Optik 127 (2016) 2670–2678

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Expression invariant face recognition using local binary patterns and contourlet transform

Hemprasad Y. Patil*, Ashwin G. Kothari, Kishor M. Bhurchandi

Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology, Nagpur 440010, Maharashtra, India

ARTICLE INFO

Article history:
Received 4 September 2014
Accepted 25 November 2015

Keywords:
Face recognition
Expression invariance
Contourlet transform
Image formation and processing
Image analysis

ABSTRACT

Face recognition is one of the most widely used biometric techniques in surveillance and security. Although many state-of-the-art systems have already been deployed across the world, the main challenge for feature extraction comes from the variations in the query images captured under uncontrolled situations. Unlike the local binary patterns or steerable pyramids which construct a feature vector strictly from spatial and transform domain, respectively, our approach built a method that exploits the features from both spatial as well as contourlet transform domain. Specifically, the contourlet transform exhibits properties like directionality and anisotropy and hence, results in extraction of significant features. Furthermore, we have proposed a novel coefficient enhancement algorithm which is applicable on the contourlet subbands to make the system more robust by enhancing skin region features. In addition, we show that the feature level fusion produces a robust feature vector, which yields competitive face recognition rates on the Cohn–Kanade (CK), Yale, JAFFE, ORL, CMU-AMP and our own face database. Finally, we benchmark our approach with other contemporary approaches and found it as most robust expression invariant face recognition technique.

1. Introduction

Face recognition lies at the heart of most of the biometric systems due to its non-intrusive behavior. Some of the classical approaches for face recognition are PCA [1], LDA [2] and ICA [3]. Currently the research focus has been shifted from holistic methods towards local pixel level feature extraction techniques. It is assumed that more robust face recognition systems can be designed using the local features that are extracted with dense descriptors [4]. Few researchers have implemented the local pixel level feature extraction based systems [4,5]. Although such systems achieve higher recognition rates than holistic feature extraction based approaches, the performance of such systems hampers due to variations like expression, pose, illumination and occlusion changes in the query image.

Facial expression variations are resulting from various emotions or verbal communications. Facial expressions are helpful for human-computer interaction (HCI) devices but act as a challenge for face recognition systems. Expression invariant face recognition problem is a prominent challenge due to drastic changes in a query image arising from facial deformations. As query images have different expressions than the training images, robust feature extraction methods play a vital role to improve the recognition rates of expression invariant face recognition systems. A thorough review of 2-D face recognition literature can be found in [6]. Expression invariant face recognition approaches can be classified into two groups: (1) Spatial domain techniques (2) Transform domain techniques. The spatial domain techniques are further divided into two categories: (i) Model based techniques and (ii) Subspace based techniques. The methods based on model based techniques employ appearance based models for facial fitting. These models utilize shape as well as texture features extracted from a query face image. Riaz et al. [7] implemented an active appearance model (AAM) based framework along with optical flow based feature extraction. The composite vector constructed from the AAM parameters and optical flow results into a robust facial image representation. Lee and Kim [8] extracted a feature vector from query image by fitting AAM. Further, they employed expression state analyzer which recognizes the expression and subsequently transform the initial feature vector into neutral facial feature vector. Although these techniques achieve promising results, fitting AAM is computationally heavy. Subspace based techniques utilize the traditional statistical procedures to analyze different attributes. Li et al. [9] implemented the notion of detached texture and geometric facial information. Both texture and geometric information are projected onto distinct principal component analysis spaces to further extract

* Corresponding author. Tel.: +91 9096019335.
E-mail address: hemprasadpatil@gmail.com (H.Y. Patil).

http://dx.doi.org/10.1016/j.ijleo.2015.11.187
0030-4026/

the relevant features. Chen and Lovell [10] employed principal component analysis (PCA) and subsequently wrap the constructed subspace to enhance the class separation performance. For adaptive subspace construction, they have performed whitening and filtering operations on the scatter matrix.

Transform domain based techniques extract the eminent features by transforming images into another domain. Vankayalapati and Kyamakya [11] extracted features from combination of the wavelet transform and Radon transform. Since the Radon transform is effective for extraction of local features and wavelet transform yields the time-frequency localization; their combination results into a robust feature vector for expression invariant face recognition. Abbas et al. [12] proposed a preprocessing technique to address expression variation problem. An original facial image in transformed into wavelet domain and detailed coefficients are equated to zero. Further, upon inverse transformation, they have obtained an image representation which is robust for expression variations. Kirtac et al. [13] computed the Gabor wavelet transform and performed nearest neighbor discriminant analysis. Gabor wavelet based features yield promising recognition rates as their kernels have similarity with the human visual cortex fields. To address the issue of more robust feature extraction, multidirectional transforms like contourlet transform [14] and curvelet transform [15] can be utilized. As contourlet transform is entirely proposed in digital domain [14], it does not suffer from critical sampling. Following approaches utilize the contourlet transform for facial feature extraction. Yi et al. [16] employed the threshold based algorithm on contourlet coefficients and performed classification of facial images using the SVM classifier. Boukabou and Bouridane [17] extracted pixel level features using contourlet transform and performed dimensionality reduction using PCA. Chen et al. [18] extracted the features from low frequency coefficients using statistical parameters. Histograms of high frequency contourlet coefficients of facial images are also exploited for efficient feature extraction. Weighted similarity measure is used for matching amongst various features. Pingfeng et al. [19] utilized pixel level features from contourlet transform coefficients and performed dimensionality reduction using coupled subspace analysis. All of these techniques do not include features obtained with dense pixel-level analysis.

We propose a novel feature level fusion technique that considers features from spatial as well as contourlet transform domain. The overall flow of proposed expression invariant face recognition system is presented in Fig. 1. Main contribution of this paper is to propose a novel boosting function for contourlet coefficient enhancement and feature level fusion. We have explored our approach on databases that contain drastic expression variations like the Cohn–Kanade (CK), Yale, JAFFE, ORL, CMU-AMP database and few of our own acquired images. The proposed algorithms are equally efficient with all the databases.

The outline of the rest of the paper is as follows. In Section 2, contourlet transform, Local binary patterns and Weber local descriptors are explained. Further, a novel coefficient enhancement function is proposed in Section 3. In Section 4, a feature extraction and feature level fusion scheme is presented along with feature matching using nearest neighbor technique. Experimentation on various face databases and results are discussed in Section 5.

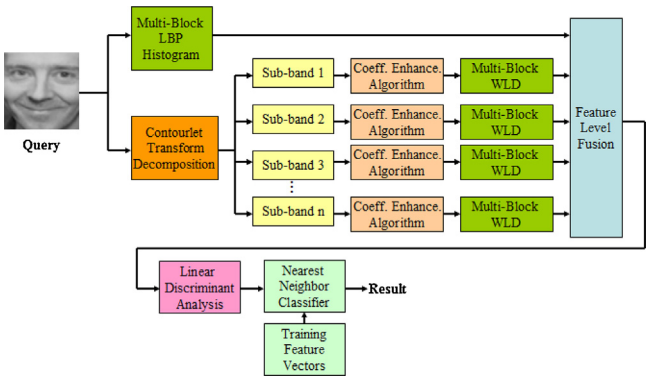

Fig. 1. Overall flow of the proposed face recognition method.

2. Materials and methods

2.1. Contourlet transform

Most of the classical frequency domain transforms such as Fourier transform, Discrete Cosine Transform and Wavelet transform are very efficient in capturing the details as far as they are exploited in one-dimensional context. In order to extract the details of an image, 2-D extensions are constructed from the 1-D separable basis functions. Such transforms are good at capturing the details when the image is a collection of 1-D piece-wise smooth signals [14]. However, practical applications like face recognition, human computer interaction demand the processing of images that involve smooth curves. For example, facial images have smooth curves that represent eyes, eyebrows, lips, nostrils etc. These smooth curves contain discriminant information which is essential for face recognition. Wavelet like transforms cannot effectively capture the features from contour like edges [14]. To address the problem of efficient representation of contour-like smooth edges, Do and Vetterli [14] proposed the contourlet transform directly in the discrete domain. Contourlet transform can successively approximate the image to have multiresolution representation. It can critically sample the original image with a small redundancy. It also represents the localization in spatial as well as frequency domain. In addition, its basis elements can be aligned to many directions which is useful to extract the directional features. Notably, its basis elements can be extended to form the shape of smooth curves. Although 2-D Gabor wavelet transform [20] and contourlet transform both can provide multidirectional and multiresolution expansions, the latter maintains critical sampling with small redundancy. The contourlet transform decomposition consists of the following stages: First, the original image is subjected to modified Laplacian pyramid (LP) transform [21] which acts as a wavelet like transform and hence utilized for edge detection. Further, on the resultant LP decomposition, modified iterated directional filter bank [22] is employed to extract smooth contour segments [14]. The modified Laplacian pyramid transform and modified iterated directional filter banks are illustrated in the following sub-sections.

2.1.1. Modified Laplacian pyramid transform

Burt and Adelson [23] proposed the Laplacian pyramid transform for image coding. At each level of the LP decomposition, one low pass downsampled image (say *LPD*) is constructed by passing the original image through low pass analysis filter $H(w)$ and further downsampling it with factor N. In addition, image *LPD* is upsampled with factor N and filtered with synthesis filter $G(w)$. This results in the image *HPD*. The difference between original image and *HPD* is encoded as bandpass image (*BP*). This decomposition procedure is illustrated in Fig. 2. The *LPD* image is further subjected to similar decomposition procedure based on number of levels of decomposition.

Do and Vetterli [21] proposed the modified Laplacian pyramid decomposition by altering the analysis and synthesis filters to orthogonal filters. A series of mappings $\{\Theta_k\}_{k\in\Gamma}$ defined in the bounds of Hilbert space H are denoted by a 'frame' if the following

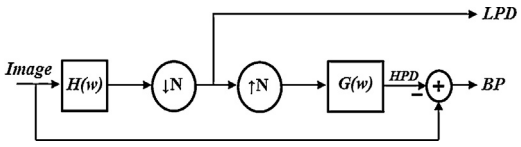

Fig. 2. Laplacian pyramid decomposition process.

conditions in (1) are true [21] for all $f \in H$, where η and ρ are frame bounds and Γ is subset of the Hilbert space.

$$\left.\begin{aligned} \eta\|f\|^2 \le \sum_{k\in\Gamma} \left|\langle f, \Theta_k\rangle\right|^2 \le \rho\|f\|^2 \\ \eta > 0, \quad \rho < \infty \end{aligned}\right\} \tag{1}$$

if $\eta = \rho$, the frame has tight bounds where no highpass frequency is overlapped with lowpass band after downsampling. The modified Laplacian pyramid transform delivers decomposition frames with critical sampling and is not affected by scrambled frequencies. In the context of face recognition problem, modified Laplacian pyramid decomposition frames of a facial image is shown in Fig. 3. Original image is decomposed at three levels. Bandpass images corresponding to level-1, level-2 and level-3 are BP_1, BP_2 and BP_3. The lowpass decomposed image at level-3 is denoted by LPD_3.

2.1.2. Modified iterated directional filter banks

Bamberger and Smith [24] proposed 2-D directional filter bank (2-D DFB) framework, which is capable of partitioning frequency spectrum into wedge like patches. 2-D DFB framework requires input image to be modulated. Modified iterated directional filter bank [14] avoids modulation of input image, for computational simplicity. It consists of combination of shearing operator framework and quincunx filter banks as shown in Fig. 4. Modified LP decomposition frame [21] is subjected to shearing operator. Due to fan like shape of quincunx filters, original frequency spectrum gets divided into parallel and perpendicular directions.

Iterated directional filter bank framework merges the point singularities acquired by modified LP transformation into linear structures. The resultant basis elements have elongated shape and are able to capture point singularities along the contour. Hence the transformation using such basis elements is named as the contourlet transform [14]. Fig. 5 illustrates the decomposition using contourlet transform on a facial image from Yale facial database. The forward 2-level decomposition is performed resulting in (4,8) subbands, respectively along with residual low passband.

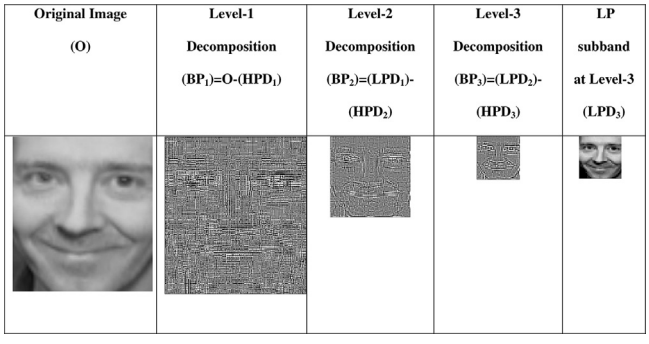

Fig. 3. Modified Laplacian pyramid decomposition frames [21] for a facial image.

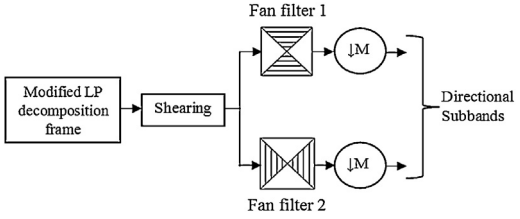

Fig. 4. Forward decomposition with shearing operator and quincunx filter bank [14]. Shearing operator controls the angular direction.

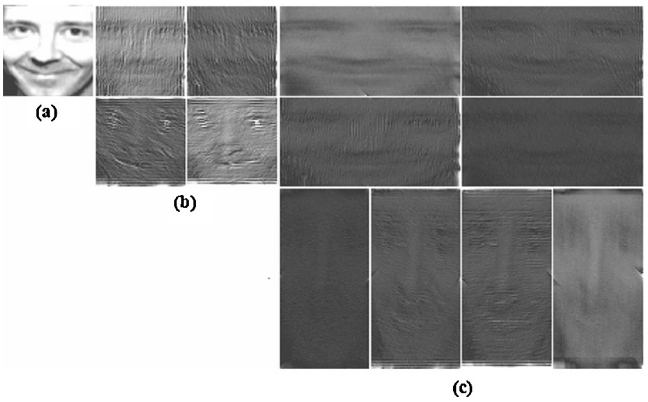

Fig. 5. Contourlet transform 2-level forward decomposition of a facial image. (a) Low pass subband (b) level-2 decomposition ($d = 2$, $2^d = 4$ directional subbands) (c) level-1 decomposition ($d = 3$, $2^d = 8$ directional subbands).

2.2. Local binary patterns (LBP)

Local binary pattern operator is widely used for texture classification. This operator creates new labeled representation which is invariant to monotonic gray-level variations. To assign a label for each pixel, its intensity value is used as threshold and compared against pixel values on 3×3-neighborhood. In general, LBP is computed with P sampling points ($g_p \in (0 \ldots P-1)$) in the neighborhood of center pixel g_{mid} at radial distance R, by Eqs. (2) and (3) [25].

$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} v_s\left(g_p - g_{\text{mid}}\right) 2^p \tag{2}$$

where

$$v_s(\text{diff}) = \begin{cases} 1, & (\text{diff}) \ge 1 \\ 0, & (\text{diff}) < 0 \end{cases} \tag{3}$$

Fig. 6 illustrates basic LBP operator and corresponding label for center pixel g_{mid}.

If sampling point p is not situated at the center of pixel grid, its value is calculated using bilinear interpolation technique [25]. Histogram of new labeled LBP representation is computed and utilized as a texture descriptor [4]. If there are no more than 2 bitwise transitions in a LBP pattern, it is called as uniform LBP [25].

2.3. Weber local descriptors (WLD)

Weber's law signifies the relationship between incremental threshold (ΔF) and surrounding intensity (F). It states that the ratio $\Delta F/F$ is always constant and denoted by Weber fraction [26]

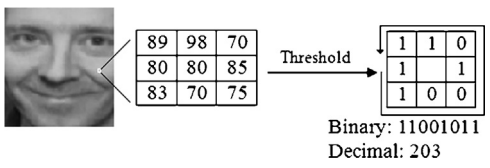

Fig. 6. LBP operator on facial image.

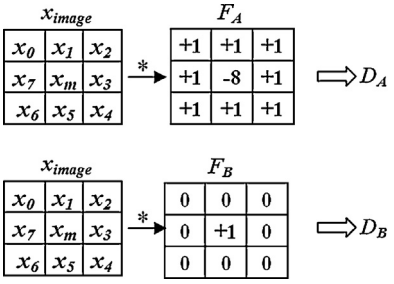

Fig. 7. Computation of differential excitation components.

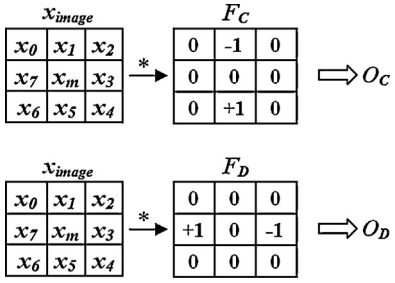

Fig. 8. Computation of orientation components O_C and O_D using F_C and F_D filters.

where, F is initial intensity in the signal/image and ΔF is the smallest change which is noticeable. For example, in signal processing context, in a crowded room, one has to speak loud in order to get heard, whereas a whispering sound is sufficient in a quiet place to get noticed. Such a phenomenon in the context of human vision is known as 'Just Noticeable Difference' (JND). Chen et al. [26] proposed the JND based Weber local descriptor (WLD) which became popular feature extraction technique for texture classification [26], face detection [26] and face recognition [27]. WLD feature vector consists of the following components: differential excitation and orientation. Steps for computing differential excitation and orientation are described in this section. Feature representation using histogram of orientations and gradients is widely used for human detection [28] and face recognition [27]. In particular, it leads to a representation close to biological human vision. The WLD histogram is constructed after mapping of differential excitation and orientation components. Further, it is projected onto 1-D space for classification efficiency.

2.3.1. Differential excitation

As per the Weber's law, differential excitation (ϑ) considers the ratio of two quantities: one is the difference in the intensities of neighborhood pixels with central pixel and another is central pixel intensity [26]. Initially, to compute the difference between the intensities of neighborhood pixels with central pixel, original image $X = \{x_{\text{image}}\}$ is filtered with mask F_A as shown in Fig. 5 resulting in D_A.

$$\begin{aligned} D_A &= \sum_{j=0}^{L-1} \left(\Delta x_j\right) \\ &= \sum_{j=0}^{L-1} \left(x_j - x_m\right) \end{aligned} \tag{4}$$

where L denotes total neighboring pixels, e.g. in an 3×3 image patch, neighboring pixels $L = 8$. The central pixel is denoted by x_m. Further, the denominator term is obtained on filtering original image with filter mask F_B. The filter coefficients illustrate the extraction of central pixel value while ignoring neighboring pixels as shown in Fig. 7.

$$D_B = x_{\text{image}} = \sum_{j=0}^{L-1} x_m \tag{5}$$

Differential excitation for the central pixel $\vartheta(x_m)$ is calculated using arctangent ratio of D_A and D_B as shown in (6)

$$\vartheta(x_m) = \arctan\left[\frac{D_A}{D_B}\right] = \arctan\left[\sum_{j=0}^{L-1}\left(\frac{x_j - x_m}{x_m}\right)\right] \tag{6}$$

The usage of arctangent function limits the drastic changes in output in correspondence with input value. The output range is confined to $[-\pi/2, \pi/2]$. $\vartheta(x_m)$ may take negative values as it does not consider absolute value of the difference $\left(x_j - x_m\right)$ [26]. Hence, it is beneficial as a descriptor when values of X are both positive and negative, specifically in the case of contourlet coefficients.

2.3.2. Orientation

Orientation extracts the changes in horizontal and vertical directions. The original image X, $X = \{x_{\text{image}}\}$ is filtered with F_C and F_D to obtain O_C and O_D as shown in Fig. 8. Filter coefficients F_C and F_D denote gradient orientation [26].

$$O_C = x_5 - x_1 \tag{7}$$

and

$$O_D = x_7 - x_3 \tag{8}$$

Overall orientation (Ξ) for central pixel x_m is computed as the ratio of horizontal and vertical gradient orientation O_D and O_C:

$$\Xi(x_m) = \arctan\left[\frac{O_D}{O_C}\right] = \chi_{(O_C, O_D)} \tag{9}$$

Arctangent function maps the ratio (O_D/O_C) within the range of $[-\pi/2, \pi/2]$. To further simplify, orientation Ξ is mapped to $\hat{\Xi}$ where, the range span of $\hat{\Xi}$ is $[0, 2\pi]$. The mapping $\Omega : \Xi \mapsto \hat{\Xi}$ is given by (10) as in [26]:

$$\begin{aligned} \hat{\Xi} &= \chi_{(O_C, O_D)} + \pi, \quad \text{and} \\ \chi_{(O_C, O_D)} &= \begin{cases} \Xi, & O_D > 0 \text{ and } O_C > 0 \\ \pi - \Xi, & O_D > 0 \text{ and } O_C < 0 \\ \Xi - \pi, & O_D < 0 \text{ and } O_C < 0 \\ -\Xi, & O_D < 0 \text{ and } O_C > 0 \end{cases} \end{aligned} \tag{10}$$

Further, $\hat{\Xi}$ is quantized into dominant orientations Φ to reduce the computational complexity. The orientation values $\hat{\Xi}$ between $[\Phi - (\pi/T), \Phi + (\pi/T)]$ are denoted by Φ_T, where T is number of dominant orientations [26].

$$\begin{aligned} \Phi_T &= fn_{\text{quan}}(\hat{\Xi}) \\ \Phi &\cong \Phi_T \end{aligned} \tag{11}$$

2.3.3. WLD histogram

Initially, a 2-D WLD histogram representation $\{\text{WLD}(\vartheta_{\text{row}}, \Phi_{\text{col}})\}$ is built using differential excitation component ϑ and gradient orientation Φ for an image or transform domain representation [26]. Such a representation allows to summarizethe frequencies of

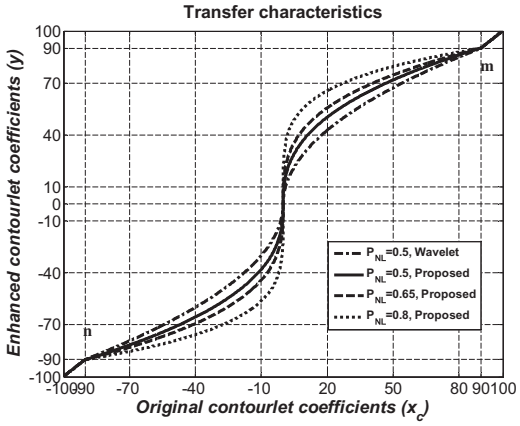

Fig. 9. Enhancement map for contourlet coefficients between original coefficients (x) and enhanced coefficients (y) using proposed algorithm for $m = 90$ and $n = -90$ and (i) $P_{NL} = 0.5$ using Wavelet based enhancement [29], (ii) $P_{NL} = 0.5$, proposed algorithm, (iii) $P_{NL} = 0.65$, proposed algorithm and (iv) $P_{NL} = 0.8$, proposed algorithm.

ϑ for a particular Φ. For classification simplicity, such a 2-D WLD histogram is mapped onto 1-D histogram.

3. Proposed coefficient enhancement function

The piecewise non-linear enhancement functions are used in conjuction with wavelet transform [29,30] and curvelet transform [31] for spatial domain image enhancement. On improvement of subband coeffients, inverse transform is computed resulting in contrast enhancement as in many image denoising methods. We propose novel functions to enhance contourlet transform coefficients and further processing for feature extraction using heuristic experimentation. Moreover, the proposed method does not require inverse transformation, hence its computational cost is less than [31]. We propose the mapping function $y(x_c)$ as in (12), which is monotonous, single valued and non-linear.

$$
\begin{aligned}
y(x_c) &= 1 && \text{if } x_c > 0 \text{ and } x_c > m \\
y(x_c) &= \left(\frac{m - \log_{10} P_{NL}}{|x_c|}\right)^{P^{(3*\sin(P_{NL}))/2}} && \text{if } x_c > 0 \text{ and } x_c \le m \\
y(x_c) &= 1 && \text{if } x_c < 0 \text{ and } x_c < n \\
y(x_c) &= \left(\frac{n - \log_{10} P_{NL}}{|x_c|}\right)^{P^{(3*\sin(P_{NL}))/2}} && \text{if } x_c < 0 \text{ and } x_c \ge n \\
y(x_c) &= 0 && \text{if } x_c = 0
\end{aligned}
\tag{12}
$$

here P_{NL} denotes degree of non-linearity of the curve and $0 < P_{NL} < 1$ [29], 'm' and 'n' are mapping thresholds on positive and negative axes. If value of x_c is greater than m or less than n, no need to enhance the value of coefficients in this range. The original contourlet transform coefficients are multiplied by mapping function $y(x_c)$ to yield enhanced contourlet coefficient $\hat{x}_c$ as,

$$\hat{x}_c = y(x_c) \times x_c \tag{13}$$

The enhancement map between original contourlet coefficients and enhanced contourlet coefficients is as shown in Fig. 9 for $m = 90$ and $n = -90$ and $P_{NL} = 0.5$,0.65 and 0.8. It should be noted that the proposed enhancement function offers more boosting than wavelet based approach [29].

Unlike existing transform domain coefficient enhancement techniques for which $|m| = |n|$ [29], the proposed contourlet coefficient enhancement algorithm is dynamic and the values of m and n can be computed based on the span of contourlet coefficients.

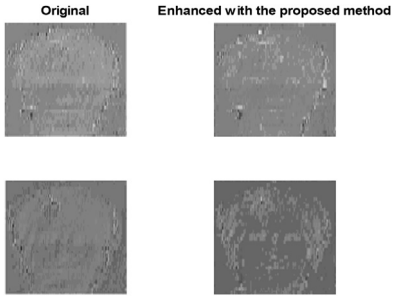

Fig. 10. Representative scaled original and enhanced contourlet coefficient subbands at level-1 decomposition using proposed mapping function for a facial image from the Yale face database.

The effectiveness of proposed contourlet coefficient enhancement technique is illustrated in Fig. 10. It displays original contourlet subband coefficients and enhanced contourlet coefficients scaled for displaying as images. From the figure, it is evident that the prominent features like eyes, nose-tip and mouth are enhanced using our proposed algorithm. Thus, this mapping function is very useful in facial feature enhancement and extraction.

4. Proposed face recognition method

In this section, we illustrate the expression invariant face recognition method based on the spatial and transform domain features and feature level fusion in details. In the proposed method, multi-block local binary patterns are first used to extract the dense local variations in spatial domain, and then original image is subjected to contourlet transform in order to obtain contourlet subbands. We compute contourlet coefficient subbands using the novel contourlet enhancement algorithm to facilitate more robust subband representations which, in turn expedites a robust feature vector for representing the facial image. The working mechanism of proposed face recognition method is as shown in Fig. 1.

We obtain n subbands after contourlet transform decomposition for S scales. Further, each subband is processed using the proposed CE algorithm as illustrated in Algorithm 1. Note that, mapping control measures κ_1 and κ_2 limit the enhancement range on both positive as well as negative axes separately.

Algorithm 1. Proposed coefficient enhancement (CE) algorithm.

Input: Contourlet transform subband X, $x_c \in X$, Mapping control measures κ_1, κ_2, Degree of non-linearity P_{NL}
Output: Enhanced contourlet coefficient subband $\hat{X}$, $\hat{x}_c \in \hat{X}$
Algorithm:
Step 1 (Initialization):
1.1 Compute maximum positive coefficient in X, denote it with ϖ.
1.2 Compute $m = \kappa_1 \cdot \varpi$
1.3 Compute minimum negative coefficient in X, denote it with Υ.
1.4 Compute $n = \kappa_2 \cdot \Upsilon$
Step 2 (Compute mapping function):
for each coefficient $x_c \in X$, calculate $y(x_c)$ as in (12).
Step 3 (Enhancement):
Compute enhanced coefficient $\hat{x}_c = y(x_c) \times x_c$
Repeat steps 2 and 3 for all x_c to obtain $\hat{X}$.

In addition, we proposed a novel face recognition method which utilizes the features from both spatial as well as contourlet transform domain features. Initially, the system is trained using Algorithm 2. The resultant feature vectors corresponding to training images are stored in the feature vector database.

Algorithm 2. Proposed training algorithm to extract expression invariant features.

Input: Train image, mapping measures κ_1, κ_2, LBP block size L_P, WLD block size W_P, contourlet transform scales S, degree of non-linearity P_{NL}
Output: Feature vector for input train image
Algorithm:
Step 1 (Preprocessing):
1.1 Read the train image I.
1.2 Perform RGB into gray transformation if the image is not grayscale.
1.3 Resize image to $[256 \times 256]$ pixels.
1.4 Denote this image as J
Step 2 (Local Binary Pattern (LBP) computation):
2.1 Divide the image J into multiple blocks P_t each with block size L_P.
for each block **do**
for each pixel within the block **do**
2.2 calculate local binary patterns representation
2.3 calculate histogram H_{Block_i} for each block where $i = [1, 2, \ldots, P_t]$
2.4 concatenate all histograms to form a feature vector H_{LBP}.
Step 3 (Computation of the contourlet transform coefficient subbands):
for each $s \in S$, compute 2^s contourlet transform coefficient subbands.
for each subband, process it using CE algorithm to obtain enhanced contourlet coefficients $\hat{X}$.
Step 4 (Weber local descriptor (WLD) computation):
4.1 Divide the enhanced subband $\hat{X}$ into multiple blocks P_q each with block size W_P.
for each block **do**
for each coefficient value within the block **do**
4.2 calculate Weber local descriptor
4.3 calculate 1-D WLD histogram
4.4 concatenate all histograms to form a feature vector $H_{\text{ECT-WLD}}$.
Step 5 (Feature level fusion):
Perform feature level fusion by concatenating H_{LBP} and $H_{\text{ECT-WLD}}$ into a feature
vector.
Step 6 (Dimensionality reduction):
Perform dimensionality reduction using linear discriminant analysis (LDA) [2]
Step 7 (Feature vector):
Save the feature vector to train feature vector database fv_{database}.

A test feature vector is compared with the train feature vector database and classified using nearest neighbor classifier. The algorithm to obtain decision regarding class of test image is shown in Algorithm 3. The parameters κ_1, κ_2, L_P, W_P, S and P_{NL} must be same for training images and a test image.

Algorithm 3. Proposed algorithm used to match the test image against training images in the database.

Input: Test image, mapping measures κ_1,κ_2, LBP block size L_P, WLD block size W_P, contourlet transform scales S, degree of non-linearity P_{NL}
Output: Class label for the test image
Algorithm:
Step 1: Repeat steps 1 to 6 of Algorithm 2 to obtain the discriminant feature vector fv_{test}.
Step 2 (Nearest neighbor classifier): Compare test image feature vector fv_{test} against multi-class feature vector from feature vector database fv_{database} to find minimum Euclidean distance. Nearest match below predefined threshold τ is declared as particular class (person in this context).

5. Experiments and discussion

In the following section, we provide a benchmarking analysis of our technique with other state-of-the-art approaches.

5.1. Databases

We evaluate the performance of proposed scheme using six facial expression databases: (i) Cohn–Kanade (CK), (ii) Yale, (iii) JAFFE, (iv) ORL, (v) CMU-AMP and (vi) Our own database.

Fig. 11 shows some sample images from these six databases.

The Cohn–Kanade (CK) face expression dataset [32], contains 8795 images (488 expression sequences (frames separated from videos) from 97 subjects performing the six universal facial expressions (happy, fear, surprise, anger, disgust and sadness) performed in mild to severe strength. It is a popular dataset for experimentation on expression invariant face recognition. The Yale face database [2], [33] consists of 165 gray scale image of 15 individuals (11 images/person). The expression variations considered are normal, wink, sleepy, happy, surprised, no glasses, center-light and sad. The Japanese Female Face Expression database (JAFFE) [34] contains a set of 213 facial images captured from 10 Japanese models. The available facial expression variations are happy, surprise, fear, disgust, sad, anger and neutral. All the images are captured against light homogeneous background and the subjects are in frontal pose. It is widely used for expression invariant face recognition experimentation and benchmarking. The ORL database of faces [35] includes a total of 400 images collected from 40 distinct subjects with 10 dissimilar images per subject. The CMU-AMP database [36] consists of facial images from 13 subjects, with 75 images per subject. This database is captured under static lighting conditions and includes only human expression variations. We collected our own facial database that contains expression variations like smile, anger, disgust, fear, surprise, sad and neutral. These are a total of 560 images in our database captured from 80 distinct subjects. All experiments are done on a computer system with Intel i7-4770, CPU@3.4 GHz, 32 GB RAM and Windows 8 operating system using MATLAB.

5.2. Implementation of the proposed algorithm

In all the databases, facial images are well normalized and resized to 256×256 pixel. In order to avoid pixel intensity changes due to small illumination variations, histogram equalization is employed. From the overall system structure, the local binary patterns (LBP) and Weber local descriptors (WLD) must be applied block-wise, as illustrated in section 1. We performed heuristic experimentation in order to obtain optimum block-size to achieve highest rank-one recognition rates. The block size values for the LBP and WLD are $L_P = 16 \times 16$ and $W_P = 8 \times 8$ pixels, respectively. Further, the CE algorithm parameters are selected as $\kappa_1 = 0.33$, $\kappa_2 = 0.3$, $P_{NL} = 0.5$, with contourlet transform scales $S = (3,4,4)$.

5.2.1. Experimentation on the Cohn–Kanade data set

The experimentation on the Cohn–Kanade (CK) data set is performed by following the two distinct protocols as defined in [7,37]. In [7], Riaz et al. considered the facial images from 62 subjects with 4000 images. Further, 66% of the selected images are labeled as a 'training' dataset and remaining as a 'testing' dataset. We report the average rank-one recognition results and benchmark with AAM based algorithm [7] in Table 1.

For performance comparison with compressed sensing based approach (B-JSM) [37], we follow the experimentation protocol as mentioned in [37]. J_T stands for the number of training images per person employed for face recognition performance evaluation. Training images are selected randomly and recognition rates are computed after performing 9 trials each. Table 2 displays that in all settings our method yields higher rank-one recognition rates.

5.2.2. Experimentation on the Yale face database

In order to evaluate the recognition performance, we arbitrarily select 3 images/person and specify the resultant rank-one recognition rate. Comparison of our method with existing approaches is presented in Table 3. We obtain the highest recognition rate with least training images (3 faces/person).

5.2.3. Experimentation on the JAFFE database

For JAFFE face database, we arbitrarily selected 4 images per person as a training dataset and remaining as a testing dataset.

Fig. 11. Representative faces with expression variations from (1) the Cohn–Kanade face database (first row), (2) the Yale face database (second row), (3) the JAFFE facial expression database (third row), (4) the ORL database (forth row), (5) the CMU-AMP database (fifth row) and (6) Our own database (sixth row).

Table 1
Benchmarking of rank-one recognition rates on the Cohn–Kanade data set with Active Appearance Model based technique [7].

Algorithm	Rank-1 recognition rate (%)	Number of training images	Number of testing images	Total images for experiment
AAM-BDT[7]	91.17	2679	1381	4060
AAM-BN [7]	98.69	2679	1381	4060
Proposed	**99.56**	2679	1381	4060

Table 2
Comparison of rank-one scores on the Cohn–Kanade data set with B-JSM [37].

Number of training images per person (J)	Rank one recognition rate	
	B-JSM (%) [37]	Proposed (%)
5	95.47	**97.16**
6	95.93	**97.86**
7	96.15	**98.49**
10	97.14	**99.17**

Table 3
Benchmarking with other approaches (the Yale database).

Algorithm	Number of training faces/person	Rank-one recognition rate (%)
Gabor-NNDA [13]	5	90.00
2D-GFD-FLD [38]	3	70.80
Wavelet-DCT [12]	3	91.82
FT-DCT-DWT [39]	3	82.50
Proposed	**3**	**99.30**

Table 4 displays the comparison of proposed approach with the state-of-the-art methods. From Table 4, it can be observed that, with least number of faces labeled as 'training' images/person, we achieve highest recognition rate as compared with the published literature.

Table 4
Benchmarking with other approaches (the JAFFE database).

Algorithm	Number of training images/person	Rank-one recognition rate (%)
LLE-Eigen [40]	18	93.93
Higher Order-SVD [41]	6	92.96
B-JSM [37]	4	95.12
PCA-FLD-ANN [42]	10	84.90
Proposed	**4**	**99.15**

5.2.4. Experimentation on the ORL database

The experimentation on ORL database is performed by selecting 3 images/person as 'training' set and remaining faces as 'testing' images. Table 5 depicts the comparison of proposed method with state-of-the-art approaches.

5.2.5. Experimentation on the CMU-AMP database

For CMU-AMP database, we achieved 100% rank-one recognition rate with 3 faces/person as well as 4 faces/person as a training dataset. Benchmarking with existing approaches is indicated in Table 6.

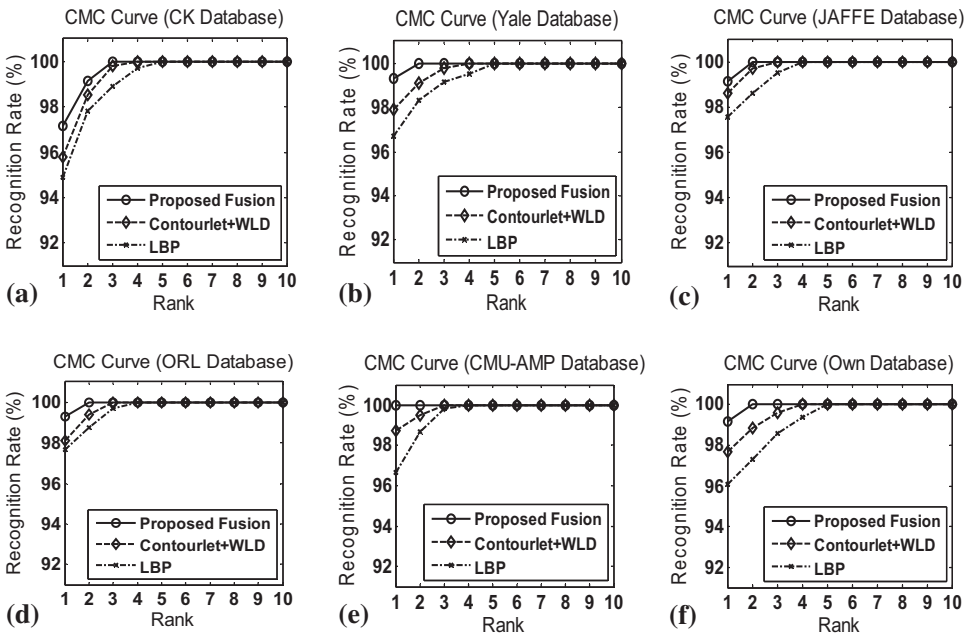

Fig. 12. Cumulative Match Curves for the (a) Cohn–Kanade (CK) database (b) Yale database (c) JAFFE database (d) ORL database (e) CMU-AMP database and (f) Own database upon feature extraction with the LBP, contourlet with WLD and proposed fusion techniques.

Table 5
Comparison of rank-one scores on the ORL database.

Algorithm	Number of training images/person	Rank-one recognition rate (%)
Wavelet-Energy [43]	5	90.50
Gabor-MPCA [44]	3	88.90
GM-DE [45]	3	95.00
DWT-DDFFE [46]	4	96.82
Proposed	**3**	**99.27**

Table 6
Comparison of rank-one recognition rates on the CMU-AMP database.

Algorithm	Number of training images/person	Rank-one recognition rate (%)
DCS [47]	4	100
B-JSM [37]	4	98.95
Proposed	**4**	**100**
Proposed	**3**	**100**

Table 7
Rank-one recognition rates using the proposed approach on our own database by varying training to testing images/person ratio.

Number of training images/person	Number of testing images/person	Rank-one recognition rate (%)
4	3	**99.16**
3	4	**98.43**
2	5	**96.25**

5.2.6. Validation on our own database

We experimented on our own database by arbitrarily choosing images to form a training set. We also have varied the ratio of number of images in gallery (training dataset) and number of probe images (testing dataset). Table 7 illustrates the rank-one recognition rates obtained using the proposed approach.

5.3. Discussion

These experimental outcomes illustrate that the proposed face recognition technique significantly improves the recognition performance when query face has different expression than the training facial images. Fig. 12 demonstrates the effectiveness of the proposed feature level fusion by performing experimentation on each database mentioned in Section 5.1. It can be noted that the proposed feature level fusion scheme outperforms the LBP and contourlet with WLD based face recognition in terms of rank-one recognition rates. The average recognition time for a single query face is 0.949 s. Although contourlet transform coefficient ranges vary drastically, the proposed CE algorithm proves to be efficient for the extraction of eminent features. The experimentation determines the effectiveness of the proposed technique even if the number of gallery images are sufficiently high.

6. Conclusion

The major contribution of this paper lies in proposing a novel approach using spatial as well as transform domain features for robust expression invariant face recognition. In this work, we proposed a contourlet coefficient enhancement algorithm along with feature level fusion method. Using a discrete domain contourlet transform we obtained discriminant features because of its ability to capture edge information along multi-resolution and multi-directions. Dense local descriptors LBP and WLD effectively extract features from spatial and transform domain. Unlike LBP, WLD proved to be advantageous to process the negative values for feature extraction from contourlet transform subbands. Based on this observation, we employed WLD to extract features from transform domain. A novel boosting function for contourlet transform feature enhancement is proposed. After rigorous experimentation, it is observed that the algorithm increases the performance of expression invariant face recognition system in terms of rank-one recognition rate. Experiments on the CK, Yale, JAFFE, ORL and CMU-AMP databases indicates the potential of the proposed method to act as an expression invariant face recognition system. The experimentation with our own grabbed images proves its generality. Further, the CE algorithm yields competitive enhancement in terms of discriminant features and hence can be utilized by various researchers for many applications in the area of facial analysis and micro expression analysis.

References

[1] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1987) 37–52.
[2] P.N. Belhumeur, J.P. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 711–720.
[3] M.S. Bartlett, Independent component representations for face recognition, in: Face Image Analysis by Unsupervised Learning, Springer, US, 2001, pp. 39–67.
[4] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 2037–2041.
[5] T. Gao, X.L. Feng, H. Lu, J.H. Zhai, A novel face feature descriptor using adaptively weighted extended LBP pyramid, Optik 124 (2013) 6286–6291.
[6] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. (CSUR) 35 (2003) 399–458.
[7] Z. Riaz, C. Mayer, M. Wimmer, M. Beetz, B. Radig, A model based approach for expressions invariant face recognition, in: M. Tistarelli, M. Nixon (Eds.), Advances in Biometrics, Springer, Berlin Heidelberg, 2009, pp. 289–298.
[8] H.S. Lee, D. Kim, Expression-invariant face recognition by facial expression transformations, Pattern Recognit. Lett. 29 (2008) 1797–1805.
[9] X. Li, G. Mori, H. Zhang, Expression-invariant face recognition with expression classification, in: The Third Canadian Conference on Computer and Robot Vision, Canada, 2006, pp. 77–79.
[10] S. Chen, B.C. Lovell, Illumination and expression invariant face recognition with one sample image, in: 17th International Conference on Pattern Recognition, Cambridge, 2004, pp. 300–303.
[11] H.D. Vankayalapati, K. Kyamakya, Nonlinear feature extraction approaches with application to face recognition over large databases, in: Second International Workshop on Nonlinear Dynamics and Synchronization, Klagenfurt, 2009, pp. 44–48.
[12] A. Abbas, M. Khalil, S. Abdel-Hay, H. Fahmy, Expression and illumination invariant preprocessing technique for face recognition, in: International Conference on Computer Engineering & Systems, IEEE, Cairo, 2008, pp. 59–64.
[13] K. Kirtac, O. Dolu, M. Gokmen, Face recognition by combining Gabor wavelets and nearest neighbor discriminant analysis, in: 23rd International Symposium on Computer and Information Sciences, Istanbul, 2008, pp. 1–5.
[14] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (2005) 2091–2106.
[15] E. Candes, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, Multiscale Model. Simul. 5 (2006) 861–899.
[16] W. Yi, L. Jian-Ping, L. Jie, L. Lin, The Contourlet Transform and SVM Classification for Face Recognition, in: International Conference on Apperceiving Computing and Intelligence Analysis, Chengdu, 2008, pp. 208–211.
[17] W.R. Boukabou, A. Bouridane, Contourlet-based feature extraction with PCA for face recognition, in: NASA/ESA Conference on Adaptive Hardware and Systems, IEEE, Noordwijk, 2008, pp. 482–486.
[18] C. Lei, W. Zhiyong, S. Bing, Z. Xingrong, S. Fei, Approach based on contourlet transform and weighted similarity measure for face recognition, in: IEEE International Conference on Intelligent Computing and Intelligent Systems, Shanghai, 2009, pp. 504–508.
[19] T. Pingfeng, G. Qu, N. Lin, W. Feifei, Face feature extraction and recognition using contourlet transform and coupled subspace analysis, in: Fifth International Conference on Biomedical Engineering and Informatics, Chongqing, 2012, pp. 270–273.
[20] T.S. Lee, Image representation using 2D Gabor wavelets, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 959–971.
[21] M.N. Do, M. Vetterli, Framing pyramids, IEEE Trans. Signal Process. 51 (2003) 2329–2342.
[22] M.N. Do, Directional Multiresolution Image Representations, Systems and Communication, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL), Lausanne, 2001, Ph.D. Thesis.
[23] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (1983) 532–540.
[24] R.H. Bamberger, M.J. Smith, A filter bank for the directional decomposition of images: theory and design, IEEE Trans. Signal Process. 40 (1992) 882–893.
[25] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 971–987.
[26] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, W. Gao, WLD: a robust local image descriptor, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1705–1720.
[27] S. Li, D. Gong, Y. Yuan, Face recognition using Weber local descriptors, Neurocomputing 122 (2013) 272–283.
[28] J. Bégard, N. Allezard, P. Sayd, Real-time human detection in urban scenes: local descriptors and classifiers selection with adaboost-like algorithms, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW'08, IEEE, Alaska, USA, 2008, pp. 1–8.
[29] K.V. Velde, Multi-scale color image enhancement, in: International Conference on Image Processing, IEEE, Japan, 1999, pp. 584–587.
[30] A. Łoza, D.R. Bull, P.R. Hill, A.M. Achim, Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients, Digital Signal Process. 23 (2013) 1856–1866.
[31] J.-L. Starck, F. Murtagh, E.J. Candès, D.L. Donoho, Gray and color image contrast enhancement by the curvelet transform, IEEE Trans. Image Process. 12 (2003) 706–717.
[32] T. Kanade, J.F. Cohn, Y. Tian, Comprehensive database for facial expression analysis, in: Fourth IEEE International Conference on Automatic Face and Gesture Recognition Grenoble, 2000, pp. 46–53.
[33] A.S. Georghiades, P.N. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 643–660.
[34] M.J. Lyons, M. Kamachi, J. Gyoba, Japanese Female Facial Expressions (JAFFE), 1997, ⟨http://kasrl.org/jaffe.html⟩, last accessed on 04/09/2014, (database of digital images).
[35] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Second IEEE Workshop on Applications of Computer Vision, Sarasota, FL, 1994, pp. 138–142.
[36] X. Liu, T. Chen, B.V.K.V. Kumar, Face authentication for multiple subjects using eigenflow, Pattern Recognit. 36 (2003) 313–328.
[37] P. Nagesh, B. Li, A compressive sensing approach for expression-invariant face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Miami, FL, 2009, pp. 1518–1525.
[38] R. Mutelo, W. Woo, S. Dlay, Discriminant analysis of the two-dimensional Gabor features for face recognition, IET Comput. Vision 2 (2008) 37–49.
[39] Z. Lihong, Z. Cheng, Z. Xili, S. Ying, Z. Yushi, Face recognition based on image transformation, in: WRI Global Congress on Intelligent Systems, IEEE, Xiamen, 2009, pp. 418–421.
[40] E.E. Abusham, D. Ngo, A. Teoh, Fusion of locally linear embedding and principal component analysis for face recognition, in: Pattern Recognition and Image Analysis, Springer, Portugal, 2005, pp. 326–333.
[41] T. Hua-Chun, Z. Yu-Jin, Expression-independent face recognition based on higher-order singular value decomposition, in: International Conference on Machine Learning and Cybernetics, IEEE, Kunming, 2008, pp. 2846–2851.
[42] P. Tsai, T. Jan, Expression-invariant face recognition system using subspace model analysis, in: IEEE International Conference on Systems, Man and Cybernetics, IEEE, Waikoloa, HI, 2005, pp. 1712–1717.
[43] C. Cunjian, Z. Jiashu, Wavelet energy entropy as a new feature extractor for face recognition, in: Fourth International Conference on Image and Graphics, Sichuan, 2007, pp. 616–619.
[44] N. Gudur, V. Asari, Gabor wavelet based modular PCA approach for expression and illumination invariant face recognition, in: 35th IEEE Applied Imagery and Pattern Recognition Workshop, Washington, DC, 2006, pp. 1–13.
[45] V. Maheshkar, S. Agarwal, V.K. Srivastava, S. Maheshkar, Face recognition using geometric measurements, directional edges and directional multiresolution information, Procedia Technol. 6 (2012) 939–946.
[46] N.L. Ajit Krisshna, V.K. Deepak, K. Manikantan, S. Ramachandran, Face recognition using transform domain feature extraction and PSO-based feature selection, Appl. Soft Comput. 22 (2014) 141–161.
[47] S. Taheri, V.M. Patel, R. Chellappa, Component-based recognition of faces and facial expressions, IEEE Trans. Affective Comput. 4 (2013) 360–371.