CLASSIFICATION OF COMPRESSED AUDIO SIGNALS VIA COMPRESSIVE SENSING IN A TELE-MONITORING CONTEXT

Bochra BOUCHHIMA [1]*, Rim AMARA* [1,2]*, Monia TURKI HADJ-ALOUANE*[1]

[1] Université de Tunis El Manar,
Ecole Nationale d'Ingénieurs de Tunis,
$03/UR/11 - 07$
Laboratoire de Signaux et systémes,
1002, Tunis, Tunisie

[2] Université de Carthage,
Institut National des Sciences
Appliquées et de Technologie
1080, Tunis, Tunisie

Home tele-monitoring was proposed in recent years as a patient management approach aimed at reducing the steadily increasing healthcare costs. It consists on profiting from artificial intelligence and technologies to make sure that patients are safe in their homes. Wireless sensor networks (WSNs), for example, provide an infrastructure capable of supporting this type of applications. The sensors collect information about what is happening in their surroundings and transmit it to the analysis center where data is interpreted and identified. This analysis phase is crucial as it provides accurate information about the safety level of the patient. Dealing with sensor networks, Compressive Sensing(CS) [1] is a quite recent attractive paradigm adapted to the low complexity requirement on the sensor side. In this work we intend to use CS to optimize the sensing and analysis process in an audio WSN.

CS aims at, jointly, acquiring and compressing data. Formally, consider $x \in \mathbf{R}^N$ as the signal to be acquired in a transmission application context. CS states that only $M < N$ samples of x are transmitted. They are given by the linear multiplication of x by a measurement matrix $\mathbf{\Phi} \in \mathbf{R}^{M \times N}$. $y = \mathbf{\Phi}x$ is called the measurement vector. Given y, the recovery of x is possible due to the sparsity property. In fact, x is said to be sparse if there exists a basis $\mathbf{\Psi} \in \mathbf{R}^{N \times K}$ in which it is expressed as $x = \mathbf{\Psi}s$ with s having at most $P \ll N$ non zeros elements. Signals in general are compressible but can be well approximated by a sparse model. Applying a successful CS scenario consists therefore in three major tasks : designing an adequate sparsity matrix for the signals of interest, a good measurement matrix and a robust algorithm to recover s from y.

As far as tele-monitoring context is concerned, the used signals are various short-duration sounds such as alarm sounds, glass breaking, door slamming, cat meowing, screams, etc. Some are related to normal situations, while others stand for emergency cases. We propose a new method for learning the sparsity matrix $\mathbf{\Psi}$. It is based on Empirical Mode Decomposition (EMD) and Hilbert Transformation. The matrix is designed to minimize the signal reconstruction error under some sparsity conditions. Depending on this matrix, the measurement matrix $\mathbf{\Phi}$ is built using a gradient algorithm. The objective is to reduce the incoherence (μ) as much as possible.

$$\mu = \frac{\displaystyle\sum_{1 \le i,j \le K, i \ne j} | \, \widetilde{d}_i^H \, \widetilde{d}_j \, |}{K(K-1)} \tag{1}$$

where $\widetilde{d}_i$ denotes the ith column of $D = \mathbf{\Phi}\mathbf{\Psi}$ scaled to unity (in l_2 norm sense). The reconstruction consists on recovering the signal x from the measurement y and is done using one of the state of the art methods (OMP, IHT, BP, ...). Results for the reconstruction are satisfactory.

Coming to the analysis phase, we intend to perform classification on compressed form of the signals. Generally, the following combination of features are well adapted to environmental sounds manifold: Zero Crossing Rate (ZCR), Short Time Energy (STE), Spectral Roll-off Point (SRF), Spectral Centroid (SC), Mel Frequency Cepstral Coefficients (MFCC) Mel-frequency Discrete Wavelet coefficients (MFDWC). They are relevant and well adapted to environmental sounds in their uncompressed forms. Using an SVM classifier, for example, we achieve high accuracy. Yet, when the signal is in a compressed form, these features seem not to be adapted. New features should be proposed and extracted from compressed form signals to achieve high recognition rates. Another approach consists in merging the features extraction and the sensing process together. Likewise, three tasks are done simultaneously by the sensors side: acquiring data, compressing data and features extraction. The sensing process is therefore made more intelligent and classification can be done faster, more directly and more easily in the analysis center.

REFERENCES

[1] D.L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.