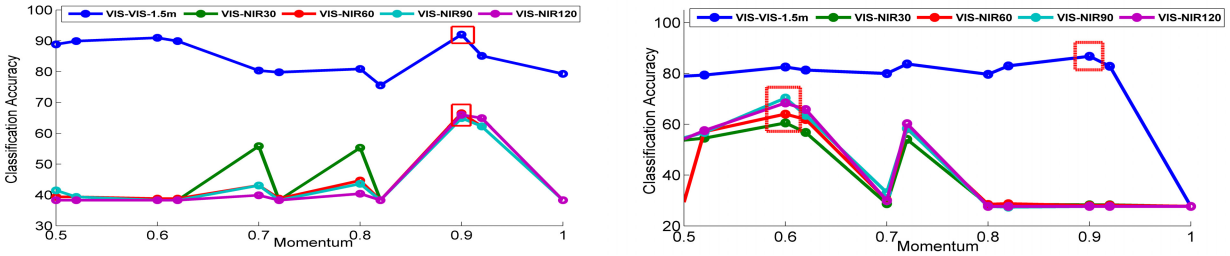
Long Distance WVU Database: The database was collected at WVU and it was used to perform the classification and face recognition experiments. NIR images were collected outdoors in night time at long standoff distances of 30, 60, 90 and 120 meters, spanning over a time period of 20 days, as shown in Fig. 1. The visible (VIS) dataset, contains subjects mug shots taken indoors and at a distance of 1.5 meter. In total, the database consists of 103 subjects (70 males and 30 females).

Long Distance Heterogeneous Face Database (LDHF): We used the database collected from Kang et al. [5], with 100 subjects to extend the training data to perform the classification experiments. It contains both VIS and NIR face images captured at distances of 60, 100, and 150 meters outdoors and at a 1 meter distance indoors.

2.2. Normalization of Data

The background is not uniform and objects such as vehicles, trees may affect the classification and recognition experiments. To deal with this problem, normalization of face images is performed. Image registration compensates for the slight perturbation in the frontal pose. It is composed of two main steps, eye detection and affine transformation [2]. The eye centers are, first, located by manual annotation and are used to geometrically normalize the images. Based on the located eye coordinates, the canonical faces are automatically constructed by applying an affine transformation. Faces are first aligned by placing the coordinates of the eyes in the same row, such that the slope between the right and left eye is zero degrees. Finally, all faces are canonicalized to the same dimension of 128×128 pixels (see Fig. 1).

Training Data: VIS Images, Testing Data: VIS and NIR Images



(a) Ethnicity Class (b) Gender Class

Figure 2. Selection of Momentum for the Ethnicity and Gender Class. For intra-spectral classification, visible images were used for training and testing. For cross-spectral classification, visible images (VIS 1.5m) were used for training (to simulate the most challenging scenario of having a gallery set composed of only visible face images) and NIR images collected at a distance of 30, 60, 90 and 120 meters distance were used for the testing.

2.3. Deep Learning for Automatic Classification

To perform the classification, we selected the visual geometry group (VGG) CNN architecture [13]. It consists of a number of convolutional layers (a bank of linear filters), followed by max pooling layers and a rectification layer, such as the rectified linear unit (ReLU) along with fully connected layers [13, 7, 12]. The pooling layer is applied to perform the down-sampling operation via computing the maximum of local region [13]. The last fully connected layer is softmax that computes the scores for each class. To train our system, we selected the batch size of 100, learning rate of 0.002 and empirically optimized the classification model in terms of the values of epoch and momentum.

- Model Architecture: In the first convolution layer, 20 filters of size 5×5 are applied to the input, followed by a max pooling layer taking the maximal value of 2×2 with two-pixel strides. The output of the previous layer is processed by a second convolution layer, with 50 filters of size 5×5, followed by a max pooling layer with the same stride value. The second layer is further processed by a third convolution layer with 50 filters of size 4×4. This layer is followed by an activation function, and ReLU [12]. Finally, a third layer is processed by the last convolution layer, with 26 filters of size 2×2. The output of this layer is fed to a soft-max layer that assign a label to each class, i.e. it will assign a label male or female in terms of gender and Asian or Caucasian in terms of ethnicity class.

-Training and Testing: In the experiments performed, the subjects in the training and test sets are different, and the images are taken at different locations and days. In this work, for the ethnicity class, two scenarios are selected while for the gender class three scenarios are selected.

Ethnicity Class: *Scenario 1*, where there is an overlap between the subjects for training and testing, which are performed using our WVU database (VIS-VIS, VIS-NIR). The VIS face images at short distance of 1m are selected to train the CNN network and the NIR face images, collected at long distances, for the testing. To perform the classification experiments, 50% of the data (subjects) is selected to train the system and the rest for testing. This process is repeated for 4 times, where for each set the different training and testing data is selected. For *Scenario 2*, we selected the LDHF database [5] and extended the training data set by using also the WVU database. For testing, the WVU database collected in NIR band at long standoff distances is selected. To perform the classification experiments, the data from the LDHF database and 50% from WVU is selected to train the system. The rest of the data is selected for testing (no overlap of subjects).

Gender Class: *Scenario 1*, where the LDHF face images in NIR band at a distance of 1m are used for training and the WVU database (NIR dataset; 30 up to 120 meters) for testing. *Scenario 2*, the VIS face images from the LDHF database are selected for training and the NIR dataset from the WVU database for testing. *Scenario 3*, the VIS face images at a short distance from both LDHF and WVU databases are used for training, while the WVU database at long standoff distances for testing. There is no overlap of subjects for the training and testing. All the database from the LDHF (70 males and 30 females) and 20% of the data (subjects) from the WVU is selected to train the system. The rest of the 80% of the data (subjects) from the WVU is selected for testing. This process is repeated twice, where for each set the different training and testing data is selected. In all these scenarios, training and test sets consist of images from different people, sensors, light conditions and locations.

- Optimization: We conducted an empirical optimization on epoch and momentum parameters that resulted in better ethnicity and gender based classification accuracy.

Selection of Epoch: We performed a series of experiments with the selected values of $4, 8, ..., 52$. Based on the results, the best classification results are achieved for epoch value 16 for ethnicity class. The same set of experiments is

performed for the gender class and an epoch value of 16 is selected.

Selection of Momentum: Based on [13], the momentum values lie in the range from 0.50 to 1. To select the best value, a series of experiments are performed with values of 0.5, 0.52, ..., 1. We ran the experiments for both the ethnicity and gender classes. As shown in Fig. 2, for the ethnicity class the highest classification accuracy is achieved when the momentum value is 0.90. For the gender class, the highest classification accuracy is achieved when the momentum value is 0.60 for NIR (30 up to 120 meters) and 0.90 for the VIS band in the testing set as shown in Fig. 2.

2.4. Face Matching

In this work, both the commercial and academic face matchers is used to perform the face recognition experiments. For the commercial software, we used the COTS face matcher (L1 systems). For the academic software, we used the publicly available academic face identification evaluation system from Colorado State University (CSU) [1]. CSU face evaluation system includes, *Principal Component Analysis (PCA)*, a combined *Principal Component Analysis (PCA) and Linear Discriminant Analysis (PCA+LDA)* and *Bayesian Intra-personal/Extra-personal (BIC)* classifier using either *Maximum Likelihood (ML)* or *Maximum Posteriori (MAP)* hypothesis [1].

3. Experimental Results

3.1. Ethnicity and Gender based Classification

In the first set of experiments, our experiments aim to illustrate how the deep learning system performs for intra-spectral band, where both the training and testing data is selected from the same band (VIS-VIS) under controlled conditions. Finally, in order to determine the extent of which the performance of the classification system is affected when the standoff distance increases, we performed cross-spectral experiments, where we selected the face images from visible band (VIS 1.5) for the training and NIR face images from a distance of 30 up to 120 meters away for testing. The epoch and momentum values are selected based on our results of the empirical optimization study discussed in Section 2.3, for both the ethnicity and gender classes. To further improve the classification accuracy results, we combined the classification results from two models with different momentum values.

Ethnicity Class: For the ethnicity class, from a set of values selected in range from 0.50 to 1, the best classification results were achieved when the momentum value is 0.90 (for 1.5 up to 120 meters) as shown in Fig. 2. For model 1, classification is performed where we trained the network when the momentum value is 0.90 and epoch value is 16. The classification accuracy is more than 85% for the Caucasian group. For model 2 (with a different set of learning parameters), the classification accuracy is more than 75% for the Asian group. Finally, when we combined the model 1 and model 2, a significant improvement in the performance results is achieved. The overall classification accuracy is almost 95% as presented in Table 1 for cross-spectral classification (VIS-NIR 30m).

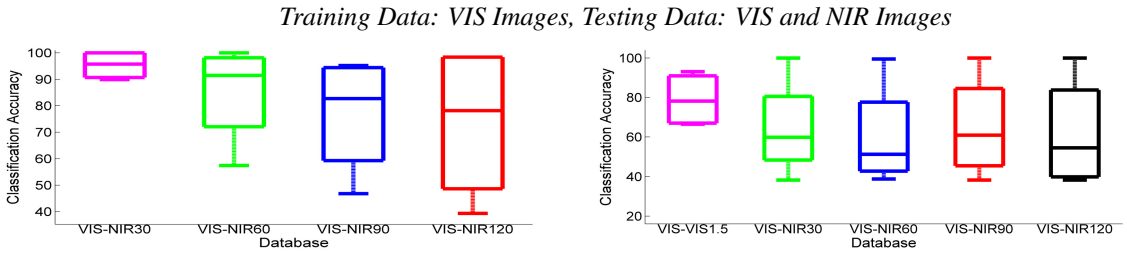
In order to examine the effectiveness of classification system (trained on parameters selected from optimization), each experiment is repeated 4 different times, where each time a different training set was randomly selected and rest of the data was used for testing (Scenario 1 and Scenario 2). To see the performance variation in the classification results with increasing distance for the testing set, we plotted the boxplots as shown in Fig. 3 (each boxplot is based on results from 4 randomly selected training and testing sets). Based on mean and variance plots, the better classification results achieved for testing data at distance of 30m for Scenario 1.

Table 1. Summary of classification results for the ethnicity and gender class based on CNN for each dataset. To perform CNN, the model is trained for the challenging testing database.

	Classification Accuracy		
Datasets	**Ethnicity Class**		**Gender Class**
Train-Test	**Scenario 1**	**Scenario 2**	**Scenario 3**
VIS-VIS 1.5m	99.04	78.98	96.41
VIS-NIR 30m	95.34	64.49	86.14
VIS-NIR 60m	85.10	60.23	89.45
VIS-NIR 90m	76.86	65.02	93.52
VIS-NIR 120m	73.53	61.83	94.12

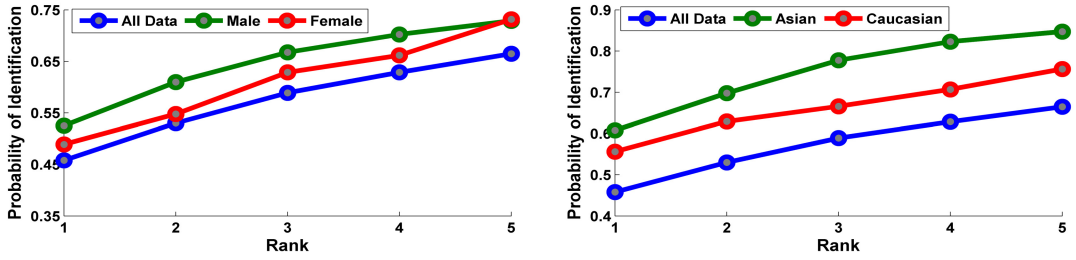
Table 1, depicts the accuracy results for the ethnicity and gender classification from deep learning system. Based on the results, we determined that (trained model on WVU database) for the ethnicity class for scenario 1, the classification accuracy is 95% when the VIS dataset was used for training while the NIR 30m dataset for testing. The system achieves promising classification performance results, when the NIR face images at a long standoff distance of 60, 90 and 120 meters are selected for testing. The decrease in the performance of the system for scenario 2 with the extended database, is due to the variation in light conditions and sensors, under which the training images were collected.

Gender Class: For the gender class, for Scenario 1 (NIR images for training) and Scenario 2 (VIS images for training), the LDHF database is used for the training and the WVU database for testing (see in Section 2.3). For Scenarios 2, the highest classification accuracy achieved was 78% as presented in Table 2. However, the results were not satisfactory (32% accuracy) when the NIR images collected from long standoff distance were used. Scenario 3, the VIS



(a) Scenario 1 (b) Scenario 2

Figure 3. Classification results for the ethnicity class. Scenario 1 (Left), is based on overlapping of subjects between training and testing. Scenario 2 (Right), is based on no overlapping of subjects between training and testing.



(a) Gender Class vs. All Data (b) Ethnicity Class vs. All Data

Figure 4. *Cross-distance matching scenarios for NIR 30m (Gallery) against NIR 120m (Probe)*: Performance results when using the academic CSU face recognition evaluation tool. Each algorithm was run several times and the mean rank scores are presented, i.e. Gender class with Male and Female Group vs. all data without any grouping (Left). Ethnicity class with Asian and Caucasian Group against all data without any grouping (Right).

band face images from both the LDHF and WVU databases are used for training, while the WVU database for testing. The model parameters are selected, where the highest classification accuracy is achieved (see Section 2.3). For model 1, classification is performed where we trained the network when the momentum value is 0.60 for the NIR class and 0.90 for the VIS class. The classification accuracy is more than 85% for the female group. For model 2 (with different set of learning parameters), the classification accuracy is more than 90% for the male group. Finally, when we combined the model 1 and model 2, a significant improvement in the performance results is achieved. The overall classification accuracy is more than 95% as presented in Table 1 for intra-spectral classification (VIS-VIS 1.5m).

Based on the classification results for the training and testing sets, without any overlap of subjects (Table 1), we determined that better classification results are achieved for the gender in comparison to the ethnicity class (Scenario 2). The most probable reason is the availability of more data (WVU database+LDHF) compared to ethnicity class, where the LDHF database consists of primarily Asian population data.

3.2. Cross-scenarios Face Matching Results

Two sets of face recognition experiments are performed using the academic and commercial face matchers. First, experiments are performed with the original FR system, namely when no gender or ethnicity classification is used. Second, we tried to determine whether the usage soft biometric traits in terms of gender and ethnicity can enhance the recognition performance.

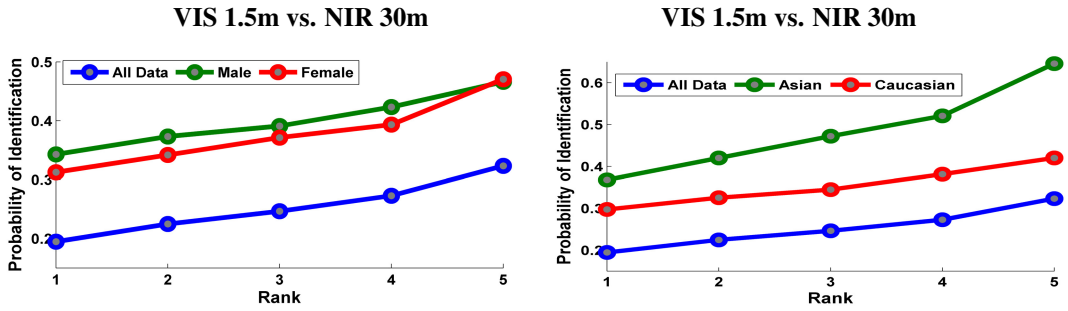
Table 3 provides an overview of the number of datasets used, as well as the cross-scenarios and face recognition algorithms selected. In this work, we selected the datasets,

Table 2. Summary of classification results for the gender class based on CNN for each dataset. To perform CNN, model is trained for the challenging testing database (the learning rate is represented as L and momentum as M).

Scenario 1: Train NIR (LDHF Database) vs. Test (Own Database)		
Parameters	**M=0.90, L=0.002**	**M=0.92, L=0.002**
VIS 1.5m	32.04	31.92
NIR 30m	32.52	32.40
NIR 60m	32.40	32.40
NIR 90m	30.83	30.95
NIR 120m	32.52	32.16
Scenario 2: Train VIS (LDHF Database) vs. Test (Own Database)		
Parameters	**M=0.90, L=0.002**	**M=0.92, L=0.002**
VIS 1.5m	77.06	78.28
NIR 30m	32.65	32.65
NIR 60m	32.77	32.52
NIR 90m	32.04	32.04
NIR 120m	32.04	32.04

Table 4. In this table we compared the *intra-spectral, cross-distance* matching scenarios for with and without grouping of data in terms of ethnicity and gender class: experimental results when running all CSU FR algorithms when using 50% of the NIR data for training and the rest of the data for testing. The experiments were run 4 times and the rank-1 scores presented here are the means.

FR Algorithm Gallery NIR 30m vs.	**Cross-Distance Rank-1 Scores**								
	NIR 60m			**NIR 90m**			**NIR 120m**		
	All Data	**Male**	**Female**	**All Data**	**Male**	**Female**	**All Data**	**Male**	**Female**
Bayesian MAP	0.85	0.92	0.78	0.52	0.69	0.53	0.36	0.40	0.38
Bayesian ML	0.86	0.93	0.82	0.60	0.74	0.56	0.38	0.47	0.42
LDA Euclidean	**0.88**	0.93	0.74	0.61	0.69	0.47	0. 38	0.45	0.38
LDA lda_Soft	**0.88**	0.93	0.75	0.64	0.69	0.46	0.40	0.47	0.39
PCA Euclidean	0.68	0.79	0.56	0.27	0.39	0.21	0.20	0.28	0.17
PCA MahaCosine	0.86	**0.94**	0.84	**0.68**	**0.79**	0.61	**0.46**	**0.52**	0.41
	All Data	**Asian**	**Caucasian**	**All Data**	**Asian**	**Caucasian**	**All Data**	**Asian**	**Caucasian**
Bayesian MAP	0.85	0.87	0.86	0.52	0.62	0.44	0.36	0.44	0.29
Bayesian ML	0.86	0.85	0.88	0.60	**0.70**	0.60	0.38	0.50	0.38
LDA Euclidean	**0.88**	0.80	**0.90**	0.61	0.56	0.58	0. 38	0.44	0.44
LDA lda_Soft	**0.88**	0.83	0.89	0.64	0.53	0.57	0.40	0.45	0.43
PCA Euclidean	0.68	0.67	0.65	0.27	0.36	0.22	0.20	0.22	0.19
PCA MahaCosine	0.88	0.87	0.89	**0.68**	0.69	0.68	**0.46**	**0.61**	**0.56**



(a) Gender Class vs. All Data (b) Ethnicity Class vs. All Data

Figure 6. *Cross-Spectral matching scenarios for VIS 1.5m (Gallery) against NIR 30m (Probe)*: Performance of academic CSU face recognition system. Each algorithm was run several times and the rank scores presented here are means. Gender class with Male and Female Group vs. all data without any grouping (Left). Ethnicity class with Asian and Caucasian Group against all data without any grouping (Right).

Table 3. Summary of total number of FR experiments are performed, with and without grouping of data in terms of ethnicity and gender class. To perform the FR experiments, two FR cross-scenarios are selected cross-distance (CD) and cross-spectral (CS).

Datasets	**Train/Test Datasets**	**# Scenarios**	**CSU, COTS**	**# Total Experiments**
All	4	3 CD + 4 CS	6 + 1	196
Caucasian	4	3 CD + 4 CS	6 + 1	196
Asian	4	3 CD + 4 CS	6 + 1	196
Male	4	3 CD + 4 CS	6 + 1	196
Female	4	3 CD + 4 CS	6 + 1	196

with and without the usage of soft biometric traits. Without grouping, all data is used to perform the face matching experiments. With grouping the ethnicity class (Caucasian and Asian) and gender class (male and female) is used. The classified database with labels; male, female, Asian and Caucasian from our developed deep learning system are used for the data with grouping. For each dataset, we randomly divided the 50% data as the training set and rest of the data is used as the testing set, with no subject overlap.

This process is repeated several times, using random selection of the training and test sets each time. Two cross-scenarios are selected: cross-distance (CD) and cross-spectral (CS). For CD, NIR vs. NIR face matching is performed for 3 sets: *30 vs. 60, 30 vs. 90 and 30 vs. 120 meters*. For CS, VIS vs. NIR face matching is performed for 4 sets: *1.5 vs. 30, 1.5 vs. 60, 1.5 vs. 90 and 1.5 vs. 120 meters*. Cross-scenarios are investigated using academic and commercial (non-training based) face matching schemes. In table 3, the total number of face matching experiments performed for each dataset are described.

Cross-Distance: In this experiment, we compared the baseline images in NIR band (30m) to NIR images captured at 60, 90 and 120 meters respectively. We compared the results with and without the usage of soft biometric traits. The identification results using both the academic and commer-

Table 5. In this table we compared the *cross-spectral, cross-distance* matching scenarios for with and without grouping of data in terms of ethnicity and gender class: experimental results when running all CSU FR algorithms when using 50% of the NIR data for training and the rest data for testing. The experiments were run 4 times and the rank-1 scores presented here are the means. COTS was also tested.

FR Algorithm Gallery VIS 1.5m vs.	**Cross-Spectral rank-1 scores for Ethnicity**											
	NIR 30m			**NIR 60m**			**NIR 90m**			**NIR 120m**		
	All Data	**Asian**	**Caucasian**	**All Data**	**Asian**	**Caucasian**	**All Data**	**Asian**	**Caucasian**	**All Data**	**Asian**	**Caucasian**
Bayesian MAP	0.19	0.34	0.29	0.15	0.31	0.19	**0.13**	0.23	0.09	**0.15**	0.26	0.14
Bayesian ML	0.18	0.37	0.30	0.16	0.31	0.23	0.12	0.20	0.14	0.13	0.24	0.12
LDA Euclidean	0.19	0.36	0.26	0.20	0.39	0.20	**0.13**	0.30	0.11	0.11	0.25	0.16
LDA lda_Soft	0.19	0.33	0.26	**0.22**	0.38	0.20	**0.13**	**0.32**	0.10	0.12	**0.28**	0.16
PCA Euclidean	0.08	0.12	0.15	0.08	0.13	0.11	0.06	0.09	0.10	0.08	0.10	0.09
PCA MahaCosine	0.12	0.24	0.23	0.10	0.27	0.19	0.09	0.27	0.14	0.08	0.23	0.13
COTS	**0.48**	**0.70**	0.47	0.03	**0.48**	**0.67**	0.02	0.04	0.06	0.01	0.05	0.05

cial face matchers are summarized in Table 4. Based on the results, we determined that for 60m, the rank-1 score is improved from 88% to 94% - male group. For 90m, the identification performance is improved from 68% to 79% and 46% to 52% for 120 meters - male group. For the female group, the identification results are similar for all the distances (for probe images at 60 up to 120 meters distance).

For grouping of the data in terms of ethnicity, the identification performance is improved from 46% to 61% for the Asian and 56% for the Caucasian groups for 120m distance data. The results are similar for 60 and 90 meters data, with and without the usage of soft biometric traits in terms of ethnicity. The CMC curves for the first 5 ranks for the gender and ethnicity class, for the best performed algorithm out of set of algorithms included in the CSU academic matcher, are represented in Fig. 4. The rank scores in CMC curves are the mean of scores from a set of experiments where each time a different training set was randomly selected.

Cross-Spectral: We matched the visible face images (1.5m) to NIR (30, 60, .., 120 meters) face images. In Table 5, we see the results with and without the usage of soft biometric traits in terms of ethnicity, using both the academic and commercial matchers. In order to examine the effectiveness of these matchers, each experiment is repeated four times, where training set is randomly selected and rest of the data is used for testing. In Table 5, the mean rank-1 scores are represented for probe images captured from 30 and up to a 120 meter distance.

Based on the results, we determined that at 30m distance, the rank-1 score accuracy is improved from 48% to 70% for the Asian group. The results are similar for the Caucasian group when using the COTS matcher. For the 60m data, the identification performance is improved from 3% to 48% for the Asian group. For the 90m data (i.e. VIS - NIR 90m), the identification performance is improved from 13% to 32% for the Asian group while similar results are achieved for the Caucasian group, when using academic face matchers. For the 120m data, the identification is improved from 15% to 28% for the Asian group when using academic matchers. For all distances, better performance results are achieved for the Asian group in comparison to the original FR system, namely when no gender or ethnicity classification is used.

The same set of experiments is repeated for the gender class. Based on the results obtained, we determined that for the 30m distance, the COTS matcher performs better than the academic ones. The identification performance is improved from 48% to 52% for the male and from 48% to 65% for the female group. For 60, 90 and 120 meters distances, better performance results are achieved from the academic matchers in comparison to the commercial one. For the 60m data, the identification performance is improved from 22% to 33% for the male and from 22% to 25% for the female group. For the 90m data, the identification performance is improved from 23% to 26% for the male group. Finally, for the 120m data, the identification performance is improved from 15% to 17% for the male group.

The CMC curves for the first 5 ranks for the gender and ethnicity classes of algorithms included in the CSU academic matcher, are represented in Fig. 6. The rank scores in CMC curves are the mean of scores from a set of experiments where each time a different training set was randomly selected.

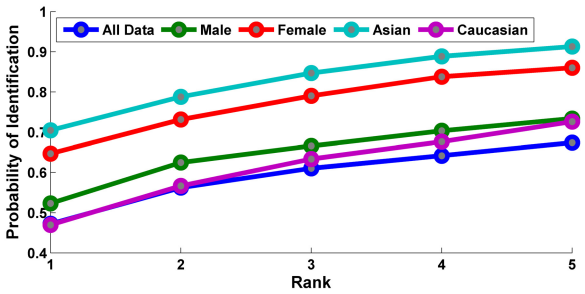


Figure 5. Cross-Spectral matching scenarios for VIS (Gallery) against NIR 30m (Probe): Performance of COTS.

Commercial matcher is used on all cross-scenarios. We obtained good results only for the cross-spectral scenario (i.e. visible 1.5m gallery against the NIR 30m probe images), where images are of good quality. The CMC curves for the first 5 ranks, for the gender and ethnicity class, are illustrated in Fig. 5.

4. Conclusions and Future Work

We investigated the advantages and limitations of the gender and ethnicity classification in heterogeneous environments, i.e. when the face images are captured at night time and at variable standoff distances. To perform the experiments, we used two different databases, i.e. WVU and the LDHF database. We proposed a deep convolutional neural network based architecture to classify the visible and multi-distance NIR face images into Asian or Caucasian as well as Male or Female groups. To perform the classification using our deep learning architecture, an empirical optimization of parameters was performed, including the epoch, momentum and learning rate, when training the model. In the experiments we performed, we trained the model when using a multi-band database, where the training data is selected from the visible band dataset (controlled conditions at a short standoff distance) and the testing data is selected from the NIR band multi-distance face images (30 up to 120 meters).

Different face matching scenarios are tested (VIS 1.5 vs. VIS 1.5, NIR 30, NIR 60, NIR 90 and NIR 120 meters). Based on an extensive set of experiments, the proposed CNN architecture provided us with significant classification results for the selected challenging databases. The experimental results show that the gender classification accuracy is better when compared to the ethnicity classification accuracy. This can be explained by the fact that, the LDHF database is used to extend the training data, which consists of both the male and female groups. The challenge was that, this database consists of primarily Asian population face images while working to solve the ethnicity-based classification problem. For ethnicity-based classification, the highest classification accuracy was achieved when a short distance data is selected for testing. Also, the classification performance of the system reduces as a function of the distance, especially at ranges greater than 60m for this class.

We performed a series of face identification experiments for cross-distance and cross-spectral scenarios. Our results provide important evidence that data grouping in terms of gender and ethnicity (Asian and Caucasian) provide significant improvement in the rank-1 identification rate for both cross-distance and cross-spectral scenarios.

Based on the experimental results we conclude that: First, a CNN can be used to classify the data in terms of ethnicity and gender class when using both constrained and unconstrained face datasets. Second, based on the face identification results, we conclude that the usage of soft biometric traits in terms of ethnicity and gender can improve the rank-1 identification rate of our FR system, i.e. 45% improvement for the Caucasian and 26% for the Asian class for the Cross-spectral scenario at 60m distance. In the future, we expect to further improve the classifications results. Thus, we plan to include more databases to train our deep learning model as well as to test alternative CNN architectures.

Acknowledgments: This material is based upon work supported by the Center for Identification Technology Research and the National Science Foundation under Grant No. 1066197.

References

[1] D. S. Bolme, J. R. Beveridge, M. L. Teixeira, and B. A. Draper. The CSU face identification evaluation system: Its purpose, features and structure. In *Proc. International Conference on Vision Systems*, pages 304 – 311, Graz, Austria, April 2003.

[2] T. Bourlai, J. V. Dollen, N. Mavridis, and C. Kolanko. Evaluating the efficiency of a nighttime, middle-range infrared sensor for applications in human detection and recognition. In *Proc. SPIE Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIII*, pages 1–12, Bellingham, WA, USA, April 2012.

[3] C. Chen and A. Ross. Evaluation of gender classification methods on thermal and near-infrared face images. In *IEEE International Joint Conference on Biometrics, IJCB*, pages 1–8, Washington, DC, USA, October 2011.

[4] C. Chen and A. Ross. Local gradient Gabor pattern (LGGP) with applications in face recognition, cross-spectral matching, and soft biometrics. In *Proc. SPIE Biometric and Surveillance Technology for Human and Activity Identification X*, pages 87120R–87120R, Baltimore, Maryland, USA, 2013.

[5] D. Kang, H. Han, A. K. Jain, and S.-W. Lee. Nighttime face recognition at large standoff: Cross-distance and cross-spectral matching. *Pattern Recognition*, 47(12):3750–3766, 2014.

[6] S. Lagree and K. W. Bowyer. Predicting ethnicity and gender from iris texture. In *Technologies for Homeland Security (HST), 2011 IEEE International Conference on*, pages 440–445, Nov 2011.

[7] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*, June 2015.

[8] J. R. Lyle, P. E. Miller, S. J. Pundlik, and D. L. Woodard. Soft biometric classification using periocular region features. In *Biometrics: Theory Applications and Systems (BTAS), Fourth IEEE International Conference on*, pages 1–7, Sept 2010.

[9] J. Mansanet, A. Albiol, and R. Paredes. Local deep neural networks for gender recognition. *Pattern Recognition Letters*, 70:80–86, 2016.

[10] D. Mery and K. Bowyer. Automatic facial attribute analysis via adaptive sparse representation of random patches. *Pattern Recognition Letters*, 2015.

[11] U. Park and A. K. Jain. Face matching and retrieval using soft biometrics. *Information Forensics and Security, IEEE Transactions on*, 5(3):406–415, 2010.

[12] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. *Proceedings of the British Machine Vision*, 1(3):6, 2015.

[13] A. Vedaldi and K. Lenc. MatConvNet – Convolutional Neural Networks for MATLAB. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pages 689–692, New York, NY, USA, 2015.