Evaluation of Gender Classification Methods on Thermal and Near-infrared Face Images

Cunjian Chen and Arun Ross

Lane Department of Computer Science and Electrical Engineering

West Virginia University, Morgantown WV 26506, USA

cchen10@mix.wvu.edu, arun.ross@mail.wvu.edu

Abstract

Automatic gender classification based on face images is receiving increased attention in the biometrics community. Most gender classification systems have been evaluated only on face images captured in the visible spectrum. In this work, the possibility of deducing gender from face images obtained in the near-infrared (NIR) and thermal (THM) spectra is established. It is observed that the use of local binary pattern histogram (LBPH) features along with discriminative classifiers results in reasonable gender classification accuracy in both the NIR and THM spectra. Further, the performance of human subjects in classifying thermal face images is studied. Experiments suggest that machine-learning methods are better suited than humans for gender classification from face images in the thermal spectrum.

1. Introduction

A variety of attributes such as gender, age and ethnicity, besides an individual's identity, can be deduced from face images [31]. Predicting gender from a face image has been actively studied in the computer vision and biometrics literature [13, 22]. The problem of gender classification from face images can be posed as a two-class problem in which the input face image is analyzed and assigned to one of two classes: male or female. Automatic gender classification has applications in surveillance, human-computer-interaction and image retrieval systems [5]. In the context of biometrics, gender is viewed as a soft biometric trait that can be used to index databases or enhance the recognition accuracy of primary traits such as face [15].

With advancements in sensing technologies, thermal and near-infrared images are beginning to be used in face-related applications. For example, face recognition in near-infrared (NIR) [17, 32] and thermal (THM) [26] spectra has been motivated by the need to determine human identity in

nighttime environments [7]. Furthermore, changes in ambient illumination have lesser impact on face images acquired in these spectra than the visible spectrum. Current gender classification systems discussed in the literature have been designed for and evaluated on face images acquired in the visible spectrum. Little attention has been given to automatic gender classification from faces in the thermal or near-infrared spectrum (Figure 1). In fact, only one publication in the vast biometric literature has dealt with the problem of gender prediction using near-infrared images [25].
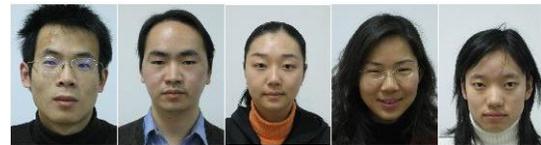
(a) Visible images [18]

(b) Thermal images

(c) Near-infrared images [17]

Figure 1. Samples of face images in different spectra. The left two columns depict male subjects while the right three columns depict female subjects. The identities in (a), (b), (c) are different. The problem of gender estimation in (b) and (c) becomes challenging when the face images are cropped and the surrounding context information, such as hair, is lost. See Figures 2 and 4.

Makinen et al. [21] evaluated gender classification methods in the visible spectrum with automatically detected and

aligned faces. Three different feature sets based on pixel intensities of low-resolution images (24×24, 36×36 and 48×48), Local Binary Patterns (LBP) and Haar-like filters were used, along with Neural Network, Support Vector Machine (SVM) and Adaboost classifiers. They found that the SVM classifier operating on 36×36 pixels face images gave the best gender classification accuracy. Since there are visual differences between face images in the visible spectrum and those acquired in the NIR or THM spectrum, it is essential to investigate and determine the most effective features for gender classification in these spectra.

In this work, we provide a systematical evaluation of different gender classification methods on both NIR and THM images. To the best of our knowledge, this is the first work that evaluates the possibility of assessing gender from thermal face images.

The paper is organized as follows. Previous literature in gender classification is briefly reviewed in Section 2. The various feature extraction and classification methods used in this work for gender assessment are presented in Section 3. Section 4 presents the results of gender classification on both thermal and near-infrared face images . A discussion of the results is provided in Section 5, and conclusions are drawn in Section 6.

2. Relevant Work

The study of automatic gender classification from face images dates back to the early 1990s and has gained increased interest recently due to renewed focus on attributes-based face recognition [16]. Most techniques for gender classification approach the problem from the perspective of machine learning, as it is essentially a two-class classification problem.

Golomb et al. [11] trained a back-propagation neural network (BPNN) to identify gender from human face images at a resolution of 30×30 pixels. An average classification rate of 91.9% on 90 exemplars was obtained compared to a human performance of 88.4%.

Later on, Moghaddam et al. [22] utilized SVM for gender classification from low-resolution thumbnail face images (21×12). The average five-fold cross-validation classification rate on 1,755 FERET face images was 96.62% when the Gaussian Radial Basis Function (RBF) kernel was used by the SVM. Their work also pointed out that the SVM classification of low-resolution face images was very effective, compared to other classifiers.

Baluja et al. [3] used the Adaboost classifier to determine the gender of a person from a low-resolution face image. The proposed system was extremely fast and yet the performance was comparable to the SVM-based classifier. They reported an accuracy of over 93% on a dataset of 2,409 FERET face images each of resolution 20×20 pixels.

Although the direct use of pixels from low-resolution face images in [11, 22, 3] resulted in very good gender classification performance, this represents a rather simple feature set that may not be robust enough in complex scenarios involving pose and illumination changes. Therefore, other types of feature sets have also been proposed. Yang et al. [31] used the Local Binary Pattern Histogram (LBPH) features and the Adaboost classifier to learn the best local features for gender classification. Experiments were performed to predict the age, gender and ethnicity of face images. A gender classification rate of 93.3% was obtained using five-fold cross validation on 8,433 snapshot face images.

In principle, a gender classification method can be divided into two components: (a) a feature extractor that extracts features from the face, and (b) a feature classifier that assigns the extracted features into one of two classes - male or female. Feature extraction methods are based on direct use of pixels from low resolution face images [22, 3], Principle Component Analysis (PCA) [2, 12], Linear Discriminant Analysis (LDA) [5, 14] and LBP [31, 10, 29]. Most gender classifiers are based on Neural Network [11], Adaboost [3, 31], and SVM [22]. A systematic overview of methods for gender classification from face images in the visible spectrum can be found in [21]. The choice of feature extraction and classification methods depends on whether the faces are aligned manually or automatically [21], and whether the train and test sets are from the same database or different databases [5]. The goal of this work is to study the following two questions:

1. Can a learning-based scheme be used to distinguish males from females in near-infrared and thermal face images? Visually, the task appears to be challenging. But can a machine be trained to make this distinction?

2. Are there specific feature extraction and classification techniques that result in better gender classification accuracy than others in these spectra?

The first question has implications in the psychology of gender recognition from faces. We have included experiments in which four human subjects were asked to distinguish between males and females in thermal images. These experimental results are also reported in this paper.

3. Methods

We selected three types of features for gender representation: LBP, PCA and pixels from low-resolution face images. These features have been tested before in the gender classification literature and have proved to be effective in the visible spectrum. The main discriminant classifiers investigated in this work include SVM, LDA, Adaboost, Random Forest, Gaussian Mixture Model (GMM) and Multi Layer Perceptron (MLP).

The LBP texture descriptor [27, 23] has been successfully used in many face-related applications. The features are calculated based on pixel intensity values within a small neighborhood. Ahonen et al. [1] first extended the use of LBP to face recognition and demonstrated that this local feature descriptor is very efficient in face representation. After that, numerous techniques based on LBP features have been adopted for gender representation [31, 10].

Earlier work on employing SVM for LBP-based gender classification often used linear or RBF kernels [21, 20]. However, from the results reported in [4] and in subsequent works, it appears that histogram intersection is an effective kernel representation scheme making it a good candidate for building gender classification systems. It has also been shown that histogram intersection has the required mathematical properties for it to be a suitable kernel function for SVMs [4] compared to the Gaussian RBF kernel. Besides, it requires fewer parameters to tune in order to achieve the optimal performance. Finally, it is much faster to compute than the Gaussian RBF kernel.

3.1. Feature Extraction

The LBP operator was first introduced and described as a texture descriptor that computes micro-patterns in an image by thresholding 3×3 neighborhoods based on the value of the center pixel, and then converting the resulting binary pattern into a decimal value. Later, it was extended to include neighborhoods of different sizes to account for textures at multiple scales [1]. LBP encodes the sign difference between the center pixel and neighborhoods, and therefore it is invariant to monotonic illumination changes.

The local neighborhood is defined as a set of sampling points evenly spaced on a circle (or rectangle). The LBP operator is described as $LBP_{P,R}^{u^2}$, where P refers to the number of sampling points placed on a circle with radius R. The symbol u^2 represents the uniform pattern which, in our case, refers to those binary patterns that have at most two bitwise transitions from 0 to 1 or 1 to 0. For instance, 10011111 is a uniform binary pattern while 10100111 is not. Uniformity is an important concept as it characterizes micro-features (structural information) such as lines, edges and corners in the image. Although only 58 out of the 256 8-bit binary patterns are reckoned to be uniform, nearly 90% of all observed local image neighbourhoods are uniform [27, 23]. We chose to use $LBP_{8,1}^{u^2}$ in all our experiments based on empirical evidence. The binary pattern for pixels lying in a circle (f_p, $p = 0, 1, \ldots, P-1$) with the center pixel f_c, is calculated as follows:

$$
S(f_p - f_c) = \begin{cases} 1 & \text{if } f_p - f_c \geq 0; \\ 0 & \text{if } f_p - f_c < 0. \end{cases} \tag{1}
$$

Then a binomial weight 2^p is assigned to each $S(f_p - f_c)$ to compute the LBP code:

$$
LBP_{P,R} = \sum_{p=0}^{P-1} S(f_p - f_c) 2^p. \tag{2}
$$

The original image is first divided into non-overlapping blocks, and the LBP histogram is computed for each block. After deriving the histogram sequence for each block, the final global representation is obtained by concatenating individual sequences [1]. This is not the only way to extract local binary pattern histogram (LBPH) features from an image. It is possible to densely sample the blocks and apply the Adaboost algorithm to select LBPH features [31].

The LBPH features derived from an image can be directly used to perform gender classification. For an image size of 126×90, which is divided into 18×15 pixels per block, the total number of blocks is 42. Each block can produce 59 bin features for a specific uniform pattern (u^2), leading to a $42 \times 59 = 2478$ dimensional feature. These parameters are fixed for all LBP-based experiments. Such a high dimensional feature vector can be effectively reduced by applying PCA or other dimension reduction schemes.

Consider a set of LBPH features derived from an image database, $\{(x_i, t_i)\}_{i=1}^{N}$, where x_i is the LBPH feature derived from the ith facial image and t_i is the associated gender class label. Here, $t_i \in \{-1, 1\}$, where -1 (+1) indicates a female (male). The PCA is performed on the covariance matrix of vectorized images:

$$
\Sigma_g = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T, \tag{3}
$$

where x_i is the derived LBPH feature vector and $\bar{x}$ is the mean vector of the training set features. The eigenvectors can be obtained through the decomposition, $\Sigma_g \Phi_g = \Phi_g \Lambda_g$, where Φ_g are the eigenvectors of Σ_g and Λ_g are the corresponding eigenvalues. The gender features are extracted by projecting the sample vector onto the subspace expanded by the eigenvectors: $s_i = \Phi_g^T (x_i - \bar{x})$. Here, s_i is the final feature vector that encodes the gender information of sample x_i.

3.2. Feature Classification

Once the LBPH feature vector x_i is reduced to a lower, but more compact dimensional vector s_i, different classifiers can be used to predict the gender label.

SVM: SVM [9] is a machine learning technique used for pattern classification and regression analysis. It is based on the concept of searching a linear boundary between two classes of patterns as follows:

$$
y(s) = w^T \phi(s) + b, \tag{4}
$$

where $\phi(s)$ denotes the transformation of the original feature-space into a higher dimensional space and b is the

bias. The training set comprises of a set of N training samples $\{s_1, \cdots, s_N\}$, with corresponding label values $\{t_1, \cdots, t_N\}$ where $t_i \in \{-1, 1\}$. The incoming new data point s is classified based on the sign of $y(s)$. We assume that the training dataset is linearly separable in the transformed feature space, which indicates that there would be a linear boundary defined by parameters $\{w, b\}$ satisfying the condition:

$$y(s_i) = \begin{cases} 1 & \text{if } t_i = +1; \\ -1 & \text{if } t_i = -1. \end{cases} \quad (5)$$

The final criteria is to make sure that $t_i \cdot y(s_i) > 1$ for all the training points ($i = 1, \ldots, N$). If there exists no hyperplane that can partition the male and female examples, a soft margin method is used to solve the following optimization problem [9]:

$$\min_{w, \varepsilon} \left\{ \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} \varepsilon_i \right\}, \quad (6)$$

subject to the constraint:

$$y_i(w^T \cdot \phi(s_i) + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0. \quad (7)$$

Here, ε_i is a variable introduced to control the trade off between a large margin and a small error penalty. C is a constant. SVM seeks a linear separating hyperplane with the maximal margin in the higher dimensional space $\phi(s_i)$. The Gaussian RBF kernel is defined as: $k(x_i, y_i) = exp(-\gamma ||x_i - y_i||^2)$. The optimum values for C and the kernel parameter γ are often obtained by a grid-search of the parameter space. In some cases, where the features are derived from a histogram, as in the case of LBP, the histogram intersection kernel might be more effective:

$$k(x_i, y_i) = \sum_{i=1}^{n} min(x_i, y_i), \quad (8)$$

where x_i and y_i are the ith histogram bin of the feature vectors x and y. The kernel is related to the transform ϕ by the equation $k(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$.

LDA: In the work of [5], the authors argued that the use of linear classification techniques is preferred in the context of limited computational resources. LDA classifier tries to maximize the separation of male and female classes based on the Fisher's criterion:

$$J(w) = \frac{w^T S_B w}{w^T S_W w}, \quad (9)$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix. $J(w)$ is the objective function that we are trying to maximize with respect to w. The maximum value (projection matrix) is obtained by solving the generalized eigenvalue problem of $S_W^{-1} S_B$.

Adaboost: Adaboost [3] is a method for constructing a strong classifier as a cascaded linear combination of simple weak classifiers. In the cascaded arrangement, the subsequent classifiers are tweaked in favor of those training patterns that are misclassified by previous classifiers. The Adaboost classifier can be denoted as:

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x), \quad (10)$$

where $h_t(x)$ refers to the weak classifier operating on the input feature set x and T is the number of weak classifiers. The sign of $f(x)$ is the output of the final strong classifier. α_t is the corresponding weight for each weak classifier. Adaboost has been observed to result in good classification even if the base weak classifiers perform poorly. We apply the Adaboost classifier to both the LBP and low-resolution feature sets. It is implemented based on decision stump weaklearners (decision tree with a single node). The default number of weak classifiers used is 100.

Random Forest: Random Forest (RF) [8] is an ensemble classifier that consists of many decision trees. Each decision tree is trained independently and successively based on a boot-strapped sampling of the training dataset. The individual learners are combined through bootstrap aggregation [8]. Given an input feature vector, it successively moves through the individual trees in the forest. The final classification (prediction) is based on a majority voting scheme over all the trees. Recent work [30] on the task of gender classification from infants to seniors has also shown the superiority of using Random Forest for feature selection.

3.3. Other Gender Classifiers

Apart from the aforementioned classifiers, other types of classifiers such as GMM [19], MLP [21] and Bayesian [28] have been used in the gender classification literature work. In the GMM or Bayesian classifier, the probability density function for each class (i.e., male and female) is modeled as a multivariate Gaussian distribution:

$$p(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\}. \quad (11)$$

Here, D is the dimension of the feature vector that encodes gender information and μ is the D-dimensional mean vector. Σ is the $D \times D$ covariance matrix and its determinant is denoted by $|\Sigma|$. The classification is done by maximizing the posterior function: $p(C_i|x_t) = p(x_t|C_i \cdot p(C_i)), i \in \{1, 2\}$.

A MLP neutral network is composed of an input layer, a hidden layer and an output layer. It utilizes a supervised learning technique (backpropagation) to train the network. The number of input nodes is equal to the dimension of the feature vector. We select 20 hidden nodes, and the number

of training cycles is set to be 40. The output value is based on the classification threshold, for instance, 0.5. An output value above the threshold is classified as male and a value below is classified as female. Some of the classifiers used in this work are adapted from [24].

4. Experiments

In order to establish the possibility of assessing gender from images obtained in non-visible spectra, we use a thermal face database and a near-infrared face database. The eye coordinates of the face images are manually located. Sometimes, a mask was necessary to exclude any useless background information.

4.1. Thermal Dataset

The thermal database contains one thermal face image each of 1003 subjects. In addition, this database contains two visible-light (VIS) images for each of these subjects. The image size is 480×640 pixels. The size of each image after alignment and cropping is 130×150 pixels. For the LBP methods, the image is resized to 126×90. There are 229 female subjects and 774 male subjects. The subjects have variations in age and ethnicity. Most of the samples are captured in the near-frontal pose (Figure 2).

Figure 2. Samples images from the thermal database. Top row shows male subjects and the bottom row shows female subjects.

To test gender classification algorithms on thermal images, the first 100 male and 100 female subjects are selected for training and the remaining are used for testing. This ensures that there is no overlapping of subjects between the training and test sets. Thus, there are 674 male subjects and 129 female subjects in the test set. The male (female) classification rate is defined as the percentage of males (females) that are correctly recognized within the male (female) group. These classification rates are used to determine if there is any bias of individual groups towards overall classification performance due to imbalanced test data sets. The ratio of males and females in the test set is close to 5:1.

As shown in Table 1, the use of histogram intersection (HI) kernel in SVM results in better performance than the linear kernel (LI) in terms of the female classification rate. The LBP+SVM (HI) method achieves 84.50% accuracy for female classification, compared to 79.07% using the LBP+SVM (LI) method. The RBF kernel ($C = 8, \gamma = 0.002$) achieves better results than both HI and LI kernels, except for female classification. However, it is much slower during the training stage as it searches a large parameter space to seek the optimum values. The performance is further enhanced with the use of PCA to derive a more compact feature descriptor, compared to LBP+SVM (LI). The reduced dimension feature vector is 60 in this experiment based on singular value decomposition (SVD). Among all tested classifiers, SVM (LBP+PCA+SVM) results in the most balanced accuracy in terms of overall male and female classification rates. Here, the linear SVM kernel is preferred, since the derived feature is no longer directly obtained from histogram bin features.

In order to show that LBPH descriptor is much more effective than PCA or low-resolution features extracted from thermal images, we perform experiments on the same dataset with the latter set of features (Figure 3). The best overall performance of 81.32% is achieved using the PCA+LDA methods. The individual performances for male and female groups are 80.71% and 84.50%, respectively.
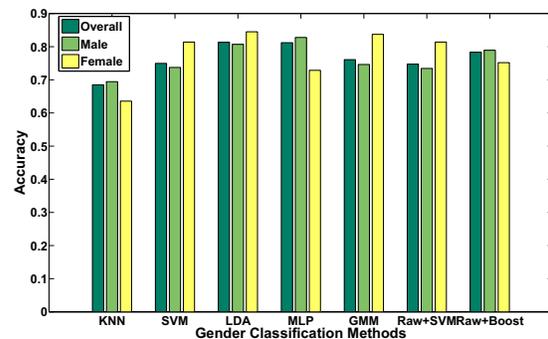

Figure 3. Performance evaluation of different gender classifiers based on PCA and low resolution features on the thermal dataset (First five methods use the PCA features).

Since the thermal database includes visible-light (RGB) images for each subject, gender classification is conducted on the visible spectra as well. It must be noted that the pairs of visible and thermal images are not co-registered. From the results in Table 2, the LBP+PCA+SVM method has the best overall performance of 90.66%. The RBF kernel ($C = 2, \gamma = 0.002$) is slightly better than the HI kernel. Meanwhile, the HI kernel is much better than the LI kernel. Compared to the LBP+SVM (LI) method, the enhanced feature reduction and classification methods such as

Table 1. Gender classification accuracy on thermal images.

Algorithm	Overall	Male	Female
LBP+SVM (HI)	0.8792	0.8858	0.8450
LBP+SVM (LI)	0.8705	0.8858	0.7907
LBP+SVM (RBF)	0.9041	0.9184	0.8295
LBP+RandomForest	0.8655	0.8665	0.8605
LBP+PCA+SVM	0.9016	0.9110	0.8527
LBP+PCA+LDA	0.8667	0.8650	0.8760
LBP+PCA+MLP	0.8804	0.8828	0.8682
LBP+PCA+GMM	0.8742	0.8783	0.8527

Table 2. Gender classification accuracy on visible images in the thermal database.

Algorithm	Overall	Male	Female
LBP+SVM (HI)	0.8842	0.8739	0.9380
LBP+SVM (LI)	0.8493	0.8398	0.8992
LBP+SVM (RBF)	0.8941	0.8828	0.9535
LBP+RandomForest	0.8667	0.8546	0.9302
LBP+PCA+SVM	0.9066	0.9036	0.9225
LBP+PCA+LDA	0.8804	0.8724	0.9225
LBP+PCA+MLP	0.8643	0.8501	0.9380
LBP+PCA+GMM	0.8269	0.8205	0.8605

LBP+PCA+SVM, LBP+PCA+LDA and LBP+PCA+MLP achieve better results. The average overall performance of all the seven algorithms on the thermal dataset is 88.03%, compared to 87.16% on the visible dataset. This suggests the feasibility of assessing gender from thermal images as well. This is the first work in the literature that establishes this possibility.

4.2. NIR Dataset

In this section, the gender classification methods are evaluated on NIR images. There is limited work conducted on gender classification exclusively in the NIR dataset. Unlike images in the thermal spectrum, facial features from NIR face images are much more visually perceptible.

The CBSR NIR Face Dataset [1] [17] contains 3,940 NIR face images of 197 persons. The original image size is 480×640 pixels. After alignment and normalization, the image size is 130×150. Each subject has 20 samples. We manually label the gender information and exclude those subjects that are not easily recognizable. In the end, 135 male subjects and 55 female subjects are used. Samples of one subject are shown in Figure 4(a) and the corresponding normalized versions are shown in Figure 4(b). The normalization approach is based on manually located eye coordinates and is described in the work of Bolme et al. [6].

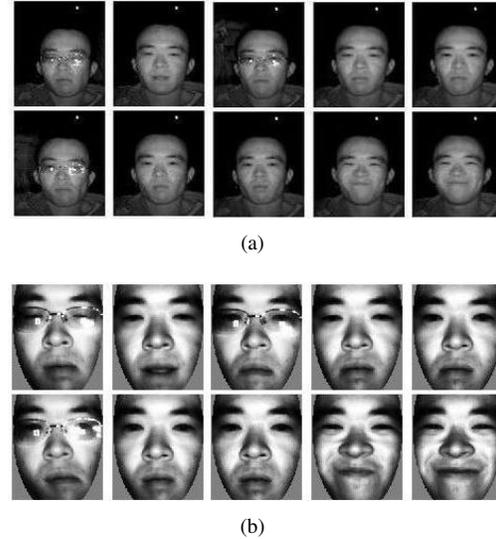[1]CBSR database: http://www.cse.ohio-state.edu/otcbvs-bench/

(a)

(b)

Figure 4. (a) Original NIR face samples of one subject from CBSR dataset; (b) Normalized NIR face samples.

To test gender classification algorithms in the NIR dataset, 15 male subjects and 15 female subjects are used for training. The total number of training samples is $15 \times 20 + 15 \times 20 = 600$. The remaining 120 male and 40 female subjects are used for testing, resulting in a total of 3200 images. The subjects in the training and test sets are mutually exclusive in this experiment also.
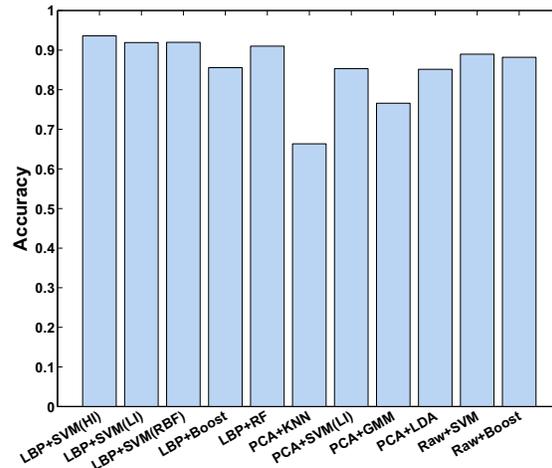

Figure 5. Comparison of different gender classification algorithms on CBSR NIR dataset based on overall accuracy.

From the results shown in Figure 5 and Table 3, the LBPH descriptor of facial features gives the best classification accuracy, with LBP+SVM (HI) resulting in 93.59%, LBP+SVM (LI) resulting in 91.91% and LBP+SVM (RBF, $C = 32, \gamma = 0.001$) resulting in 91.97%. The results indi-

Table 3. Gender classification accuracy on near-infrared images using different feature sets and classifiers.

Algorithm	Overall	Male	Female
LBP+SVM (HI)	0.9359	0.9417	0.9187
LBP+SVM (LI)	0.9191	0.9208	0.9137
LBP+SVM (RBF)	0.9197	0.9375	0.8662
LBP+Adaboost	0.8556	0.8458	0.8850
LBP+RandomForest	0.9100	0.9208	0.8775
PCA+KNN	0.6634	0.6358	0.7462
PCA+SVM (LI)	0.8531	0.8721	0.7963
PCA+GMM	0.7659	0.8013	0.6600
PCA+LDA	0.8516	0.8533	0.8462
RawPixels+SVM [22]	0.8897	0.9083	0.8337
RawPixels+Adaboost [3]	0.8819	0.9033	0.8175

cate that the HI kernel is superior to both the LI and RBF kernels in this case. The combination of LBP and Adaboost resulted in a classification rate of 84.58% for males and 88.5% for females. For the Random Forest classifier with LBP features (LBP+RandomForest), an overall accuracy of 91% is obtained. The individual accuracy of this method for male and female classes were 92.08% and 87.75%, respectively.

We also tested other type of facial features such as PCA and low-resolution image pixels (See Table 3). For the PCA method, SVD is used to reduce the feature dimensionality to 60. The image was also resized to 32×32 pixels. Among all the PCA-based methods, the SVM (PCA+SVM(LI)) and LDA (PCA+LDA) classifiers result in the best performance, although this is still lower than the LBP-based methods. The rest of the classifiers, such as GMM classifier (PCA+GMM) and KNN do not result in good performance.

The low-resolution feature representation has been observed to perform well in the visible spectrum [22, 3]. We apply the same methods used in the work of [22] and [3] on the NIR dataset. Each image was resized to 20×20. The number of weak classifiers used by the Adaboost classifier was 625 (Table 3). The SVM classifier used a Gaussian RBF kernel with $\gamma = 0.001$ and $C = 1$, based on a grid-search of the parameter space. Individual classification rates of 88.97% and 88.19% were obtained for SVM and Adaboost, respectively.

The above experiments show the effectiveness of employing LBP features for gender classification in NIR spectrum, compared to PCA or low resolution (RawPixels) features [22, 3].

5. Discussion

In Section 4.1, various gender classification methods have been evaluated on thermal face images. The visual

Table 4. Gender classification accuracy reported on thermal images based on human perception. Subject A and B are male observers and subject C and D are female observers (Four subjects in total). Note that the subjects were not very good at classifying female face images. The subjects possibly classified all ambiguous images as male.

Observers	Overall	Male	Female
Subject A	0.9141	0.9585	0.6822
Subject B	0.8966	0.9748	0.4884
Subject C	0.8917	0.9496	0.5891
Subject D	0.9078	0.9896	0.4806
Machine	0.9016	0.9110	0.8527

appearance of facial thermal images is remarkably different from that of visible-light and near-infrared face images. This poses a big challenge for humans to recognize gender information from thermal face images (cropped) if the entire upper-body image is not available (See Figure 2). However, machine learning approaches treat a face image as a bunch of pixels and automatically select the most relevant features from an image to perform gender classification.

In order to compare the performance of the proposed approach against that of humans, we asked several human subjects (observers) to perform gender prediction independently. Only the cropped version of thermal images were provided to human subjects and they were asked to assign one of two labels to each image: male or female. Human observers tend to classify males or females based on presence or absence of mustache, beard and eyelids, which are still observable in the thermal spectrum. Ambiguities arise when such distinctive features are not available in the face images. As seen in Table 4, more females are misclassified as males while fewer males are misclassified as females. This is expected since females are usually much more difficult to classify than males when important facial features are missing. On the other hand, the machine learning approach automatically selects relevant facial features from thermal images to make this distinction. This experiment demonstrates the advantage of using machine learning approaches for gender classification in complex scenarios involving non-traditional spectrum. The machine algorithm performance reported in Table 4 is based on the LBP+PCA+SVM method in Table 1.

6. Conclusion

In this paper, we evaluated different gender classification methods in both thermal and near-infrared face databases. Our work demonstrates the importance of SVM, Adaboost and LDA classifiers for gender recognition. It was shown that SVM for histogram-based gender classification results in much better performance than PCA features or low-

resolution pixel intensity values in both the thermal and near-infrared spectra. Further, machine-learning based gender classification achieves better results compared to that of humans in the thermal spectrum. This is the first work that evaluates various gender classifiers on near-infrared and thermal images.

Acknowledgements

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: application to face recognition. *IEEE Trans. on PAMI*, 28:2037–2041, 2006.

[2] K. Balci and V. Atalay. PCA for gender estimation: which eigenvectors contribute? In *ICPR*, pages 363–366, 2002.

[3] S. Baluja and H. A. Rowley. Boosting sex identification performance. *IJCV*, 71:111–119, 2007.

[4] A. Barla, F. Odone, and A. Verri. Histogram intersection kernel for image classification. In *ICIP*, pages 513–516, 2003.

[5] J. Bekios Calfa, J. Buenaposada, and L. Baumela. Revisiting linear discriminant techniques in gender recognition. *IEEE Trans. on PAMI*, 33:858–864, 2011.

[6] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper. The CSU face identification evaluation system: Its purpose, features, and structure. In *ICCVS*, pages 304–313, 2003.

[7] T. Bourlai, N. D. Kalka, D. Cao, B. Decann, Z. Jafri, F. Nicolo, C. Whitelam, J. Zuo, D. A. Adjeroh, B. Cukic, D. Jeremy, H. Larry, A. Ross, and N. Schmid. Ascertaining human identity in night environments. In *Distributed Video Sensor Networks*, pages 471–478, 2011.

[8] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang. Gender recognition from body. In *ACM Multimedia*, pages 725–728, 2008.

[9] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:1–27, 2011. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[10] G.Guo, C. Dyer, Y. Fu, and T. Huang. Is gender recognition affected by age? In *ICCV (Workshops)*, 2009.

[11] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *NIPS*, pages 572–577, 1990.

[12] A. B. A. Graf and F. A. Wichmann. Gender classification of human faces. In *Biologically Motivated Computer Vision*, pages 491–500, 2002.

[13] S. Gutta, H. Wechsler, and P. J. Phillips. Gender and ethnic classification of face images. In *International Conference on Face & Gesture Recognition*, pages 194 –199, 1998.

[14] A. Jain and J. Huang. Integrating independent components and linear discriminant analysis for gender classification. In *AFGR*, pages 159–163, 2004.

[15] A. K. Jain, S. C. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition? In *SPIE International Symposium on Defense and Security : Biometric Technology for Human Identification*, pages 561–572, 2004.

[16] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, pages 365–372, 2009.

[17] S. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. on PAMI*, 29:627–639, 2007.

[18] S. Z. Li, Z. Lei, and M. Ao. The HFB face database for heterogeneous face biometrics research. In *IEEE Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum*, pages 1–8, 2009.

[19] Z. Li, X. Zhou, and T. S. Huang. Spatial gaussian mixture model for gender recognition. In *ICIP*, pages 45–48, 2009.

[20] J. R. Lyle., P. E. Miller., S. J. Pundlik., and D. L. Woodard. Soft biometric classification using periocular region features. In *BTAS*, pages 1–7, 2010.

[21] E. Makinen and R. Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Trans. on PAMI*, 30:541–547, 2008.

[22] B. Moghaddam and M. Yang. Learning gender with support faces. *IEEE Trans. on PAMI*, 24:707–711, 2002.

[23] T. Ojala, M. Pietikinen, and T. Menp. Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on PAMI*, 24:971–987, 2002.

[24] Y. Rong. Matlabarsenal: A matlab package for classification algorithms. http://www.informedia.cs.cmu.edu/yanrong.

[25] A. Ross and C. Chen. Can gender be predicted from near-infrared face images? In *ICIAR*, pages 120–129, 2011.

[26] D. A. Socolinsky and A. Selinger. A comparative analysis of face recognition performance with visible and thermal infrared imagery. In *ICPR*, volume 4, pages 217–222, 2002.

[27] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. on Image Processing*, 19:1635–1650, 2010.

[28] M. Toews and T. Arbel. Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE Trans. on PAMI*, 31:1567–1581, 2009.

[29] R. Verschae, J. R. del solar, and M. Correa. Gender classification of faces using adaboost. In *Proceeding of Progress in Pattern Recognition, Image Analysis and Applications*, pages 68–78, 2006.

[30] Y. Wang, R. K., C. Chen, and Y. Chang. Gender classification from infants to seniors. In *BTAS*, pages 1–6, 2010.

[31] Z. Yang and H. Ai. Demographic classification with local binary patterns. In *ICB*, pages 464–473, 2007.

[32] B. Zhang, L. Zhang, D. Zhang, and L. Shen. Directional binary code with application to PolyU near-infrared face database. *Pattern Recogn. Lett.*, 31:2337–2344, 2010.