

## Image Feature Extraction Using 2D Mel-Cepstrum

Serdar ÇAKIR, A. Enis ÇETİN\*

*Department of Electrical and Electronics Engineering  
Bilkent University, 06800, Ankara, Turkey  
{cakir,cetin}@bilkent.edu.tr*

### Abstract

*In this paper, a feature extraction method based on two-dimensional (2D) mel-cepstrum is introduced. Feature matrices resulting from the 2D mel-cepstrum, Fourier LDA approach and original image matrices are individually applied to the Common Matrix Approach (CMA) based face recognition system. For each of these feature extraction methods, recognition rates are obtained in the AR face database, ORL database and Yale database. Experimental results indicate that recognition rates obtained by the 2D mel-cepstrum method is superior to the recognition rates obtained using Fourier LDA approach and raw image matrices. This indicates that 2D mel-cepstral analysis can be used in image feature extraction problems.*

### 1. Introduction

Mel-cepstral analysis is one of the most widely used 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, 10, 16, 6]. To the best of our knowledge 2-D mel-cepstrum which is a variant of 2D cepstrum is not used in image feature extraction, classification and recognition problems. The goal of this paper is to define the 2-D mel-cepstrum and show that it is a viable image representation tool. 2D cepstrum is a queffreny domain method and it is computed using 2D FFT. As a result it is a computationally efficient method. It is also independent of pixel amplitude variations and translational shifts. 2D mel-cepstrum which is based on logarithmic decomposition of frequency domain also has the same shift and amplitude invariance properties as 2D cepstrum.

\*This work is supported by European Commission Seventh Framework Program with EU Grant: 244088(FIRESENSE)

In this article, the 2D mel-cepstrum based feature extraction method is applied to the face recognition problem. It should be pointed out that our aim is not the development of a complete face recognition system but to illustrate the advantages of the 2-D mel-cepstrum. 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 the 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 [18, 5]. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are well known techniques that were used in face recognition [7, 13]. 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 illuminations changes in the face images. LDA with some extensions such as quadratic LDA [11], Fisher's LDA [2], and direct, exact LDA [17] were proposed. LDA is also proposed as a Fourier domain application in order to select appropriate Fourier frequency bands for the problem of image recognition [9]. In 2D mel-cepstrum, the logarithmic division of the 2D DFT grid provides the dimensionality reduction. This is also an intuitively valid representation as most natural images are low-pass in nature. Unlike the Fourier or DCT domain features high-frequency DFT and DCT coefficients are not discarded in an ad-hoc manner. They are simply combined in bins in a logarithmic manner during mel-cepstrum computation.

The rest of the paper is organized as follows. In Section 2, proposed 2D mel-cepstrum based feature extraction method is described. In Section 3, a subspace based pattern recognition method called the Common Matrix Approach (CMA) is explained. The 2D mel-cepstrum matrices obtained from facial images are classified using the CMA which is successfully used in a face recog-

nition application [8]. In Section 4, experimental results are presented.

## 2. The 2D Mel-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, 10, 16]. In this article, 2D mel-cepstrum is used for representing face images.

2D cepstrum  $\hat{y}(p, q)$  of a 2D image  $y(n_1, n_2)$  is defined as follows

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

where  $(p, q)$  denotes 2D cepstral quefrequency coordinates,  $F_2^{-1}$  denotes 2D Inverse Discrete-Time Fourier transform (IDTFT) and  $Y(u, v)$  is the 2D Discrete-Time Fourier transform (DTFT) of the image  $y(n_1, n_2)$ . In practice, Fast Fourier Transform (FFT) algorithm is used to compute DTFT.

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

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

where  $B(m, n)$  is the  $(m, n)$ -th cell of the logarithmic grid. Cell or bin sizes are smaller at low frequencies compared to high-frequencies. This approach is similar to the mel-cepstrum 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 to 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

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

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 face images are 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.

- The non-uniform DTFT grid is applied to the resultant DFT matrix and energy  $|G(m, n)|^2$  of each cell is computed. Each cell of the grid is also 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 IDCT of the  $M$  by  $M$  data is computed to get the  $M$  by  $M$  mel-cepstrum sequence.

The flow diagram of the 2D cepstrum feature extraction technique is given in Figure 1. In a face image, edges

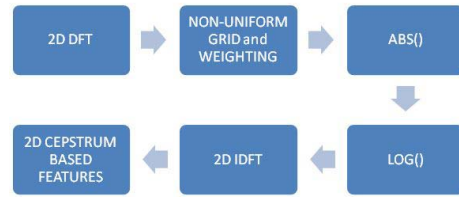


Figure 1. 2D Cepstrum Based Feature Extraction Algorithm.

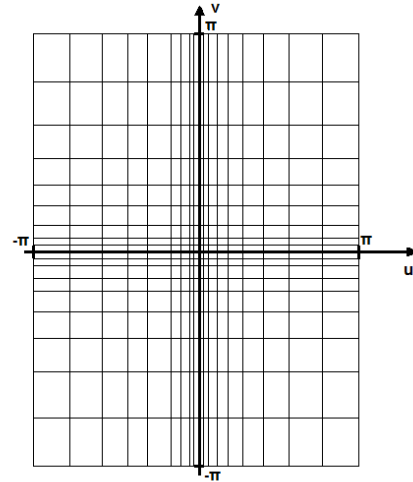


Figure 2. A representative 2D mel-cepstrum Grid in the DTFT domain. Cell sizes are smaller at low frequencies compared to high frequencies.

and 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 to low frequency component bins in the grid. As a result, high frequency components are further emphasized.

Invariance of cepstrum to the pixel amplitude changes is an important feature. 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

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

and the corresponding cepstrum is given as follows

$$\psi(p, q) = \hat{a}\delta(p, q) + \hat{y}(p, q) \quad (5)$$

where  $\delta(p, q) = 1$  for  $p = q = 0$  and  $\delta(p, q) = 0$  otherwise. Therefore, the cepstrum values except at  $(0, 0)$  location (DC Term) do not vary with the amplitude changes. Since the Fourier Transform 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.

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 2-D cepstrum or MxM 2-D mel-cepstrum coefficients are enough when IDFT is used.

### 3. Common Matrix Approach

The Common Matrix Approach (CMA) is a 2D extension of Common Vector Approach (CVA), which is a subspace based pattern recognition method [8]. The CVA was successfully used in finite vocabulary speech recognition [4]. The CMA is used as a classification engine in this article. In order to train the CMA, common matrices belonging to each subject (class) are computed. In an image dataset, there are  $C$  classes that contain  $p$  face images. Let  $y_i^c$  denote the  $i^{th}$  image matrix belonging to the class  $c$ . The 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 the difference subspace of the corresponding class. Let  $B_i^c$  denote the orthonormal basis matrices belonging to class  $c$  where  $i = 1, 2, \dots, p-1$  and  $c = 1, 2, \dots, C$ . Any image matrix  $y_i^c$  belonging to class  $c$  can be projected onto the corresponding difference subspaces in order to calculate difference matrices.

The difference matrices are determined as follows

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

Next, common matrices are calculated for each image class:

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

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

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

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

## 4. Experimental Results

### 4.1 Database

In this paper, AR Face Database [12], ORL Face Database [14] and Yale Face Database [15] are used. AR face database contains 4000 facial images of 126 subjects. In this work, 14 non-occluded poses of 50 subjects are used. In experimental work, images are converted to gray scale and cropped to have a size of 100x85. ORL database contains 40 subject and each of the subjects has 10 poses. In this work 9 poses of each subject are used. In ORL face database, the images are all in gray scale with dimensions of 112x92. Yale database contains gray scale facial images with the sizes of 152x126. The database contains 165 facial images belonging to 15 subjects.

### 4.2 Procedure and Experimental Work

In order to compare performances of various features, 2D mel-cepstrum based feature matrices, raw image matrices, and Fourier LDA based feature matrices are applied to CMA as inputs.

In order to achieve robustness in recognition results, leave-one-out procedure is used. Let  $k$  denote number of poses for each person in a database. In the test part of the CMA, one pose of each person is used for testing. Remaining  $k-1$  poses for each person are used in the training part of the CMA. In the leave-one-out procedure, the test pose is changed in each turn and the

algorithm is trained with the new k-1 images. At the end, a final recognition rate is obtained by averaging the recognition rates for each selection of test pose.

In the Table 1, average recognition rates of each leave-one-out step is given for the three feature extraction methods in each database.

**Table 1. Recognition Rates (RR)**

Features	Face Databases					
	AR		ORL		YALE	
	RR	Size	RR	Size	RR	Size
Original Images	97.42%	100 × 85	98.33%	112 × 92	71.52%	152 × 126
Fourier LDA	97.42%	100 × 10	98.88%	112 × 10	73.33%	152 × 10
2D mel-cepstrum	99%	18 × 35	100%	18 × 35	74.54%	18 × 35

Based on the above experiments, the Fourier LDA based features do not provide better results than the proposed 2D mel-cepstrum features. 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.

## 5. Conclusion

In this article, a 2D mel-cepstrum based feature extraction technique is proposed for image representation. Invariance to amplitude changes and translational shifts are important properties of 2D mel-cepstrum and 2D cepstrum. 2D mel-cepstrum based features provide not only better recognition rates due to their robustness to illumination changes but also dimensionality reduction in feature matrix sizes in the face recognition problem. Our experimental studies indicate that 2D mel-cepstrum method is superior to classical feature extraction baseline methods in image representation and in terms of computational complexity. On the other hand, 2D mel-cepstrum features are not robust to rotational changes and scaling. One possible solution is the use of Fourier-Mellin transform before computing the cepstral features [3]. This will lead to robustness to both rotation and scale changes.

## References

[1] B. Ugur Toreyin, A. Enis Cetin. Shadow detection using 2D cepstrum. *Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIII*, 7338(1):733809, 2009.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern*

*Analysis and Machine Intelligence*, 19(7):711–720, August 1997.

[3] J. Bertrand, P. Bertrand, and J. Ovarlez. Discrete mellin transform for signal analysis. In *IEEE T Acoust. Speech, 1990. ICASSP-90.*, pages 1603–1606 vol.3, apr 1990.

[4] M. Bilginer Gulmezoglu, V. Dzhafarov, M. Keskin, and A. Barkana. A novel approach to isolated word recognition. *Speech and Audio Processing, IEEE Transactions on*, 7(6):620–628, Nov 1999.

[5] R. Brunelli and T. Poggio. Face recognition: features versus templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(10):1042–1052, 1993.

[6] A. E. Çetin and R. Ansari. Convolution-based framework for signal recovery and applications. *J. Opt. Soc. Am. A*, 5(8):1193–1200, 1988.

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

[8] M. Gulmezoglu, V. Dzhafarov, and A. Barkana. The common vector approach and its relation to principal component analysis. *Speech and Audio Processing, IEEE Transactions on*, 9(6):655–662, Sep 2001.

[9] X.-Y. Jing, Y.-Y. Tang, and D. Zhang. A Fourier-LDA approach for image recognition. *Pattern Recognition*, 38(3):453 – 457, 2005.

[10] J. K. Lee, M. Kabrisky, M. E. Oxley, S. K. Rogers, and D. W. Ruck. The complex cepstrum applied to two-dimensional images. *Pattern Recognition*, 26(10):1579 – 1592, 1993.

[11] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Regularized discriminant analysis for the small sample size problem in face recognition. *Pattern Recogn. Lett.*, 24(16):3079–3087, 2003.

[12] A. Martinez and R. Benavente. The AR face database. CVC Tech. Report # 24, 1998.

[13] A. Martinez and A. Kak. PCA versus LDA. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):228–233, Feb 2001.

[14] F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pages 138–142, August 2002.

[15] Yale. Yale Face Database. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>, 1997.

[16] Y. Yeshurun and E. Schwartz. Cepstral filtering on a columnar image architecture: a fast algorithm for binocular stereo segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):759–767, Jul 1989.

[17] H. Yu. A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 34(10):2067–2070, October 2001.

[18] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.