MEL-CEPSTRAL METHODS FOR IMAGE FEATURE EXTRACTION

Serdar ÇAKIR, A. Enis ÇETIN

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

ABSTRACT

A feature extraction method based on two-dimensional (2D) mel-cepstrum is introduced. The concept of one-dimensional (1D) mel-cepstrum which is widely used in speech recognition is extended to 2D in this article. Feature matrices resulting from the 2D mel-cepstrum, Fourier LDA, 2D PCA and original image matrices are converted to feature vectors and individually applied to a Support Vector Machine (SVM) classification engine for comparison. The AR face database, ORL database, Yale database and FRGC version 2 database are used in experimental studies, which indicate that recognition rates obtained by the 2D mel-cepstrum method is superior to the recognition rates obtained using Fourier LDA, 2D PCA and ordinary image matrix based face recognition. This indicates that 2D mel-cepstral analysis can be used in image feature extraction problems.

Index Terms— 2D mel-cepstrum, cepstral features, image feature extraction, face recognition

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, 2, 3, 4]. 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. Ordinary 2D cepstrum of a 2D signal is defined as the inverse Fourier Transform of the logarithmic spectrum of the signal 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 grid also has the same shift and amplitude invariance properties as the 2D cepstrum.

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 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 [5, 6]. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are well known techniques that were used in face recognition [7, 8]. 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 [9], Fisher's LDA [10], and direct, exact LDA [11] were proposed. LDA was also proposed to select appropriate frequency bands in the Fourier domain [12]. 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 of frequency values in a logarithmic manner during mel-cepstrum computation. The proposed feature extraction method outperform classical baseline PCA, since it does not eliminate any feature detail information by means of frequency values.

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, the well-known classification method SVM is briefly explained. The 2D mel-cepstrum matrices obtained from facial images are converted into vectors and classified using the SVM which was successfully used in face recognition applications [13, 14]. In Section 4, experimental results are presented.

This work is supported by European Commission Seventh Framework Program with EU Grant: 244088(FIRESENSE-Fire Detection and Management through a Multi-Sensor Network for the Protection of Cultural Heritage Areas from the Risk of Fire and Extreme Weather Conditions)

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, 2, 3]. 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

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

where (p,q) denotes 2D cepstral quefrency 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 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 $B(m,n)$ 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

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

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 $N = 2^r > dimension(y(n_1,n_2))$ 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 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 Fig. 2.

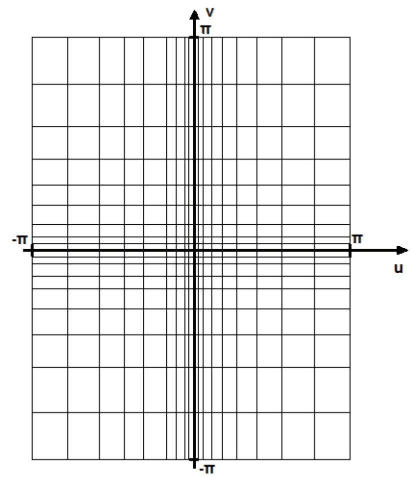

Fig. 1. A representative 2D mel-cepstrum Grid in the DTFT domain. Cell sizes are smaller at low frequencies compared to high frequencies.

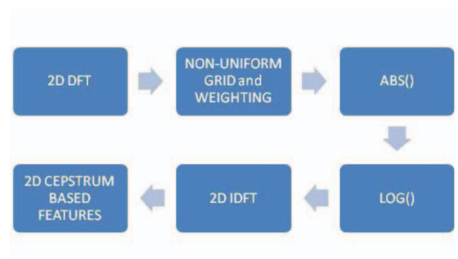

Fig. 2. 2D Cepstrum Based Feature Extraction Algorithm.

In a face image, edges 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. $cy(n_1,n_2)$ has a DTFT $cY(u,V)$ for any real constant c. The log spectrum of $cY(u,V)$ is given as follows

$$log(|cY(u,v)|) = log(|c|) + log(|Y(u,v)|) \tag{4}$$

and the corresponding cepstrum is given as follows

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

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.

Due to symmetry and shift invariance properties of DFT, 2D cepstrum and mel-cepstrum are also shift invariant and symmetric features. 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. SUPPORT VECTOR MACHINE BASED CLASSIFICATION

SVM is a supervised machine learning method based on the statistical learning theory and developed by Vladimir Vapnik [15]. 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, SVM with a multi class classification support namely C-SVC [16] with RBF kernel is used. The multi-class SVM uses "one-against-one" strategy [17]. In the experiments SVM parameters are set as "$cost = 1000$, $gamma = 0.008$", after performing a cross validation process.

4. EXPERIMENTAL RESULTS

4.1. Database

In this paper, AR Face Database [18], ORL Face Database [19] and Yale Face Database [20] and FRGC version 2 database [21] are used. AR face database contains 4000 facial images of 126 subjects. In this work, 14 non-occluded poses of 50 subjects are used. Images are converted to gray scale and cropped to have a size of 100x85. ORL database contains 40 subjects 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. The FRGC version 2 database [21] contains 12776 images belonging to 222 subjects. In our experiments, 32 controlled pose of each subject is randomly selected from the image subset previously used in the Experiment 1 [21]. The selected images are cropped by using a simple face detector algorithm [22] and the cropped images are resized to 50×50.

4.2. Procedure and Experimental Work

In order to compare performances of various features, 2D mel-cepstrum based feature matrices, actual image matrices, Fourier LDA and 2D PCA based feature matrices are converted into feature vectors and individually applied to SVM 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 SVM, one pose of each person is used for testing. Remaining K-1 poses for each person are used in the training part of the SVM. 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)

Databases	Features							
	Original Images		2D PCA		Fourier LDA		2D mel-cepstrum	
	RR	Size	RR	Size	RR	Size	RR	Size
AR	96.85%	8500	96.85%	1200	97.42%	1000	98.71%	630
ORL	98.05%	10304	98.33%	1680	98.88%	1120	98.61%	630
YALE	88.00%	19152	87.87%	1368	88.00%	1520	96.96%	630
FRGC v.2	93.22%	2500	93.80%	300	94.63%	500	96.18%	630

Based on the experimental results listed in Table 1, the Fourier LDA and the 2D PCA based features do not provide better results than the proposed 2D mel-cepstrum features. Moreover, the computational complexity of 2D PCA features are higher than 2D cepstrum based features which are computed using FFT. Recall that, K denote the number of poses for each person in a database. The computational cost of 2D PCA for an P by Q image is $(P^2Q)K + P^3 + SP^2$ where S denotes the number of eigenvectors that corresponds to largest eigenvalues in order to construct linear transformation matrix. The cost of computing a 2D mel-cepstrum sequence for an N by N image is $O(N^2log(N) + M^2log(M))$ and an additional $M^2/2$ logarithm computations which can be implemented using a look-up table where $N > (P,Q) > M$. It can be observed from the computations that the cost of the 2DPCA is clearly much more than 2D mel-cepstrum.

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 good recognition rates 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 method PCA in image representation and in terms of computational complexity.

6. REFERENCES

[1] B. U. Toreyin and A. E. Cetin, "Shadow detection using 2D cepstrum," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, May 2009, vol. 7338.

[2] James K. Lee, Matthew Kabrisky, Mark E. Oxley, Steven K. Rogers, and Dennis W. Ruck, "The complex cepstrum applied to two-dimensional images," *Pattern Recognition*, vol. 26, no. 10, pp. 1579 – 1592, 1993.

[3] Y. Yeshurun and E.L. Schwartz, "Cepstral filtering on a columnar image architecture: a fast algorithm for binocular stereo segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 759–767, Jul 1989.

[4] A. Enis Çetin and Rashid Ansari, "Convolution-based framework for signal recovery and applications," *J. Opt. Soc. Am. A*, vol. 5, no. 8, pp. 1193–1200, 1988.

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

[6] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 10, pp. 1042–1052, 1993.

[7] Li-Fen Chen, Hong-Yuan Mark Liao, Ming-Tat Ko, Ja-Chen Lin, and Gwo-Jong Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713 – 1726, 2000.

[8] A.M. Martinez and A.C. Kak, "PCA versus LDA," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 2, pp. 228–233, Feb 2001.

[9] Juwei Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recogn. Lett.*, vol. 24, no. 16, pp. 3079–3087, 2003.

[10] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, August 1997.

[11] Hua Yu, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067–2070, October 2001.

[12] Xiao-Yuan Jing, Yuan-Yan Tang, and David Zhang, "A Fourier-LDA approach for image recognition," *Pattern Recognition*, vol. 38, no. 3, pp. 453 – 457, 2005.

[13] Jun Qin and Zhong-Shi He, "A svm face recognition method based on gabor-featured key points," in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, Aug. 2005, vol. 8, pp. 5144–5149 Vol. 8.

[14] Guang Dai and Changle Zhou, "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*, Oct.-2 Nov. 2003, pp. 49–53.

[15] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik, "A training algorithm for optimal margin classifiers," in *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory*, New York, NY, USA, 1992, pp. 144–152, ACM.

[16] Chih C. Chang and Chih J. Lin, *LIBSVM: a library for support vector machines*, 2001.

[17] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing: Algorithms, Architectures and Applications*, J. Fogelman, Ed. 1990, Springer-Verlag.

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

[19] 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*, August 2002, pp. 138–142.

[20] Yale, "Yale Face Database," `http://cvc.yale.edu/projects/yalefaces/yalefaces.html`, 1997.

[21] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek, "Overview of the face recognition grand challenge," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 947–954.

[22] M. Nilsson, J. Nordberg, and I. Claesson, "Face detection using local smqt features and split up snow classifier," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 15-20 2007, vol. 2, pp. II–589 –II–592.