Mel- and Mellin-cepstral Feature Extraction Algorithms for Face Recognition

SERDAR CAKIR* AND A. ENIS CETIN

Department of Electrical and Electronics Engineering, Bilkent University, 06800 Ankara, Turkey
Corresponding author: cakir@bilkent.edu.tr

In this article, an image feature extraction method based on two-dimensional (2D) Mellin cepstrum is introduced. The concept of one-dimensional (1D) mel-cepstrum that is widely used in speech recognition is extended to two-dimensions using both the ordinary 2D Fourier transform and the Mellin transform. The resultant feature matrices are applied to two different classifiers such as common matrix approach and support vector machine to test the performance of the mel-cepstrum- and Mellin-cepstrum-based features. The AR face image database, ORL database, Yale database and FRGC database are used in experimental studies, which indicate that recognition rates obtained by the 2D mel-cepstrum-based method are superior to that obtained using 2D principal component analysis, 2D Fourier-Mellin transform and ordinary image matrix-based face recognition in both classifiers. Experimental results indicate that 2D cepstral analysis can also be used in other image feature extraction problems.

1. INTRODUCTION

Mel-cepstral analysis is one of the most popular feature extraction technique in speech processing applications, including speech and sound recognition and speaker identification. Two-dimensional (2D) cepstrum is also used in image registration and filtering applications [1–4]. To the best of our knowledge, 2D mel-cepstrum which is a variant of 2D cepstrum has not been used in image feature extraction, classification and recognition problems. The goals of this paper were to define the 2D mel-cepstrum and 2D Mellin-cepstrum and to show that they are viable image representation tools.

Fourier–Mellin transform (FMT) is a mathematical feature extraction tool which is used in some pattern recognition applications [5, 6]. The FMT is generally implemented by performing a log-polar mapping followed by the Fourier transform (FT) [6]. The main idea behind this approach is to represent rotation and scaling as translations along some axes and to take advantage of the translation invariance property of the FT.

Ordinary 2D cepstrum of a 2D signal is defined as the inverse FT of the logarithmic spectrum of the signal and it is computed using 2D fast Fourier transform (FFT). As a result, it is independent of pixel amplitude variations or gray-scale changes, which leads to robustness against illumination variations. Since it is an FT-based method, it is also independent of translational shifts. 2D mel-cepstrum which is based on logarithmic decomposition of frequency domain grid also has the same shift and amplitude invariance properties as the 2D cepstrum. In addition, 2D Mellin-cepstrum has rotation and scale invariance properties.

In this article, the 2D mel-cepstrum- and Mellin-cepstrum-based feature extraction methods are proposed. Mellin-cepstrum is a rotational and amplitude invariant. The proposed feature extraction technique is applied to the face recognition problem. It should be pointed out that our aim is not to develope a complete face recognition system but to illustrate the advantages of the 2D cepstral domain features.

Face recognition is still an active and popular area of research due to its various practical applications such as

security, surveillance and identification systems. Significant variations in the images of same faces and slight variations in the images of different faces make it difficult to recognize human faces. Feature extraction from facial images is one of the key steps in most face recognition systems [7, 8]. Principal component analysis (PCA) and linear discriminant analysis (LDA) are well-known techniques that were used in face recognition [9–11]. Although PCA is used as a successful dimensional reduction technique in face recognition, direct LDA-based methods cannot provide good performance when there are large variations and illumination changes in the face images. LDA with some extensions such as quadratic LDA [12], Fisher's LDA [13] and direct, exact LDA [14] were proposed. PCA is known as a popular linear feature extraction method that is used in one of the most famous techniques called eigenfaces. In the eigenface method, image space is simply projected to a low-dimensional feature space [15]. That is how the dimensional reduction is achieved.

In 2D mel-cepstrum and Mellin-cepstrum, the logarithmic division of the 2D discrete Fourier transform (DFT) grid provides the dimensionality reduction. This is also an intuitively valid representation as most natural images are low-pass in nature. Unlike the ordinary Fourier or discrete cosine transform (DCT) domain features, high-frequency DFT and DCT coefficients are not discarded in an ad-hoc manner. They are combined in bins of frequency values in a logarithmic manner during the 2D mel-cepstrum computation.

The rest of the paper is organized as follows. In Section 2, the proposed 2D cepstral domain feature extraction methods are described. In Section 3, a subspace-based pattern recognition method called common matrix approach (CMA) and the well-known classification method called support vector machine (SVM) are briefly explained. The 2D mel-cepstrum and Mellin-cepstrum matrices obtained from facial images are converted into vectors and classified using the SVM which was successfully used in face recognition applications [16–18]. In the CMA which is used as a pattern recognition engine in a face recognition application [19], matrices are directly used as the feature set. In Section 4, experimental results are presented.

2. THE 2D MEL- AND MELLIN-CEPSTRUM

In the literature, the 2D cepstrum was used for shadow detection, echo removal, automatic intensity control, enhancement of repetitive features and cepstral filtering [1–3]. In this article, 2D mel-cepstrum and Mellin-cepstrum are used for representing images or image regions.

We first introduce the 2D mel-cepstrum using the definition of 2D cepstrum which is defined as follows. 2D cepstrum $\hat{y}(p, q)$ of a 2D image $y(n_1, n_2)$ is given by

$$\hat{y}(p, q) = F_2^{-1}(\log(|Y(u, v)|^2)) \tag{1}$$

where (p, q) denotes 2D cepstral quefrency coordinates [32], F_2^{-1} denotes 2D inverse discrete-time fourier transform

(IDTFT) and $Y(u, v)$ is the 2D discrete-time Fourier transform (DTFT) of the image $y(n_1, n_2)$. Cepstrum sequence is an infinite extent sequence; however, it decays very fast [20]. In our implementation, the ranges of quefrency coordinates (p, q) are simply the same as the ranges of 2D input signal. As it can be observed from the Fig. 4a and b, the mel- and Mellin-cepstrum coefficients decay faster than Fourier coefficients. In practice, FFT algorithm is used to compute DTFT.

In 2D mel-cepstrum, the DTFT domain data are divided into non-uniform bins in a logarithmic manner as shown in Fig. 1 and the energy $|G(m, n)|^2$ of each bin is computed as follows:

$$|G(m, n)|^2 = \sum_{k, l \in B(m, n)} |Y(k, l)|^2 \tag{2}$$

where $Y(k, l)$ is the DFT of $y(n_1, n_2)$, and $B(m, n)$ is the (m, n)-th cell of the logarithmic grid. The range of cell numbers (m, n) depends on the non-uniform grid that is used in the computation of cepstral features. The frequency coefficients in a cell are grouped together to represent the corresponding cell. In each non-uniform grid, the number of cells and cell sizes differ in order to extract features with different frequency characteristics. The non-uniform grids used in cepstral feature computation have the ranges of 49, 39, 35 and 29. Cell or bin sizes are smaller at low frequencies compared with high frequencies. This approach is similar to the mel-cepstrum
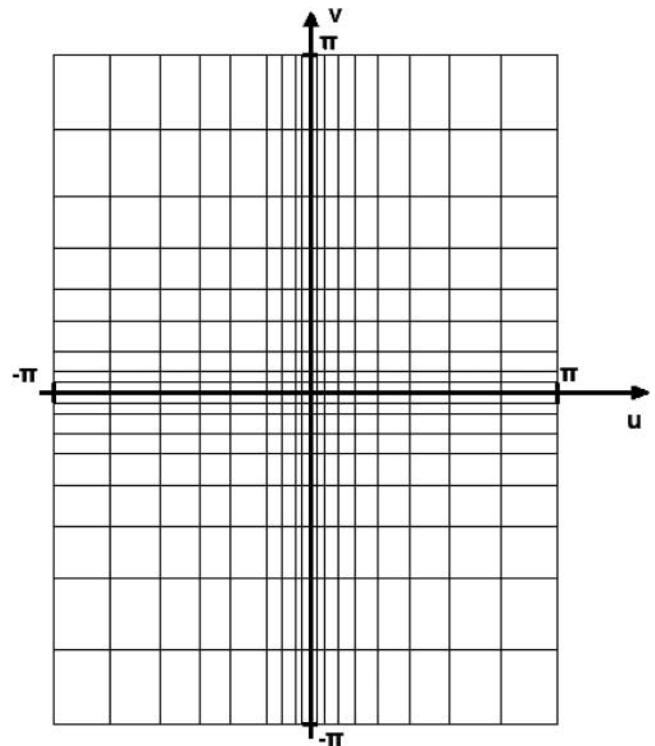

FIGURE 1. A representative 2D mel-cepstrum grid in the DTFT domain. Cell sizes are smaller at low frequencies compared with high frequencies.
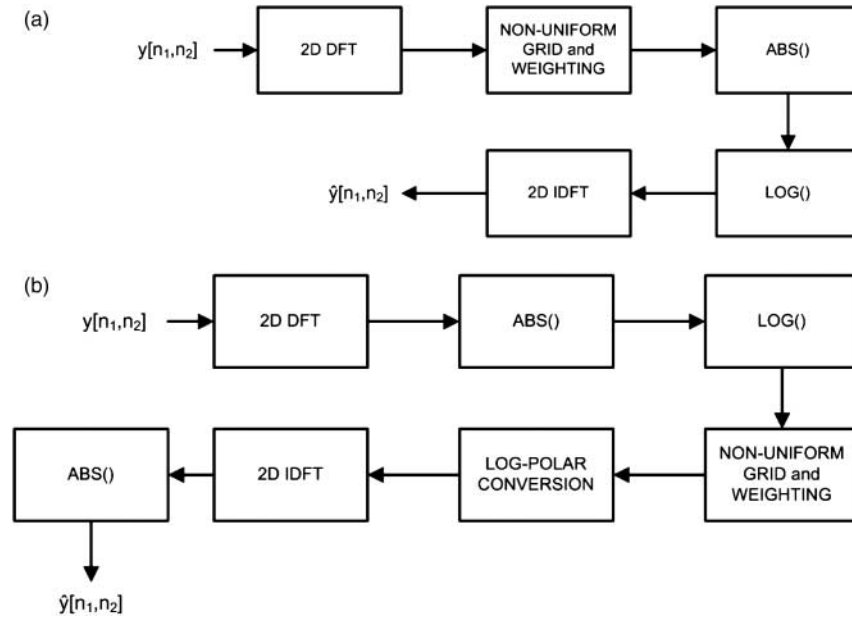
FIGURE 2. Flow diagram of cepstral feature extraction methods. (a) 2D Mel-cepstrum-based feature extraction algorithm and (b) 2D Mellin-cepstrum-based feature extraction algorithm.

computation in speech processing. Similar to speech signals, most natural images including face images are low-pass in nature. Therefore, there is more signal energy at low frequencies compared with high frequencies. Logarithmic division of the DFT grid emphasizes high frequencies. After this step, 2D mel-frequency cepstral coefficients $\hat{y}_m(p, q)$ are computed, using either inverse DFT or DCT, as follows:

$$\hat{y}_m(p, q) = F_2^{-1}(\log(|G(m, n)|^2)) \qquad (3)$$

The size of the IDFT is smaller than the size of the forward DFT used to compute $Y(k, l)$ because of the logarithmic grid shown in Fig. 1. It is also possible to apply different weights to different bins to emphasize certain bands as in speech processing. Since several DFT values are grouped together in each cell, the resulting 2D mel-cepstrum sequence computed using the IDFT has smaller dimensions than the original image.

Steps of the 2D mel-cepstrum-based feature extraction scheme is summarized below:

- N-by-N 2D DFT of input images is calculated. The DFT size N should be larger than the image size. To take advantage of the FFT algorithm during DFT computation, it is better to select $N = 2^r$ such that $\min_r\{2^r > \max\{W, H\}\}$ where W and H are the width and the height of the input image, respectively.
- The non-uniform DTFT grid is applied to the resultant DFT matrix and the energy $|G(m, n)|^2$ of each cell is computed. Each cell of the grid can also be weighted with a coefficient. The new data size is M by M, where $M \leq N$.
- Logarithm of cell energies $|G(m, n)|^2$ are computed.

- 2D IDFT or 2D inverse discrete cosine transform of the M-by-M data is computed to get the M-by-M mel-cepstrum sequence.

The flow diagram of the 2D mel-cepstrum feature extraction technique is shown in Fig. 2a.

The 2D Mellin-cepstrum feature extraction technique is a modified version of the 2D mel-cepstrum algorithm. It takes advantage of the Mellin transform and provides rotation, scale and illumination invariant features [6]. 2D Mellin-cepstrum feature extraction algorithm is explained below and the flow diagram of this technique is shown in Fig. 2b. Fourier–Mellin features are rotation, scale and translation invariants. By taking the logarithm of Fourier domain magnitudes, it is possible to achieve an illumination invariance in cepstral domain. This is explained at the end of this section. Steps of 2D Mellin computation is summarized below:

- N-by-N 2D DFT of input images is calculated. The DFT size N should be larger than the image size. It is better to select $N = 2^r > \text{dimension}(y(n_1, n_2))$ to take advantage of the FFT algorithm during DFT computation.
- Logarithm of magnitudes of the DFT coefficients are computed.
- The non-uniform DFT grid is applied to the resultant matrix and the mean of each cell is computed. Each cell of the grid is represented with this mean and the cell is weighted with a coefficient. The new data size is M by M, where $M \leq N$.
- Cartesian to log-polar conversion is performed using bilinear interpolation. This is the key step of the FMT providing rotation and scale invariance.
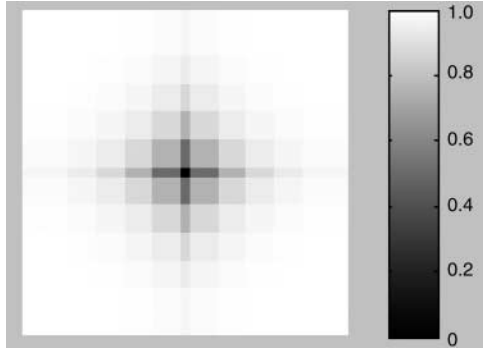
FIGURE 3. $M \times M$ normalized weights for emphasizing high frequencies. ($M = 35$ for this figure.)

- 2D IDFT of the M-by-M data is computed.
- Absolute value or energy of the IDFT coefficients is calculated to get the M-by-M Mellin-cepstrum sequence.

In a face image, edges and important facial features generally contribute to high frequencies. In order to extract better representative features, high-frequency component cells of the 2D DFT grid is multiplied with higher weights compared with low-frequency component bins in the grid. As a result, high-frequency components are further emphasized. In order to emphasize the high-frequency component cells of the 2D DFT grid further, the $M \times M$ normalized weights are organized as in Fig. 3. In Fig. 3, white values corresponds to 1 and black values corresponds to 0. The smallest value used in Fig. 3 is 0.005. In this paper, the weights are organized as linear weights. To select appropriate weights corresponding to certain frequency cells that contain more discriminative power, a research is still going on about an automatic weight selection algorithm.

Invariance of cepstrum to the pixel amplitude changes is an important feature. In this way, it is possible to achieve robustness to illumination invariance. Let $Y(u, v)$ denote the 2D DTFT of a given image matrix $y(n_1, n_2)$ and $cy(n_1, n_2)$ has a DTFT $cY(u, V)$ for any real constant c. The log spectrum of $cY(u, V)$ is given as follows:

$$\log(|cY(u, v)|) = \log(|c|) + \log(|Y(u, v)|) \qquad (4)$$

and the corresponding cepstrum is given as follows:

$$\psi(p, q) = \hat{a}\delta(p, q) + \hat{y}(p, q)$$

where

$$\delta(p, q) = \begin{cases} 1 & p = q = 0 \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

Therefore, the cepstrum values except at $(0, 0)$ location (DC term) do not vary with the changes in amplitude. Since FT magnitudes of $y(n_1, n_2)$ and $y(n_1 - k_1, n_2 - k_2)$ are the same, the 2D cepstrum and mel-cepstrum are shift invariant features.
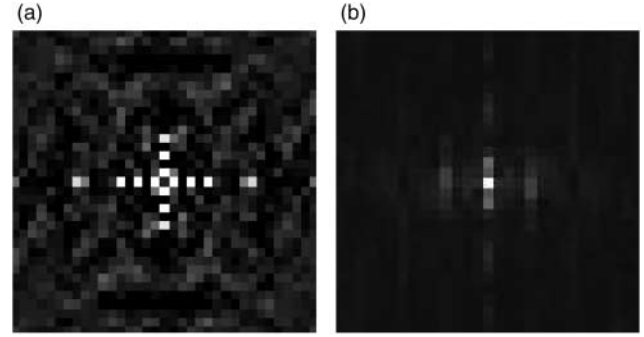

FIGURE 4. Illustration of cepstral domain features. (a) Magnitude of 35-by-35 2D mel-cepstrum of a face and (b) 35-by-35 2D Mellin-cepstrum of the face image matrix.

Another important characteristic of 2D cepstrum is symmetry with respect to $\hat{y}[n_1, n_2] = \hat{y}[-n_1, -n_2]$. As a result, only a half of the 2D cepstrum or $M \times M$ 2D mel-cepstrum coefficients are enough when IDFT is used.

In this paper, the dimensions of the 2D mel-cepstrum and Mellin-cepstrum matrices are selected as 49, 39, 35 and 29 to represent various size face images. A 35-by-35 2D mel-cepstrum of a face image is displayed in Fig. 4a and a 35-by-35 2D Mellin-cepstrum of the same face image is displayed in Fig. 4b. The symmetric structure of these cepstral domain features can be observed in the figure.

3. FEATURE CLASSIFICATION

In this article, CMA and multi-class SVM are used in feature classification. The CMA directly uses feature matrices as input. On the other hand, the SVM needs a matrix to vector conversion process to convert the 2D cepstral domain feature matrices to vectors. In the next section, the CMA method is described. In Section 3.2, the SVM method is described.

3.1. Common matrix approach

The CMA is a 2D extension of common vector approach (CVA), which is a subspace-based pattern recognition method [19]. The CVA was successfully used in finite vocabulary speech recognition [21]. In this article, the CMA is used as a classification engine. In order to train the CMA, common matrices belonging to each subject (class) are computed. In an image data set, there are C classes that contain p face images. Let y_i^c denote the ith image matrix belonging to the class c. The common matrix calculation process starts with selecting a reference image for each class. Then, the reference images are subtracted from the remaining $p - 1$ images of each subject. After the subtraction, the remaining matrices of each class are orthogonalized by using Gram–Schmidt orthogonalization. The orthogonalized matrices are orthonormalized by dividing each matrix to its frobenius norm. These orthonormalized

matrices span different subspace of the corresponding class. Let B_i^c denote the orthonormal-based matrices belonging to class c, where $i = 1, 2, \ldots, p - 1$ and $c = 1, 2, \ldots, C$. Any image matrix y_i^c belonging to class c can be projected onto the corresponding different subspaces in order to calculate difference matrices. The difference matrices are determined as follows:

$$y_{\text{diff},i}^c = \sum_{s=1}^{p-1} \langle y_i^c, B_s^c \rangle B_s^c \qquad (6)$$

Next, common matrices are calculated for each image class:

$$y_{\text{com}}^c = y_i^c - y_{\text{diff},i}^c \qquad (7)$$

In the test part of the CMA algorithm, the test image T is projected onto the difference subspaces of each class, then the projection is subtracted from the test image matrix.

$$
\begin{aligned}
D_1 &= T - \sum_{s=1}^{p-1} \langle T, B_s^1 \rangle B_s^1 \\
D_2 &= T - \sum_{s=1}^{p-1} \langle T, B_s^2 \rangle B_s^2 \\
&\quad\cdot \\
&\quad\cdot \\
D_C &= T - \sum_{s=1}^{p-1} \langle T, B_s^C \rangle B_s^C
\end{aligned} \qquad (8)
$$

The test image T is assigned to the class c where the distance $\|D_c - y_{\text{com}}^c\|^2$ is minimum.

3.2. Multi-class SVM

The SVM is a supervised machine-learning method, developed by Vladimir Vapnik [22], based on the statistical learning theory. The method constructs a hyperplane or a set of hyperplanes in a high-dimensional space that can be used in classification tasks. In this work, the SVM with a multi-class classification support [23] with radial basis function kernel is used. The multi-class classification method uses 'one-against-one' strategy [24].

2D cepstral domain feature matrices are converted to vectors in a raster scan approach before training and classification. The raster scan starts with $\hat{y}_m(0, 0)$. If there is pixel intensity variations, $\hat{y}_m(0, 0)$ is ignored and the scan starts from $\hat{y}_m(0, 1)$.

4. EXPERIMENTAL RESULTS

4.1. Database

In this paper, AR Face Image Database [25], ORL Face Database [26], Yale Face Database [27] and FRGC Version 2 database [28] are used in experimental studies.

The AR face database, created by Aleix Martinez and Robert Benavente, contains 4000 facial images of 126 subjects. Seventy of these subjects are male and the remaining 56 subjects are female. Each subject has different poses, including different facial expressions, illumination conditions and occlusions (sun glasses and scarf). The face images are all in the dimensions of 768-by-576 pixels. In this work, 14 non-occluded poses of 50 subjects are used. Face images are converted to gray scale, normalized and manually cropped to have a size of 100×85. Then, the cropped faces are aligned. Some data sets may include badly aligned face images. In this case, a simple face detector, i.e. [29], can be implemented and recognition process is performed after face detection. Sample poses for randomly selected five subjects from the AR face database are shown in Fig. 5.

The second database used in this work is ORL face database. The ORL database contains 40 subject and each subject has 10 poses. The images are captured at different time periods, different lighting conditions and different accessories for some of the subjects. In this work, nine poses of each subject are used. In ORL face database, images are all in gray scale with dimensions of 112×92. Sample images from the ORL face database are shown in Fig. 6. The third database used in this work is the Yale Face Database. The Yale database contains gray-scale facial images with sizes of 152×126. The database contains 165 facial images belonging to 15 subjects. Each pose of the subjects has different facial expressions and illuminations. Sample images from Yale database are shown in Fig. 7.

The FRGC Version 2 database [28] contains 12 776 images belonging to 222 subjects. In our experiments, the image set previously used in Experiment 1 [28] is used. The subset images of Experiment 1 are taken in a controlled environment

FIGURE 5. Sample images from the AR Face Database.

FIGURE 6. Sample images from the ORL Face Database.

FIGURE 7. Sample images from the Yale Face Database.

FIGURE 8. Sample images from Experiment 1 subset of FRGC Version 2 Database.

under different illumination and facial expressions. The data set contains 16 028 images of 225 subjects and the number of poses for each person varies between 32 and 88. In order to have a data set with equal number of poses for each subject, 32 poses of each subject are randomly selected. At the end, we have a subset containing 7200 facial images. The images are cropped and resized to 50×50 using a simple face detector algorithm [29]. Sample images of FRGC Version 2 database are displayed in Fig. 8.

All of the subjects in ORL, YALE and FRGC version 2 face databases are used in the experiments; however, in the AR face database, 50 of 126 subjects are randomly selected. In [30, 31], the authors also used 50 of 126 subjects. To have a fair comparison with [30, 31], we follow the same strategy with [30, 31].

4.2. Procedure and experimental work

In order to compare performances of various features, the proposed 2D cepstral domain features, 2D FMT-based features, actual image pixel matrices and 2D PCA-based features are applied to CMA and multi-class SVM as inputs.

In order to achieve robustness in recognition results, leave-one-out procedure is used. Let p denote the number of poses for each person in a database. In the test part of the classifier, one pose of each person is used for testing. Remaining $p - 1$ poses for each person are used in the training part of the classifier. In the leave-one-out procedure, the test pose is changed in each turn and the algorithm is trained with the new $p-1$ images. At the end, a final recognition rate (RR) is obtained by averaging the RRs for each selection of a test pose.

In the calculation of 2D cepstral domain features, different non-uniform grids are used. Due to these different non-uniform grids, quefrency coefficients in each bin are grouped together in different numbers. Therefore, different M by M 2D cepstrum based features are generated ($M = 49, 39, 35, 29$). In Tables 1 and 2, the cepstral features obtained using four different non-uniform grids are denoted with NG_i where $i = 1, 2, 3, 4$. The highest RR achieved for a database is indicated with bold in Tables 1 and 2.

4.2.1. AR face database
In Tables 1 and 2, average RRs of each leave-one-out step is given for each classifier when different 2D cepstrum-based features are used. The CMA produces slightly higher results than the SVM-based classifier. The proposed 2D cepstrum-based features outperform the PCA-, FMT- and actual image pixel value-based features, as shown in Tables 1 and 2 in all the four databases including the AR face database.

4.2.2. ORL face database
In the ORL face database, recognition experiments were repeated using different DTFT domain grids in the calculation of 2D cepstrum-based features. These features are applied to the CMA and SVM. The RRs corresponding to each classifier are shown in Tables 1 and 2.

In both ORL and AR face image databases, the CMA provides slightly better results than the SVM-based classifier.

4.2.3. Yale face database
The Yale face database contains images captured under different illumination conditions. Since there exist illumination changes

TABLE 1. RRs of CMA classifier with different databases and feature sets. NG stands for the non-uniform grid used in the cepstral feature computation.

Face databases	RR and feature sizes	Features			Proposed 2D mel-cepstrum				Proposed 2D Mellin-cepstrum			
		Raw images	2D PCA	2D FMT	NG_1	NG_2	NG_3	NG_4	NG_1	NG_2	NG_3	NG_4
AR	RR (%)	97.42	97.71	98.28	98.85	99.00	97.57	97.00	**99.28**	99.14	99.00	98.57
	Feature size	100×85	100×12	60×60	25×49	20×39	18×35	15×29	49×49	39×39	35×35	29×29
ORL	RR (%)	98.33	98.33	98.61	98.61	98.61	99.44	98.61	98.88	99.16	99.44	**100**
	Feature size	112×92	112×15	60×60	25×49	20×39	18×35	15×29	49×49	39×39	35×35	29×29
YALE	RR (%)	71.52	71.52	73.33	**77.57**	**77.57**	76.36	75.15	**77.57**	**77.57**	**77.57**	76.96
	Feature size	152×126	152×9	60×60	25×49	20×39	18×35	15×29	49×49	39×39	35×35	29×29
FRGC	RR (%)	92.58	93.22	93.80	96.34	95.90	95.23	95.58	**96.75**	96.50	96.43	96.50
	Feature size	50×50	50×6	60×60	25×49	20×39	18×35	15×29	49×49	39×39	35×35	29×29

TABLE 2. RRs of SVM-based classifier with different databases and feature sets. NG stands for non-uniform grid used in the cepstral feature computation.

Face databases	RR and feature sizes	Features			Proposed 2D mel-cepstrum				Proposed 2D Mellin-cepstrum			
		Raw images	2D PCA	2D FMT	NG_1	NG_2	NG_3	NG_4	NG_1	NG_2	NG_3	NG_4
AR	RR (%)	96.85	96.85	97.85	98.71	98.71	98.71	98.42	98.71	98.71	**98.85**	98.42
	Feature size	8500	1200	3600	1225	780	630	435	2401	1521	1225	841
ORL	RR (%)	98.05	98.33	98.61	98.05	98.61	98.61	99.16	98.05	98.88	98.88	**99.44**
	Feature size	10 304	1680	3600	1225	780	630	435	2401	1521	1225	841
YALE	RR (%)	88.00	87.87	90.90	**98.18**	96.96	96.96	96.96	95.75	96.96	95.75	95.15
	Feature size	19 152	1368	3600	1225	780	630	435	2401	1521	1225	841
FRGC	RR (%)	93.22	93.67	93.80	**96.18**	95.58	**96.18**	93.80	93.67	94.87	**96.18**	94.63
	Feature size	2500	300	3600	1225	780	630	435	2401	1521	1225	841

in the Yale face database, we simply set the (0,0) value of the 2D mel-cepstrum $\hat{y}(0, 0) = 1$ in all cases to normalize illumination changes. As a result, the average RR of the Yale face database increased from 74.54 to 77.57% when CMA-based classifier is used. Similarly, we observed an increase in RRs when multi-class SVM-based classifier is used by setting $\hat{y}(0, 0) = 1$; the average RR increased from 94.54 to 98.18% in Yale face database as a result of cepstral normalization.

The performance of different-sized 2D cepstral features using Yale face database images are presented in Table 1 for CMA and in Table 2 for multi-class SVM. In this database, the SVM classifier significantly outperforms the CMA classifier, which has a low recognition rate of 77.57% since CMA cannot cope with large illumination changes.

4.2.4. FRGC Version 2 Database

The recognition experiments were repeated using 'Experiment 1' subset of the FRGC Version 2 database. The average RR of each leave-one-out step is calculated when different features and classifiers are used. According to the results listed in Tables 1 and 2, cepstral features outperform classical feature extraction baselines such as 2D PCA and Mellin transform.

Based on the RRs presented in Tables 1 and 2, NG_1 provides better representative features in the AR Face database. NG_4 provides an increase in RR when the ORL face database is used. In the Yale face database, cepstral features extracted using NG_1 and NG_2 provides better performance than other non-uniform grids. In each face database, edges and important features may lie in different frequency bands. Since different non-uniform grids combine the frequency coefficients in different manner,

the RRs due to these non-uniform grids differ. Therefore, several non-uniform grids are studied in the experiments in order to determine the appropriate frequency bands for each face database.

The 2D PCA- and 2D FMT-based features do not provide better results than the proposed cepstrum-based features. Moreover, highest recognition results are obtained using the 2D Mellin-cepstrum.

According to the results given in Tables 1 and 2, the 2D PCA feature extraction method seems to be slightly increased the RR, but this technique is included in the paper since it is a baseline method that provides dimensional reduction.

The computational complexity of 2D PCA-based features are higher than 2D mel-cepstrum-based features which are computed using FFT. The cost of computing a 2D mel-cepstrum sequence for an N-by-N image is $O(N^2 \log(N) + M^2 \log(M))$ and an additional $M^2/2$ logarithm computations which can be implemented using a look-up table. 2D Mellin-cepstrum requires an additional log-polar conversion step in the Fourier domain; therefore, its cost is lower than PCA but higher than the FMT because of the additional logarithm computation in the Fourier domain.

5. CONCLUSION

In this article, 2D mel-cepstrum- and Mellin-cepstrum-based feature extraction techniques are proposed for image representation. Illumination invariance and invariance to translational shifts are important properties of 2D mel-cepstrum and 2D cepstrum. In addition, 2D Mellin-cepstrum provides robustness against rotation and scale invariance.

2D cepstral domain features extraction techniques provide not only better recognition rates but also dimensionality reduction in feature matrix sizes in the face recognition problem. Our experimental studies indicate that 2D cepstral methods are superior to classical feature extraction baseline methods in facial image representation with lower computational complexity.

FUNDING

REFERENCES

[1] Toreyin, B.U. and Cetin, A.E. (2009) Shadow detection using 2d cepstrum. In *Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIII*, Orlando, FL, USA, 733809. SPIE.

[2] Yeshurun, Y. and Schwartz, E. (1989) Cepstral filtering on a columnar image architecture: a fast algorithm for binocular stereo segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, **11**, 759–767.

[3] Lee, J.K., Kabrisky, M., Oxley, M.E., Rogers, S.K. and Ruck, D.W. (1993) The complex cepstrum applied to two-dimensional images. *Pattern Recognit.*, **26**, 1579–1592.

[4] Çetin, A.E. and Ansari, R. (1988) Convolution-based framework for signal recovery and applications. *J. Opt. Soc. Am. A*, **5**, 1193–1200.

[5] Bertrand, J., Bertrand, P. and Ovarlez, J. (1990) Discrete Mellin transform for signal analysis. In *IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP)*, Albuquerque, New Mexico, USA, April, vol. 3, pp. 1603–1606.

[6] Gueham, M., Bouridane, A., Crookes, D. and Nibouche, O. (2008) Automatic recognition of shoeprints using Fourier–Mellin transform. In *NASA/ESA Conference on Adaptive Hardware and Systems*, June, pp. 487–491.

[7] Zhao, W., Chellappa, R., Rosenfeld, A. and Phillips, P.J. (2003) Face recognition: a literature survey. *ACM Comput. Surv.*, **35**, 399–458.

[8] Brunelli, R. and Poggio, T. (1993) Face recognition: features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, **15**, 1042–1052.

[9] Etemad, K. and Chellappa, R. (1997) Discriminant analysis for recognition of human face images. *J. Opt. Soc. Am. A*, **14**, 1724–1733.

[10] Chen, L.-F., Liao, H.-Y.M., Ko, M.-T., Lin, J.-C. and Yu, G.-J. (2000) A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognit.*, **33**, 1713–1726.

[11] Martinez, A. and Kak, A. (2001) PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**, 228–233.

[12] Lu, J., Plataniotis, K.N. and Venetsanopoulos, A.N. (2003) Regularized discriminant analysis for the small sample size problem in face recognition. *Pattern Recognit. Lett.*, **24**, 3079–3087.

[13] Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J. (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, **19**, 711–720.

[14] Yu, H. (2001) A direct LDA algorithm for high-dimensional data—with application to face recognition. *Pattern Recognit.*, **34**, 2067–2070.

[15] Turk, M. and Pentland, A. (1991) Eigenfaces for recognition. *J. Cogn. Neurosci.*, **3**, 71–86.

[16] Qin, J. and He, Z.-S. (2005) A SVM face recognition method based on Gabor-featured key points. In *Proc. Int. Conf. Machine Learning and Cybernetics*, Guangzhou, China, August, pp. 5144–5149.

[17] Dai, G. and Zhou, C. (2003) Face recognition using support vector machines with the robust feature. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, Milbrae, California, USA, November, pp. 49–53.

[18] Guo, G., Li, S.Z. and Chan, K. (2000) Face recognition by support vector machines. In *Proc. Fourth IEEE Int. Conf. Automatic Face and Gesture Recognition*, Washington, DC, USA 196. IEEE Computer Society.

[19] Gulmezoglu, M.B., Dzhafarov, V. and Barkana, A. (2001) The common vector approach and its relation to principal component analysis. *IEEE Trans. Speech Audio Process.*, **9**, 655–662.

[20] Oppenheim, A.V., Schafer, R.W. and Buck, J.R. (1999) *Discrete-time Signal Processing* (2nd edn). Prentice-Hall, Upper Saddle River, NJ, USA.

[21] Gulmezoglu, M.B., Dzhafarov, V., Keskin, M. and Barkana, A. (1999) A novel approach to isolated word recognition. *IEEE Trans. Speech Audio Process.*, **7**, 620–628.

[22] Boser, B.E., Guyon, I.M. and Vapnik, V.N. (1992) A training algorithm for optimal margin classifiers. In *COLT '92: Proc. Fifth Annual Workshop on Computational Learning Theory*, New York, NY, USA, pp. 144–152. ACM.

[23] Chang, C.C. and Lin, C.J. (2001) *LIBSVM*: a library for support vector machines.

[24] Knerr, S., Personnaz, L. and Dreyfus, G. (1990) Single-layer learning revisited: a stepwise procedure for building and training a neural network. In Fogelman, J. (ed.), *Neurocomputing: Algorithms, Architectures and Applications*. Springer.

[25] Martinez, A. and Benavente, R. (1998) The AR Face Database. CVC Tech. Report No. 24.

[26] Samaria, F.S. and Harter, A.C. (1994) Parameterisation of a stochastic model for human face identification. In *IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, USA, August, pp. 138–142.

[27] Yale (1997) *Yale Face Database.* http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[28] Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J. and Worek, W. (2005) Overview of the face recognition grand challenge. In *Proc. 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, USA, pp. 947–954. IEEE Computer Society.

[29] Nilsson, M., Nordberg, J. and Claesson, I. (2007) Face detection using local SMQT features and split up snow classifier. In *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawai'i, USA, pp. 15–20, II-589–II-592.

[30] Cevikalp, H., Member, S., Neamtu, M., Wilkes, M. and Barkana, A. (2005) Discriminative common vectors for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**, 4–13.

[31] Çiğdem Turhal, U., Gülmezoğlu, M.B. and Barkana, A. (2005) Face recognition using common matrix approach. In *Proc. 13th European Signal Processing Conf.*

[32] Bogert, B., Healy, M. and Tukey, J. (1963) The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In *Proc. Symp. Time Series Analysis*, pp. 209–243. J. Wiley and Sons Inc., Newyork.