Gender Recognition Using Nonsubsampled Contourlet Transform and WLD Descriptor

Muhammad Hussain[1], Sarah Al-Otaibi[1], Ghulam Muhammad[1], Hatim Aboalsamh[1], George Bebis[2], and Anwar M. Mirza[1]

[1] College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
[2] Department of Computer Science and Engineering, University of Nevada at Reno

Abstract. Gender recognition using facial images plays an important role in biometric technology. Multiscale texture descriptors perform better in gender recognition because they encode the multiscale facial microstructures in a better way. We present a gender recognition system that uses SVM, two-stage feature selection and multiscale texture feature based on Nonsubsampled Contourlet Transform and Weber law descriptor (NSCT-WLD). The proposed system has better recognition rate (99.50%) than the state-of-the-art methods on FERET database. This research also reveals that in NSCT decomposition what is essential for face recognition and what is important for other tasks like age detection.

Keywords: Gender recognition, Face recognition, WLD Descriptor, Nonsubsampled Contourlet Transform, Support Vector Machines.

1 Introduction

In category specific face recognition, facial images are first categorized based on visual cues like gender and race etc. and then face recognition is performed using category specific feature descriptors. This approach improves [17] the performance but the bottleneck is categorization. In this paper, we address the problem of face categorization based on gender i.e. gender recognition problem. Gender recognition is important from other aspects as well; it can enhance the performance of a wide range of applications such as search engine retrieval, demographic data collection, human-computer interaction, access control, and surveillance that are based on facial images.

This problem attracted the attention of a number of researchers. One of the important steps of a gender recognition system is to represent facial images with the most discriminative features; different types of features have been used such as gray scale, shape, multiscale, texture etc. [1]. Another important step of the system is classification; different techniques like support vector machines (SVM) [2], artificial neural networks (ANN) [3], and nearest neighbor (NN) [4] have been used for classification. One of the earlier gender recognition systems was proposed by Gutta et. al. [3]. Moghaddam and Yang [2] used SVM gender classification and obtained a 96.62% accuracy on FERET database. Baluja and Rowley [5] used an AdaBoost based method for

gender recognition; for validation, they used FERET database and reported the best accuracy of 94.4%. Lu and Shi [6] fused the features from left eye, upper face region and nose in their gender recognition approach; their results showed that the fusion of features from different regions outperformed the whole face approach. Inspired by this idea, Alexandre [7] used a fusion approach based on three different types of features from multiple scales. He worked on normalized images of resolutions (20 x 20, 36 x 36 and 128 x 128) to extract shape and texture features (LBP). He achieved an accuracy of 99.07% on FERET database. Zheng and Lu [1] gave a comparison of 6 types of features using three classifiers and showed that for FERET database the best accuracy (99.1%) was obtained with features based on local Gabor binary pattern and LAD (LGBP-LDA) and SVM with automatic confidence (SVMAC). Ihsan et al. [4] used multiscale texture features based on dyadic wavelet transform and LBP (DyWT-LBP) and NN classifier and reported an accuracy of 99.25% on FERET.

Motivated by the success of multiscale texture features and SVM, in this paper we propose a new technique that employs SVM and multiscale texture features based on NSCT and Weber law descriptor (NSCT-WLD); this method has better recognition rate (99.50%) than the stat-of-the-art methods. In addition, this research reveals that it is only low subband in NSCT decomposition that is important for gender recognition; high frequency information contained in high subbands is not important, it might be interesting for detection tasks like age detection.

The rest of the paper is organized as follows. Section 2 gives the detail of the prosed system. In Section 3 the detail of experimental setup is given. Section 4 presents experimental results and their discussion. In Section 5, paper is concluded.

2   Methodology

In this section we describe the detail of the proposed Gender recognition system, which is a standard pattern recognition system and is shown in Figure 1.
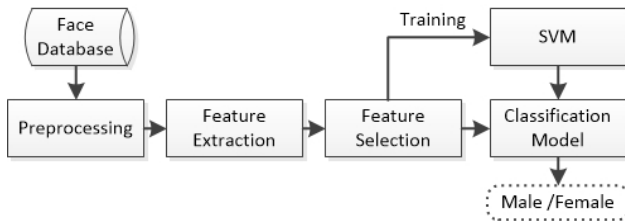

Fig. 1. System diagram for gender recognition

2.1   Preprocessing

Colored facial images are converted to YCbCr color system, where Y is luminance component; human eyes are sensitive to this component, as such it plays key role in the recognition process.   We used Y component for gender recognition.

2.2   Feature Extraction

In this section, we introduce a new feature extraction technique based on Nonsubsampled Contourlet Transform (NSCT) and WLD descriptor. First we give an overview of NSCT and WLD, and then describe our approach for feature extraction.

Nonsubsampled Contourlet Transform (NSCT)
Zhou at el. [8] proposed NSCT that decomposes an image into sub-bands of various scales and orientations and was applied for the enhancement of images. It involves two operations: Nonsubsampled Pyramid (NSP) decomposition and directional decomposition using Nonsubsampled Directional Filter Bank (NDFB). First, an image is decomposed into lowpass (low) and highpass (high) subbands using NSP decomposition. Second, applying NDFB, the high subband is decomposed into directional subbands. Figure 2 shows the steps of NSCT. This process is repeated on low subband to create a hierarchy of directional subbands of various scales.
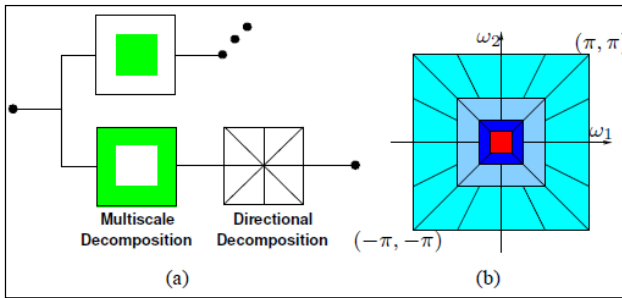
Fig. 2. NSCT decomposition. (a) NSP decomposition and (b) directional decomposition of high sub-band using NDFB [8].
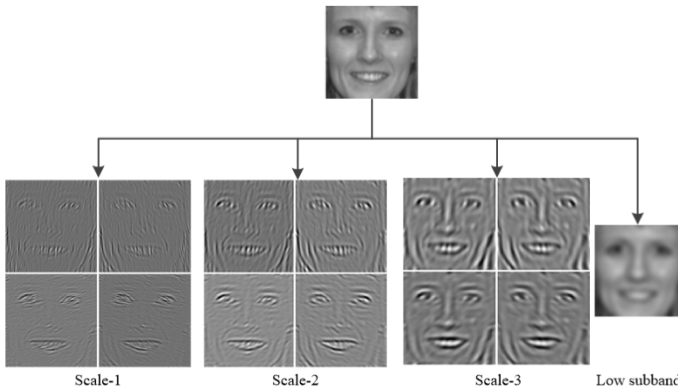
Fig. 3. Decomposition of an image using NSCT into three scales. At each scale, the image has been decomposed into 4 directional high subbands.

Figure 3 shows the multiscale hierarchy of an image created with NSCT; the image has been decomposed into 13 (one low and 12 high) subbands; there are three scales and at each scale there are 4 directional subbands.
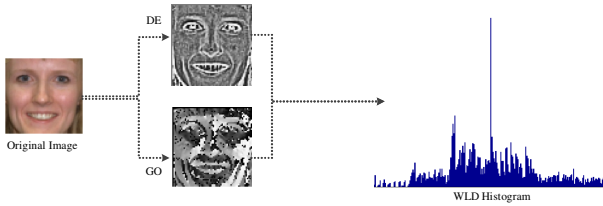
WLD Descriptor

This descriptor is based on Weber's law, which states that the ratio of the increment threshold to the background intensity is constant [9]. Its construction involves three main steps: Differential Excitation (DE), Gradient Orientation (GO) and WLD histogram. For each pixel x_c, DE is calculated using the following equation:

$$\zeta(x_c) = \arctan\left[\sum_{i=1}^{p}\left(\frac{x_i - x_c}{x_c}\right)\right]$$

where x_i, $i = 1, 2, …, p$, are pixels in the neighbourhood of the pixel x_c, as shown in Figure 4(a) for $p = 8$.
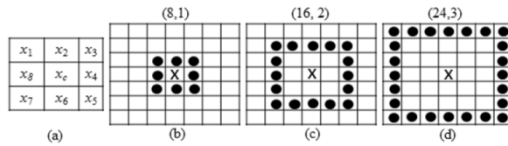
Fig. 4. Neighborhoods of a pixel for computing DE

Fig. 5. The computation of WLD descriptor. First DE and GO are computed, then using both, WLD histogram is calculated.

The DE $\zeta(x_c)$ takes positive and negative values. The arctangent function is used to prevent quick increase or decrease in the output when the input values become very small or very large.

The next important step is the computation of GO. For each pixel x_c, the gradient orientation $\theta(x_c)$ is calculated as follows:

$$\theta = \arctan\left(\frac{x_8 - x_4}{x_6 - x_2}\right)$$

and is mapped to the range $[0, 2\pi]$. Then the θ values corresponding to all pixels are quantized to T dominant gradient orientations Φ_t, $t = 1, 2, …, T$; a $\theta \in \left[\Phi_t - \frac{\pi}{T}, \Phi_t + \frac{\pi}{T}\right]$ is quantized to Φ_t.

The last step is building WLD histogram. Using both DE and GO for each pixel, WLD histogram is calculated, see Figure 5. The first step is to compute sub-histograms H_t: $t = 0, 1, 2, …, T\text{-}1$ of DEs corresponding to each dominant GO ϕ_t: $t = 0, 1, 2, …, T\text{-}1$; H_t is built using DE values of all pixels having dominant GO ϕ_t.

Then each sub-histogram H_t: $t = 0, 1, 2, …, T$-1 is further partitioned into M sub-histograms $H_{m,t}$: $m = 0, 1, 2, …, M$-1, each with S bins. In this way we get a histogram matrix $H_{m,t}$: $m = 0, 1, 2, …, M$-1, $t = 0, 1, 2, …, T$-1, where each column corresponds to a dominant GO ϕ_t. By concatenating the sub-histograms in each row m, a sub-histogram $H_m = \{ H_{m,t}: t = 0, 1, 2, …, T-1\}$ is built. Subsequently, sub-histograms H_m: $m = 0, 1, 2, …, M$-1 are concatenated into a histogram $H = \{ H_m: m = 0, 1, 2, …, M-1\}$. This histogram forms the representation of the image and is referred to as WLD descriptor. This descriptor involves three free parameters:

- T,   the number of dominant GOs ϕ_t: $t = 0, 1, 2, …, T$-1,
- M, the number of segments $H_{m,t}$ of each sub-histogram H_t corresponding to a dominant GO ϕ_t, and
- S, the number of bins in each sub-histogram $H_{m,t}$.

By incorporating the free parameters, we represent a WLD operator that computes WLD descriptor by *WLD* (T, M, S).

Multi-Scale WLD
The WLD descriptor described above uses 3x3 fixed size neighbourhood. This representation is unable to discriminate local salient patterns in various granularities. To represent local salient patterns at various scales, it is extended to multi-scale WLD, which is calculated using neighbourhoods (P, R) of various sizes and shapes; we used three square neighbourhoods (P, R) with P = 8, 16, 24 and R = 1, 2, 3, as shown in Figure 4(b-d). Multi-scale WLD descriptor is built by concatenating the WLD histograms calculated with three neighbourhood sizes.

NSCT-WLD Feature Descriptor
Now we describe our approach for feature extraction using NSCT and WLD; we call it *NSCT-WLD Feature Descriptor*. First a facial image is decomposed using NSCT into one low (L) and n high subbands: H_{ij}, where $i = 1, 2, …, s, j = 1, 2, …, o, s$ is the number of scales, o is the number of orientations at each scale and $n = sxo$; H_{ij} represents a high subband with scale i and orientation j. The next step is to compute multi-scale WLD descriptor from each subband. WLD described above is not spatially enhanced; it does not take into account the spatial locality of microstructures; to improve its discriminatory power, we incorporate spatial locality of microstructures by dividing each subband into m non-overlapping blocks. The detailed description of feature extraction from each subband is given below.

Step-1. Partition each subband H into m non-overlapping blocks $B_1, B_2, …, B_m$.
Step-2. For each block B_i, compute WLD histograms h_{i1}, h_{i2}, h_{i3} using neighbourhoods (8, 1), (16, 2), (24, 3), respectively,   and concatenate them to build multi-scale WLD histogram $h_i = [h_{i1}, h_{i2}, h_{i3}]$.
Step-3. Concatenate multi-scale WLD histograms h_i, $i = 1, 2, …, m$ corresponding to each block B_i and build the descriptor h for H i.e. $h =[h_1, h_2, …, h_m]$.

We investigated the effect of NSCT-WLD feature descriptors obtained from each subband and also using decision level fusion on the recognition performance.

2.3   Feature Selection

The dimension of the feature space generated by NSCT-WLD descriptor becomes high, there are a large number of redundant features, which not only increase the computational complexity but also cause to decline the recognition accuracy. For this problem we applied two-stage feature selection. In the first stage, we used Fisher method [10], which evaluates the significance of features using the class information, and ranks them; features are selected based on their ranks. The selected features are not independent because Fisher method does not take into account the dependence of features. To remove the redundancy due to feature dependence, we used Local Learning Based (LLB) method [11]; this method further reduces the dimension of the feature space by removing features that are not independent.

2.4   Classification

For classification, we used Support Vector Machines (SVM) because it is based on large margin theory and has better generalization. Given the training samples $\{(x_i, y_i)| 1 \leq i \leq N\}$, where $y_i \in \{+1, -1\}$ is the class label, the classification of a test sample x is performed using

$$g(x) = \sum_{i=1}^{N_s} \lambda_i\, y_i K(x_i, x) + w_0$$

where λ_i are Lagrange multipliers and N_s is the number of support vectors i.e. the training samples corresponding to non-zero λ_i's, $K(x_i, x)$ is the kernel function that expresses the inner product of data samples in the higher dimensional space. Two common kernel functions are radial basis function (RBF) and polynomial function. As RBF gives good performance for many applications, we used RBF in our gender recognition system. For implementation, we used LIBSVM.

3   Experimental Setup

In this section, we give the detail of the set-up used to perform the experiments.

3.1   Database

We used FERET [12] database to measure the performance and efficiency of the proposed system. FERET is the most popular and difficult database which was developed for research related to face recognition. The number of each subject's facial images is different, these images differ from each other based on expression, illumination, and pose. We used *fa* and *fb* sets of the database, the set *fb* contains facial images of the same subjects but with variations of expression, illumination and pose; both sets contain 1797 colored facial images of 403 male and 403 female subjects.

3.2   Parameter Selection

NSCT-WLD descriptor involves many parameters that have significant effect on performance. NSCT decomposition has two parameters: s - the number of scales and o - the number of orientations at each scale; for experiments, we used $s = 3$ and $o = 8$.

WLD involves two types of parameters: (T, M, S) and the number of non-overlapping blocks (b). For (T, M, S), we tested (4, 4, 5), (4, 6, 5), (12, 6, 15) and (12, 4, 5) and for b, we examined 4x4, 5x5, 6x6, 7x7, and 8x8 blocks. We found that (12, 4, 5) and 6x6 blocks gave the best results.

SVM with RBF kernel contain two parameters: γ and C; for finding the optimal parameter values, we used 5-fold cross validation and grid search. First a loose grid search was conducted to find the potential optimal region in (γ, C) space.   Then fine grid search was performed in this region to find the optimal parameter values.

4   Results and Discussion

To validate the performance of the proposed gender recognition system, first we present the results for NSCT and WLD descriptors. Then we give the recognition results obtained using NSCT-WLD descriptor. Finally, a comparison with stat-of-the-art methods is presented. For validation, we used 10-fold cross-validation and widely used performance measures: accuracy and AUC (area under ROC curve).

Results with NSCT descriptor
The Y component of each facial image is decomposed into one low and 24 high subbands H_{ij}, $i = 1, 2, 3$, $j = 1, 2, …, 8$. The coefficients in each subband are treated as features and the subband is converted into a feature vector. Then feature selection is applied to remove the redundant features. The recognition accuracy with each subband is shown in Figure 6; the best accuracy (95.3%) is obtained by low subband.
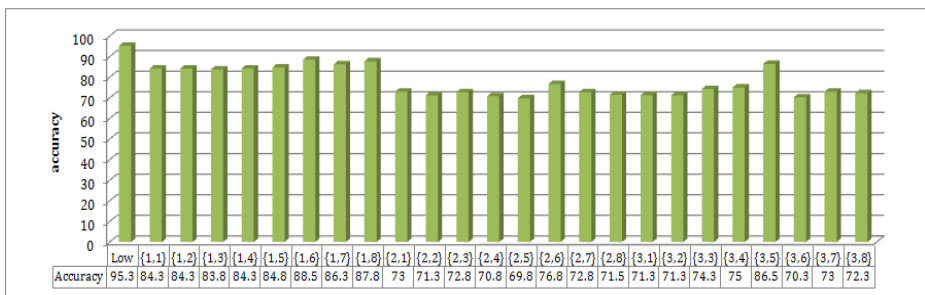

	Low	{1,1}	{1,2}	{1,3}	{1,4}	{1,5}	{1,6}	{1,7}	{1,8}	{2,1}	{2,2}	{2,3}	{2,4}	{2,5}	{2,6}	{2,7}	{2,8}	{3,1}	{3,2}	{3,3}	{3,4}	{3,5}	{3,6}	{3,7}	{3,8}
Accuracy	95.3	84.3	84.3	83.8	84.3	84.8	88.5	86.3	87.8	73	71.3	72.8	70.8	69.8	76.8	72.8	71.5	71.3	71.3	74.3	75	86.5	70.3	73	72.3

Fig. 6. Recognition rates of low and high subbands, $\{i, j\}$ means the high subband H_{ij}

Results with WLD descriptor
The Y component is divided into 6x6 blocks and multi-scale WLD histogram is computed from each block using three neighborhoods (8,1), (16, 2) and (24, 3). We tested four combinations of T, M, S parameters: (4, 4, 5), (4, 6, 5), (12, 6, 15) and (12, 4, 5).

To reduce the dimension of the feature space, two-stage feature selection was applied. Table 2 shows the recognition results by multi-scale WLD descriptor. The best accuracy (99.00%) was obtained with (12, 6, 15) and the best AUC was given by (12, 4, 5); note that in the former case the number of features is 29333 whereas in the latter case it is 9647.   In view of this, for NSCT-WLD descriptor, we employed (12, 4, 5).

Table 1. Recognition rates using multi-scale WLD descriptor

(T, M, S)	Feature Selection	Acc ± std	AUC± Std	#Features
(4, 4, 5)	Without selection	95.75±3.34	0.96±0.30	8640
	Fisher then LLB	97.75±2.18	0.97±0.02	4501
(4, 6, 5)	Without selection	95.50±3.87	0.95±0.03	12960
	Fisher then LLB	98.00±2.10	0.97±0.03	1142
(12, 6, 15)	Without selection	93.50±5.19	0.95±0.03	116640
	Fisher then LLB	**99.00±1.29**	**0.98±0.02**	**29333**
(12, 4, 5)	Without selection	95.75±4.41	0.95±0.04	25920
	Fisher then LLB	**98.75±1.31**	**0.99±0.009**	**9647**

Results with NSCT-WLD descriptor

For NSCT-WLD descriptor, we decomposed Y component of each facial image into one low and 24 high subbands H_{ij}, $i = 1, 2, 3, j = 1, 2, …, 8$. In view of the experiments with multi-scale WLD descriptor discussed in the previous paragraphs, we used T=12, M=4, S=5, and 6x6 blocks for NSCT-WLD descriptor. The dimension of NSCT-WLD descriptor is very big and several features are redundant; to reduce the dimension of the feature space, we applied two-stage feature selection. Figure 7 shows the recognition results for individual subbands. Further detail for seven subbands resulting in the highest accuracies is given in Table 3. It is obvious that feature selection not only reduces significantly the dimension of the feature space but also increases the recognition performance.
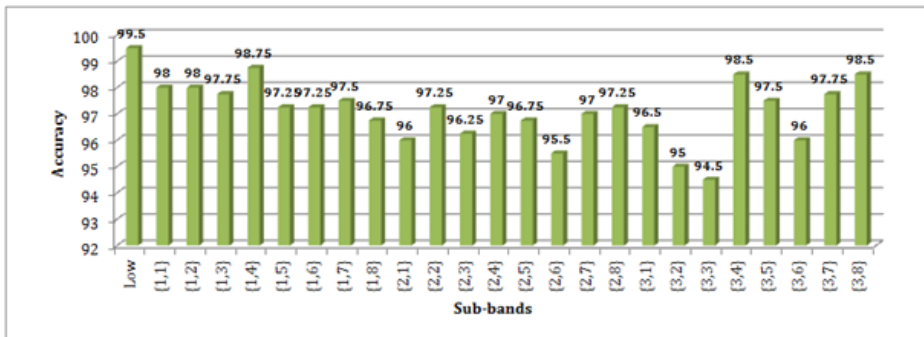
Fig. 7. Recognition accuracies obtained with NSCT-WLD descriptor. Here low means low subband, and $\{i, j\}$ stands for high subband H_{ij}.

Table 2. Recognitio rates with NSCT-WLD descriptor

| Subband	Feature Selection	Acc ± std	AUC± std	# Features
Low	Without Selection	98.25±2.05	0.98±0.02	25920
	Fisher then LLB	**99.50±1.05**	**0.99±0.01**	**299**
H_{14}	Without Selection	95.75±2.90	0.95±0.03	25920
	Fisher then LLB	98.75±2.42	0.98±0.01	2751
H_{34}	Without Selection	85.50±3.68	0.85±0.05	25920
	Fisher then LLB	98.50±1.74	0.98±0.02	3770

Table 3. Recognition performance with decision level fusion with majority vote

| # Subbands	Subbands	Acc ± std	AUC± std
3	Low, H_{14}, H_{34}	**99.50±1.5**	**0.993±0.02**
5	Low, H_{11}, H_{14}, H_{34}, H_{38}	99.00±2.1	0.980±0.02
7	Low, H_{11}, H_{12}, H_{13}, H_{14}, H_{34}, H_{38}	98.75±2.1	0.987±0.02

In addition, we performed decision level fusion using majority vote. We fused one low and different numbers (2, 4, and 6) of high subbands which have the highest recognition rates. The results are shown in Table 4 that indicate that fusion does not improve the recognition rate; instead accuracy declines with increasing the number of high subbands. The best accuracy (**99.50±1.58**) is obtained with the fusion of one low and two high subbands, which is not better than what (**99.50±1.05**) was obtained with low subband alone. This leads us to conclude that it is only low subband that plays key role in gender recognition; this is in accordance with our common experience that we can recognize a person's gender from a distance when most of high frequency features like wrinkles are not visible. High frequency detail is not important for gender recognition, it may be important for other recognition tasks like age estimation.

Comparison of NSCT, WLD, and NSCT-WLD descriptors
Figure 8 shows the ROC curves and AUC for the three methods; it indicates that NSCT-WLD outperforms the other two techniques.
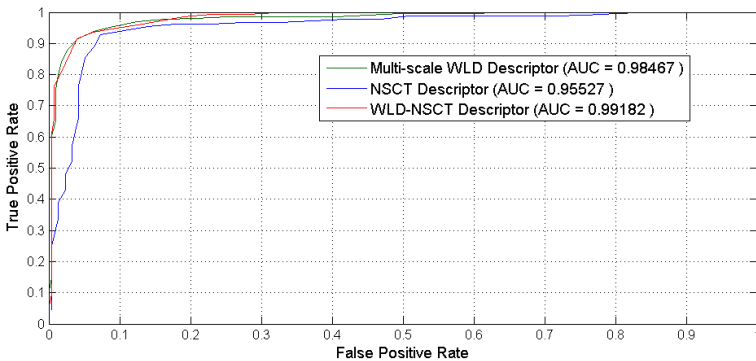
Fig. 8. ROC curves   for NSCT, WLD and NSCT-WLD descriptors
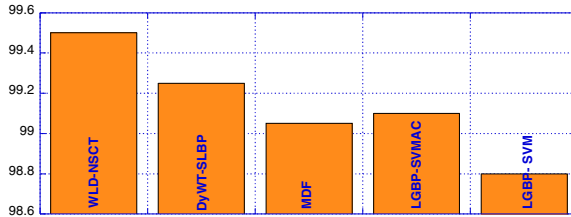
Fig. 9. Comparison with stat-of-the-art methods

Comparison with State-of-the-art Methods

For comparison, we selected recent gender recognition systems: Local Gabor Binary Pattern with LDA and SVMAC (LGBP-SVMAC) [1], Local Gabor Binary Pattern with LDA and SVM (LGBP- SVM) [1], Multi-resolution Decision Fusion method (MDF) [7], and the method based on dyadic wavelet transform and LBP (DyWT-LBP) [4]. The results in Figure 9 show that the proposed system performs better.

5   Conclusion

We proposed a gender recognition system that employs SVM, two-stage feature selection and multiscale texture features based on NSCT and WLD. We analyzed the impact on gender recognition of low and high subbands independently in the multiscale hierarchy of each facial image; also we examined the effect of fusing low and high subbands that gave the highest recognition rate. We found that it is only low subband that plays significant role in gender recognition; high subbands contain high frequency information that is not important for gender or face recognition; it is in accordance our common observation that we can recognize a person from a distance where high frequency information like wrinkles is not visible. The high frequency information contained in high subbands might be useful for detecting special effects like aging. The recognition accuracy achieved using NSCT-WLD features is not only better than that obtained by NSCT or WLD features but also the state-of-the-art methods. The two-stage feature selection approach reduces the dimension of the feature space and increase the recognition rate. The proposed system achieved an accuracy of 99.5% on FERET database which is the best accuracy obtained so far. In our future work we will explore the effect of NSCT-WLD descriptor on age detection.

References

1. Zang, J., Lu, B.L.: A support vector machine classifier with automatic confidence and its application to gender classification. Neurocomputing 74, 1926–1935 (2011)
2. Moghaddam, B., Yang, M.-H.: Gender classification with support vector machines. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 306–311 (March 2000)

3. Gutta, S., Wechsler, H., Phillips, P.: Gender and ethnic classification of face images. In: Third IEEE International Conference on Automatic Face and Gesture Recognition (FG 1998), pp. 194–199 (1998)
4. Ullah, I., Hussain, M., Aboalsamh, H., Muhammad, G., Mirza, A.M., Bebis, G.: Gender Recognition from Face Images with Dyadic Wavelet Transform and Local Binary Pattern. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Fowlkes, C., Wang, S., Choi, M.-H., Mantler, S., Schulze, J., Acevedo, D., Mueller, K., Papka, M. (eds.) ISVC 2012, Part II. LNCS, vol. 7432, pp. 409–419. Springer, Heidelberg (2012)
5. Baluja, S., Rowley, H.: Boosting sex identification performance. International Journal of Computer Vision 71(1), 111–119 (2007)
6. Lu, L., Shi, P.: Fusion of multiple facial regions for expression-invariant gender classification. IEICE Electronic Express 6(10), 587–593 (2009)
7. Alexandre, L.A.: Gender recognition: A multiscale decision fusion approach. Pattern Recognition Letters 31, 1422–1427 (2010)
8. Zhou, J., Cunha, A.L., Do, M.N.: Nonsubsampledcontourlet transform: construction and application in enhancement. In: Proc. ICIP 2005, pp. I 469-72 (2005)
9. Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W.: WLD: A robust local image descriptor. IEEE TPAMI 32(9), 1705–1720 (2010)
10. Hart, P.E., Duda, R.O., Stork, D.G.: Pattern Classification. Wiley-Interscience Publication (2001)
11. Sun, Y., Todorovic, S., Goodison, S.: Local-learning-based feature selection for high-dimensional data analysis. IEEE TPAMI 32(9), 1610–1626 (2010)
12. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE TPAMI 22(10), 1090–1104 (2000)
13. Veropoulos, K., Bebis, G., Webster, M.A.: Investigating the impact of face cate-gorization on recognition performance. In: Bebis, G., Boyle, R., Koracin, D., Parvin, B. (eds.) ISVC 2005. LNCS, vol. 3804, pp. 207–218. Springer, Heidelberg (2005)