# Identification of Satellite Images Based on Mel Frequency Cepstral Coefficients

T. M. Talal, National Authority of Remote Sensing and Space Science (Egypt), Ayman El-Sayed Menofia University, Faculty of Electronic Engineering (Egypt)

Abstract—MFCC technique is an efficient technique which can be used for speech signals' classification as MFCC can be applied for 1-D signals. This paper suggests a new application for MFCC technique as it can be used for classification of satellite images, which are 2-D objects. Applying MFCC to images, through transforming images to 1-D vectors is an innovation in its own right. The used images for training and testing are acquired by high-resolution satellites such as IKONOS, QuickBird, and SPOT. This image set can be applied before and after roads extraction approach that is called morphological direction and length filters approach.

Keywords- classification, direction filter, identification,

## length filter, MFCC, morphology, road, and satellite images.

## I. INTRODUCTION

bject identification is one of the main applications of satellite images, which is very important for agriculture, water resources, geology, civilization planning, and weather forecasting...etc. Identification can be applied through two stages: feature extraction and classification [1, 2]. Feature extraction is the stage of assigning specific print for each feature or object by using a set of training data. In other word, feature extraction means removing all usefulness information that are associated with the feature or object. MFCC can be considered as one of the most common techniques that are used for feature extraction [2,3]. The main application of this technique is a 1-D speech or voice signals as it deals with frames of the data in 1-D. Classification is the stage of constructing a model for the system and feature matching for testing the performance of this model by applying a set of testing data. One of the most common techniques that can be used for classification stage is the Artificial Neural Networks (ANNs) [4-6].

This paper suggests a new application of MFCC theory as its applying to satellite images identification. The main criterion of images is its representation as a 2-D object such as gray-scale images or a 3-D object such as stereoscopic images. This criterion is not suitable for applying MFCC technique. So, transformation of 2-D matrix into 1-D vector is proposed in this paper. The rest of the identification process is common.

Training and testing image sets are normal high resolution satellites' images and segmented images after applying a road extraction approach that is called morphological direction and length filters (MDLFs) approach [7-9].

The rest of paper is organized as follows. Section 2 gives an overview of the structure of the identification system that will be used. Section 3 discusses the feature extraction stage. Classification stage is discussed in section 4. Section 5 gives the experimental results. Finally, section 6 includes conclusions and future work.

**II. IDENTIFICATION SYSTEM OF SATELLITE** 

## IMAGES

The identification consists of two stages to perform both the training of the input images models and the evaluation of the testing image sets, which are feature extraction and classification stages as shown in Fig. 1 [3].

If we consider first the feature extraction stage, we can find that this stage is concerned with converting an input training image into a series of vectors, which contain the discriminative information for the main image features. This is achieved by framing and windowing the input images after transformation of the signals from 2-D matrix into 1-D vector, then applying the feature extraction algorithm to each frame. The output of the feature extraction stage is the feature matrix which combines the feature vectors produced by all frames. Rows of this matrix correspond to the frame number while columns correspond to the feature vector coefficient [1, 2].

The classification process employs this feature matrix as an input to perform two processes. For model training, the classifier uses a collection of feature matrices from different training images to construct some type of model [1, 4, 5]. After that, testing of this model should be performed using test image set to verify the efficiency of the constructed model and reach to the optimum model allowing identification to take place.



#### Fig. 1 Identification system overview

## III. FEATURE EXTRACTION

There are a lot of techniques that can be used for extracting image features in the form of coefficients such as the linear prediction coefficients (LPCs), the Mel-Frequency Cepstral Coefficients (MFCCs) and the Linear Prediction Cepstral Coefficients (LPCCs) [2, 3]. One of the most common techniques that will be used in this paper is MFCC theory.

MFCC features are derived through cepstral analysis and warped according to the Mel-scale which emphasizes low frequency components over high frequency components [1, 3].

Extraction of Mel-Frequency Cepstral features is performed as follow: framing and windowing of the input image signals using a Hamming window technique, applying the Fourier Transform to each frame and taking the magnitude of this spectrum to a series of Mel-filter banks to perform the warping. The log of this Mel-spectrum is then processed finally by Inverse Discrete Cosine Transform (inverse DCT) to produce the cepstral coefficients as shown in Fig. 2 [3]. The Mel is a unit that is used for the measurement of the perceived frequency. The Mel-scale is therefore a mapping between the real frequency scale in Hz and the perceived frequency scale in Mels. So, The Mapping is virtually linear below 1 kHz and logarithmic above as given by [1]:

$$Mel(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)$$

(1)

where: Mel(f) gives the mel-scale frequency corresponding to actual frequency f.



Fig. 2 Standard static Mel-Frequency Cepstral Coefficients feature extraction

For this implementation, the first 13 coefficients are taken and then cepstral mean subtraction is performed to enhance the feature's robustness to channel noise. So, a 13 point feature vector is resulted for each frame.

The above description is the procedure for extraction of the standard static MFCC features; however, this feature may be enhanced by adding delta and acceleration coefficients to those static coefficients. These dynamic features can be obtained by deriving the first and second derivatives of the 13 static coefficients by regression techniques and concatenating the resulting coefficients to form a 39 element vector for each frame. The structure of the dynamic MFCC feature vector for each frame is [3]:

$$DMFCCVector = [c_1 \cdots c_{13}, d_1 \cdots d_{13}, a_1 \cdots a_{13}]$$
(2)

where:  $c_i$  is static cepstral coefficient,  $d_i$  is delta coefficient, and  $a_i$  is acceleration coefficient.

## IV. CLASSIFICATION

As mentioned previously that classification stage is composed of two processes: building a model that simulates the identification system and feature matching process that evaluates performance of the model by using a test set of images. In the modeling step, the image is recorded to the system using features that are extracted during training. When an unknown set of images arrive, a feature matching technique is applied to map the features from this set to the model. The combination of a model and a matching technique is called a classifier [1, 4].

There are a lot of classification techniques that can be used such as Markov Random Fields (MRFs), Gaussian Mixture Models (GMMs), Vector Quantization (VQ), and ANNs [4, 5].

In this paper, ANNs technique will be included because of its high efficiency and throughput. The main advantages of the neural network approach are as follow [6]:

> 1. Less number of training examples are required for a problem than other classification and pattern recognition approaches.

> 2. Capability of implementing more complex partitioning of feature space.

3. Capability of high performance parallelprocessing implementations.

The classification step in automatic image identification systems is in fact a feature matching process between the features of a new image and the features saved in the database. Neural Networks are widely used for feature matching. Multi-Layer Perceptrons (MLPs) consisting of an input layer, one or more hidden layers and an output layer can be used for this purpose. Fig. 3 [4] shows an MLP having an input layer, a single hidden layer and an output layer. A single neuron only of the output layer is shown for simplicity. This structure will be used for feature matching because it is suitable for the problem considered in this paper.

Each *neuron* in the neural network is characterized by two parameters: activation function and bias, and each *connection* between any two neurons is characterized by weight factor. In this paper, the neurons from the input and output layers have linear activation functions and hidden neurons have sigmoid activation function  $F(u) = 1/(1 + e^{-u})$ . Therefore, for an input vector **X**, the neural network output vector **Y** can be obtained according to the following matrix equation [4, 5]:

$$Y_o = W_2 * F(W_1 * X + B_1) + B_2$$
(3)

where  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are the weight matrices between input and the hidden layer and between the hidden layer and output layer, respectively, and  $\mathbf{B}_1$  and  $\mathbf{B}_2$  are bias matrices for the hidden and the output layer, respectively.



Training of the neural network means an adjustment of its weights using a training algorithm. This algorithm adapts weights by minimizing the sum of square errors between the desired and actual outputs as [4, 5]:

$$E = \frac{1}{2} \sum_{n=1}^{N} (D_n - Y_n)^2$$
(4)

where:  $D_n$  and  $Y_n$  are the desired and actual outputs of the  $n^{th}$  output neuron for N output neurons.

The algorithm is stopped either when satisfied small value of E or a maximum number of iterations is reached. Error back-propagation algorithm can be used for this task efficiently.

#### V. PROPOSED IMAGE IDENTIFICATION

## APPROACH

In case of existence of noise or any type of degradations, the image identification process becomes an obstacle task. Overcoming these problems is the main challenge task of the proposed method. Another challenge task is the application of the proposed method for satellite images identification which is a novel subject that should pay attention to!!

The discrete wavelet transform (DWT) can be a useful tool to overcome the degradation problems. Performing one level of DWT for an image signal decomposes the signal into approximation and detail coefficients. Features can be extracted from the DWT of an image and added to the feature vector extracted from the image itself to obtain a large feature vector suitable for image identification in the presence of degradations. Wavelet denoising is another technique that can be used to reduce the effect of image noises prior to identification. The proposed approach for feature extraction in the presence of degradations is illustrated in Fig. 4.



Fig. 4 The proposed approach for feature extraction in the presence of degradations

## A Images Transformation from 2-D Matrix to 1-D Vector

Applying of MFCC-based feature extraction technique to satellite images is a new topic as MFCC technique is mainly applied to speech or voice recognition. The reason is that speech or voice is 1-D signals as MFCC can deal with directly. So, transformation of 2-D matrix to 1-D vector is the primary step of our proposed algorithm. Then, DWT with or without denoising is applied. After that, the usual procedure of MFCC is executed followed by concatenating all coefficients in the feature vector for feature extraction. Finally, feature matching using ANNs technique is performed [4, 5].

## **B** Discrete Wavelet Transform

The DWT is a very popular tool for analyzing nonstationary signals. It can be regarded as filtering the image signal with a bank of bandpass filters, whose impulse responses are all approximately given by scaled versions of a mother wavelet. The scaling factor between adjacent filters is usually 2:1 leading to octave bandwidths and center frequencies that are one octave apart [10-12]. The output of filters are usually maximally decimated so that the number of DWT output samples equals the number of input samples and thus no redundancy occurs in this transformation [1].

For one-level DWT decomposition, two filters are applied to the signal, filter  $H_0$  is a low pass filter whose output is an *approximation* part of input signal x(n) while filter  $H_1$  is a high pass filter whose output is the high frequency or *detail* part of x(n). All filtering is performed in the time domain by convolving each filter's input with its impulse response [13].

## C Wavelet Denoising

Wavelet denoising is a simple operation which aims to noise reduction of noisy signals. It is performed by choosing a threshold that is sufficiently a large multiple of the standard deviation of the noise in the signal. Most of the noise power is removed by thresholding the detail coefficients of the wavelet transformed speech signal. There are two types of thresholding; hard and soft thresholding. Hard thresholding is given by [14]:

$$f_{hard}(x) = \begin{cases} x & |x| \ge T \\ 0 & |x| < T \end{cases}$$
(5)

On the other hand, that of soft thresholding is given by:

$$f_{soft}(x) = \begin{cases} x & |x| \ge T \\ 2x - T & T/2 \le x < T \\ T + 2x & -T < x \le -T/2 \\ 0 & |x| < T/2 \end{cases}$$

where T denotes the threshold value and x represents the detail coefficients of DWT.

## VI. EXPERIMENTAL RESULTS

In this section, identification experiments are carried out for three cases. Number of images that are used for each experiment is eight. For the first case, an eight satellite images of high resolution satellites (IKONOS, SPOT-5, and QuickBird) without any processing are used. Second case, the same eight satellite images after applying roads extraction approach that is called morphological direction and length filters (MDLFs) approach are used [7,8]. Finally, for the third case, wavelet-Based MDLFs roads extraction approach is applied to the eight satellite images [9]. Each image is 400x400 pixels resolution.

Four methods for extracting features are adopted in the paper. In the first method, MFCCs are extracted from the image signals only. In the second one, the features are extracted from the DWT of the image signals. In the third method, the features are extracted from both the original image signals and the DWT of these signals and concatenated in a single feature vector. In the last method, denoising is applied and features are extracted from both the denoised signals and DWT of the denoised signals.



Stop Training

(6)

**Fig. 5 Recognition Rate vs. SNR for normal satellite images** A comparison study is done between the four methods for the above mentioned cases and results are given in Figs.Fig. 5 to Fig. 7. For first case of normal satellite images, it is clear from Fig. 5 that features extracted from both the image signals and the DWT of these signals achieve the highest recognition rates at moderate and high signal to noise ratios (SNRs). At low SNRs, features' extraction from the DWT of the original images is the highest. Wavelet denoising here does not have any effect as it does not improve the recognition rate.

For the second case of images after applying the MDLFs roads extraction approach studied in Fig. 6, the best performance is achieved by features matching of the original images only. It is also clear from this figure that Wavelet denoising is necessary at low SNRs.

For the last case of images after applying wavelet-based MDLFs roads extraction technique, it is clear from Fig. 7 that features extracted from both the image signals and the DWT of these signals achieve the highest recognition rates at moderate and high signal to noise ratios (SNRs). DWT features' matching is required at the low SNRs.



Fig. 6 Recognition Rate vs. SNR for images after applying the MDLFs roads extraction approach

## VII. CONCLUSION AND FUTURE WORK

This paper has presented a robust identification method for satellite images based on MFCC and the wavelet transform techniques. In this method, MFCCs coefficients are extracted from the images after transformation to 1-D signals and the DWT of these signals and concatenated to form a large feature vector. Experimental results have shown that the proposed method is useful for feature matching of images as a new application for MFCC technique as it is always used for speech or voice recognition.

The best performance is achieved by features extracted from both the image signals and the DWT of these signals for most cases.



## Fig. 7 Recognition Rate vs. SNR for images after applying wavelet-based MDLFs roads extraction approach

Future work is related to capability of MFCC to deal with 2-D matrix directly instead of its transformation to a 1-D vector before dealing with, as a 2-D matrix have more real features that can discriminate images for example more efficiently. So, the images identification or recognition rate will be higher.

## ACKNOWLEDGMENT

We would like to thank Dr. Fathi E. Abd El-Samie and Dr. Moawad Dessouky from the Faculty of Electronic Engineering for his efforts during the course of this work and the arrangement of this paper.

## REFERENCES

- A. Shafik, S. M. Elhalafawy, S. M. Diab, B. M. Sallam and F. E. Abd El-samie," Wavelet Based Approach for Speaker Identification from Degraded Speech," *IJCNIS*, 2009.
- [2] D. Pullella, "Speaker Identification Using Higher Order Spectra," Dissertation of Bachelor of Electrical and Electronic Engineering, University of Western Australia, 2006.
- [3] K. Sri Rama Murty and B. Yegnanarayana, "Combining Evidence from Residual Phase and

MFCC Features for Speaker Recognition", *IEEE Signal Processing Letters*, VOL. 13, NO. 1, January 2006.

- [4] A. I. Galushkin, "Neural Networks Theory," Springer-Verlag Berlin Heidelberg, 2007.
- [5] G. Dreyfus, "Neural Networks Methodology and Applications," Springer-Verlag Berlin Heidelberg, 2005.
- [6] Colin Fyfe, "Artificial Neural Networks," Edition 1.1, 1996.
- [7] T. M. Talal, M. I. Dessouky, A. El-Sayed, M. Hebaishy and F. E. Abd El-Samie, "Road Extraction from High Resolution Satellite Images by Morphological Direction Filtering and Length Filtering," *Proc. ICCTA'08*, pp. 137-141, October 2008.
- [8] T.Tateyama, Z.Nakao, X.-Y.Zeng, and Y.-W.Chen, "Segmentation of high resolution satellite images by direction and morphological filters," *Proc. HIS'04*, VOL. 2291, No. 2, pp. 482-487, December 2004.
- [9] H.Kasban, O.Zahran, M.El-Kordy, Sayed M. S. Elaraby and F. E. Abd El-Samie, "Automatic Object Detection from Images Scanned by LDV Based Acoustic to Seismic Landmines Detection System", *Proc ICCTA'08*, pp. 434-439, October 2008.
- [10] I. Daubechies, "Where Do Wavelets Come From?—A Personal Point of View," Proc. IEEE, Vol. 84, No. 4, pp. 510- 513, 1996.
- [11] A. Cohen and J. Kovacevec, "Wavelets: The Mathematical Background," *Proc. IEEE*, Vol. 84, No. 4, pp. 514- 522, 1996.
- [12] N. H. Nielsen and M. V. Wickerhauser, "Wavelets and Time-Frequency Analysis," *Proc. IEEE*, Vol. 84, No. 4, pp. 523-540, 1996.
- [13] R.C.Gonzalez and R.E.Woods, "Digital Image Processing," 2nd Edition, Prentice-Hall Inc., 2002.
- [14] J. S. Walker, "A Primer on Wavelets and Their Scientific Applications," CRC Press LLC, 1999.



**T. M. Talal** received his B.Sc. (Very Good) from the faculty of Electronic Engineering, Menofia University, Menouf, Egypt, in 1999. He joined Egyptian Space Program (ESP) at National Authority of Remote Sensing and Space Science (NARSS), in 2001, to be a member of system engineering and

image processing groups until now. He is M.Sc. student at faculty of Electronic Engineering, Menofia University, Menouf, Egypt. His current research areas of interest include satellite imaging system, pattern recognition and image segmentation.



**Ayman El-Sayed** received the B. Sc. and M. Sc. degrees from Faculty of Electronic Engineering, Menofia University, Menouf, Egypt in 1994, and 2000, respectively, and Ph.D. degree from INRIA Rhône-Alpes and INPG, France in 2004. He is currently

an IEEE member and Head of Computer Science and information System department, Faculty of Community in ALqeeyah,King Saud University, KSA. His current research areas of interest include Multicast, Mobile IP, Mobile Network, IP TraceBack, and Image Processing.