ADHD identification based on a linear projection and clustering

P.A. Castro-Cabrera[a,*], D.H. Peluffo-Ordóñez[a], F. Restrepo de Mejía[b], C.G. Castellanos-Domínguez[a]

[a] *Universidad Nacional de Colombia*
Signal Processing and Recognition Group
Manizales, Colombia
[b] *Universidad Autónoma de Manizales*
Grupo de Neuro-aprendizaje
Manizales, Colombia

Abstract

Event-related potentials (ERPs) are electrical signals from brain generated as a response to an external sensorial stimulus. This kind of signals are widely used to diagnose neurological disorders, such as Attention-deficit hyperactivity disorder (ADHD).

In this paper, a novel methodology for ADHD discrimination is proposed, which consist of obtaining a new data representation by means of a re-characterization of initial feature space. Such re-characterization is done through the distances between data and centroids obtained from k-means algorithm. This methodology also includes pre-clustering and liner projection stages. In addition, this paper explores the use of morphological and spectral features as descriptive patterns of ERP signal in order to discriminate between normal subjects and ADHD patients. Experimental results show that the morphological features, in contrast with the remaining features considered in this study, are those that more contribute to classification performance, reaching 86 % for the original feature set.

Keywords: ADHD, ERP, clustering, linear projection

1. Introduction

Attention-deficit hyperactivity disorder (ADHD) is a prevalent disorder diagnosed on the basis of persistent and developmentally-inappropriate levels of

*pacastroc@unal.edu.co

overactivity, inattention and impulsivity. It is one of the most common psychiatry disorders in childhood [1]. Currently its diagnosis is based on the clinical criteria of DSM-IV or ICD-10, helped by the conduct outlined in questionnaires applied to parents and teachers; however, there are not biological markers or conclusive tests that diagnose this behavioral disorder with a high reliability [2].

Event-related potentials (ERPs) are brain electrical signals generated as a response to an external sensorial stimulus. They have been useful in investigations of perceptual and cognitive-processing deficits, specially in children with ADHD, given that these potentials are physiologically correlated with neurocognitive functions. The most popular assessed features on ERPs for interpretation of cognitive processes are the areas and the peaks of the ERP components, defined by the mean and peak to peak voltages, respectively, which are computed in certain windows in the time domain. This parameters are determined by visual inspection of the averaged ERP waveforms [3].

The ERPs comprise of a number of characteristic peaks and trough which basic research has shown to correspond to certain underlying processes. P300 component is perhaps the most studied ERP component in investigations of selective attention and information processing, due partly to its relatively large amplitude and facile elicitation in experimental contexts [4]. Although the quantification of ERP components by areas and peaks is the standard procedure in fundamental ERP research, the conventional approach has two drawbacks:

Firstly, ERPs are time-varying signals reflecting the sum of underlying neural events during stimulus processing, operating on different time scales ranging from milliseconds to seconds. Various procedures such as ERP subtraction or statistical methods have been employed to separate functionally meaningful events that partly or completely overlap in time. However, the reliable identification of these components in the ERP waveforms still remains as a problem.

Secondly, analysis in the frequency domain has revealed that EEG/ERP components in different bands (delta, theta, alpha, beta, gamma) are functionally related to information processing and behavior. However, the Fourier transform (FT) of ERP lacks the information about the time localization of transient neural events. Therefore, efficient algorithms for analyzing a signal in time-frequency plane are very important in extracting and relating distinct functional components.

These limitations, as well as the ones related with time–invariant methods, can be solved by using the wavelet formalism. The wavelet transform

(WT) is a time-frequency representation, that has an optimal resolution both in the time and frequency domains and has been successfully applied to the study of EEG–ERP signals [**?**][5]. Although ERP feature extraction from the time–frequency domain based on the discrete WT (DWT) has been growing increasingly popular, this approach can result unhelpful to pathology detection purposes, particularly, to ADHD identification.

In this paper a novel methodology is proposed that consist of a re-characterization of initial feature space through the distances between data and centroids corresponding to clusters obtained with k-means algorithm. To this end, original data are first selected by means of a pre-clustering and linearly projected. In addition, this paper explores the use of morphological and spectral features as descriptive patterns of ERP signal in order to discriminate between normal subjects and ADHD patients. Experimental results show that the morphological features, in contrast with the remaining features considered in this study, are those that more contribute to classification performance, reaching 86 % for the original feature set.

2. Theoretical framework

In terms of spectral clustering, in particular, graph-partitioning clustering, affinity matrix represents the relation degree between observations or nodes. In other words, affinity measure denotes the association or similitude degree between two nodes and then it is a non-negative value.

Let $\mathbf{A} = \{a_{ij}\}$ be the affinity matrix that is composed by all the relations among nodes, that satisfies the following conditions $a_{ij} \geq 0$ and $a_{ij} = a_{ji}$. Then, matrix $\mathbf{A}$ is symmetric and positive semi-definite.

Let $\boldsymbol{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{n \times p}$ be the data matrix where $\mathbf{x}_i$ is a p-dimensional vector that corresponds to considered features for i-th subject. To guarantee the scale coherence in representation of data, matrix $\mathbf{X}$ is normalized , using: $\mathbf{x}_i \leftarrow (\mathbf{x}_i - \mu(\mathbf{x}_i))/\sigma(\mathbf{x}_i)$, where $\mu(\cdot)$ and $\sigma(\cdot)$ are a mean and a standard deviation operator, respectively.

A trivial form to establish the affinity measures corresponds to $\mathbf{A} = \mathbf{X}\mathbf{X}^\mathsf{T}$. For some clustering methods, this kind of affinity results to be useful because it contains the inner products between all data points or observations.

2.1. *Data truncated-projection*

In general, when using spectral techniques the clustering procedure is carried out in a low dimensional space, named eigen-space [6]. Denoting the eigen-space as $\mathbf{U}$ that corresponds to the eigenvectors of $\mathbf{A}$ and its corresponding subspace as $\widetilde{\mathbf{U}} \in \mathbb{R}^{n \times m}$, where $m < n$, that corresponds to the first m columns of $\mathbf{U}$.

In this work, it is proposed to employ $\widetilde{\mathbf{U}}$ as a rotation matrix, but given that n is significantly bigger than p the use of matrix $\mathbf{V}$ that represents the eigenvectors of $\boldsymbol{X}^\top \boldsymbol{X}$ is preferred. This can be done because the first p eigenvalues of $\mathbf{X}\mathbf{X}^\top$ correspond to the eigenvalues of $\mathbf{X}^\top\mathbf{X}$ when $||\mathbf{u}_i|| = ||\mathbf{v}_i|| = 1, i = 1, \dots, p$.In addition, it is easy to prove that there exists a linear relation between $\mathbf{u}_i$ and $\mathbf{v}_i$, which is: $\mathbf{u}_i = \mathbf{X}\mathbf{v}_i$.

Given this, data linear projection is:

$$\mathbf{Y} = \mathbf{X}\mathbf{V} \tag{1}$$

Then, truncated linear projection is obtained as follows:

$$\widetilde{\mathbf{Y}} = \mathbf{X}\widetilde{\mathbf{V}} \tag{2}$$

where matrix $\widetilde{\mathbf{V}} \in \mathbb{R}^{p \times q}$ $(q < p)$ is composed by the first q columns of $\mathbf{V}$.

2.2. *Clustering-based representation*

For most spectral clustering approaches, once the new representation space is obtained, a conventional clustering algorithm is applied to group homogeneous observations [7]. In this work, a clustering-based representation is proposed. To this end, a centroid-based clustering is used to obtain a new data representation $\mathbf{Z} = \{z_{ij}\}$, where each observation is represented by means of its distance with the centroids corresponding to each group, i.e.,

$$z_{ij} = \mathrm{d}(\hat{\mathbf{y}}_i, \mathbf{q}_j), \qquad i = 1, \dots, n; \;\; j = 1, \dots, k \tag{3}$$

where k is the number of groups, $\mathbf{q}_j$ denotes the j-th centroid and $\mathrm{d}(\cdot, \cdot)$ is a distance operator.

Centroids are obtained with k-means algorithm [8] and as a distance measure the euclidian norm is used.

2.3. Heuristic search

To improve the performance classification and determine the relevance features, a feature selection stage is done. In this particular case, a heuristic search type *wrapper*, called sequential forward floating selection (*sequential forward floating selection* - SFFS) [9]. In this technique, for each stage a new variable is included using a forward sequential procedure, then less significant variables are excluded one at a time until the percentage of accurate classifications increases. Once this search can no longer continue excluding variables, another step forward is done to include another variable and if possible the exclusion of variables procedure is applied again. The process is iterated until no more steps forward can be done because a high percentage of classification accuracy has been achieved.

3. Materials and methods

3.1. Data Base

The experiments were carried out with 120 children belonging to educational institutions of the metropolitan area of the Manizales (60 of the healthy control group and 60 of the ADHD group). The subjects, with ages between 4 and 15 years old, were medically diagnosed based on clinical criteria of DSM-IV and minikid criteria by a multidisciplinary specialist team consisting of a general physician, psychologist, neuropsychologist and experts in children psychiatric disorders. Both groups were tested under the same lighting and noise conditions, and were defined by the following inclusion criteria: non abnormality physical examination, normal visual and hearing ability, intellectual coefficient greater than 80 and, if necessary, pharmacologic management previously suspended. Subjects were verified to be free of some evidence of other neurological disorder.

Recordings were acquired by means of electrodes located in the head midline (Fz, Cz, Pz) according to $10-20$ international system, with a sampling frequency of 640 samples per second. Signals acquisition took 1 s before and after stimulus presentation. The evaluation protocol applied was the oddball paradigm in auditory and visual modalities. The first procedure involves the emission of 80 dB tone lasting 50 ms, with a frequency of 1,000 Hz for frequent stimulus and 3,000 Hz for target stimulus, presented randomly every 1,5 s. In the visual modality of the test, the subject is asked to watch a monitor placed 1 m away that shows an image with a consistent pattern (a checkerboard of 16 squares), which is the frequent stimulus. The rare stimulus is the presentation of a target in the center of the screen with the same common pattern in the background; the subject must press a button each time the unusual stimulus appears. The experiment consists of 200 stimuli, of which 80 % are non-target and 20 % remaining are target stimuli.

3.2. Experimental setup

Methodology applied in the experiments can be seen graphically in the block diagram shown in figure 3.2.

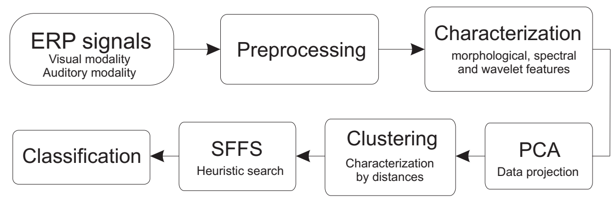

Figura 1: Proposed methodology for ERP signal analysis

Proposed methodology consist of some procedures described below. Database counts with 6 recordings per patient, corresponding to acquisitions of Fz, Cz and Pz electrodes in the auditory and visual modalities. In this work, it is reported only the results of Pz auditory record, since this is the location of the scalp where the generators of the ERP components act more clearly.

Data matrix $\mathbf{X}$ is defined as suggested in [10], which consists of three groups of features of different nature: The first group comprises 17 morphological features, which consist of parameters measured over the whole signal and are related to its shape. This set is formed by the following characteristics: latency, amplitude, latency/amplitude ratio, absolute amplitude, absolute latency/amplitude ratio, positive area, negative area, total area, absolute total area, total absolute area, average absolute signal slope, peak-peak value, peak-peak value in a time window, peak-peak slope, zero crossings, zero crossings density and slope sign alterations.

The second set of features is defined by three frequency characteristics: mode frequency, median frequency and mean frequency, which are calculated as described in [10]. Using the discrete wavelet transform, we obtain the third set of characteristics, which corresponds to the wavelet coefficients from the previous levels of decomposition.

After characterizing, the corresponding processing is performed on the matrix $\mathbf{X}$ using the following procedure: centralization and standardization of data, outlier detection and verification of univariate gaussivity. In addition to the above procedure, it is performed a data pre-clustering with the methodology used in [11], thus ensuring consistency in data and facilitate the analysis.

Now, the projection of the data is done by using the technique explained in Section 2.1. The criterion used to determine q is an accumulated variance value greater than 90 %

Subsequently, data representation is redefined through the distance between the data and the centroids of the formed groups by applying the clustering technique. To this end, it has been implemented the traditional algorithm of *k-means* and used Euclidean distance as a dissimilarity measure.

Once projected data are re-characterized calculating the distances between the data and the centroids, a heuristic search algorithm is applied. In this case, a sequential floating forward selection (SFFS) is considered. This is done in order to perform a supervised reduction that may lead to find the smallest number of features that allow sufficiently data classification. The implemented SFFS algorithm utilizes as a classification assessment function a Bayesian classifier, since each probability density function is modeled as Gaussian. In addition, the method was improved by a hypothesis test (*t-test*) and an evaluation of information loss stage [12].

The following algorithm describes mathematically and sequentially the stages of the proposed methodology previously explained.

Algorithm 1 Re-characterization of ERP signals through dissimilarity measures

Input: $\mathbf{X}_{n\times p}$.

1. A pre-clustering stage is applied over data matrix: $\hat{\mathbf{X}} = \mathrm{preclustering}\{\mathbf{X}\}$, where matrix $\hat{\mathbf{X}}$ is $h \times p$ dimensional and $h < n$.
2. Estimate the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$.
3. Compute the eigenvalues $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ and eigenvectors $\mathbf{V} = [\boldsymbol{v}_1|\ldots|\boldsymbol{v}_p]$ of $\boldsymbol{\Sigma}_{\mathbf{X}}$ decreasingly organized, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$.
4. Determine the value of q ($q < p$) through an accumulated variance, greater than 90 %
5. Obtain the truncated linear projection: $\hat{\mathbf{Y}} \in \mathbb{R}^{h\times q} = \hat{\mathbf{X}}\hat{\mathbf{V}}$.
6. Cluster data and obtain final centroids: $\mathbf{Q} = [\mathbf{q}_1^{\mathsf{T}}|\cdots|\mathbf{q}_p^{\mathsf{T}}] = \mathrm{kmedias}(\hat{\mathbf{Y}})$
7. Re-characterize projected data: $\boldsymbol{B} \in \mathbb{R}^{h\times k} = \{b_{ij}\} = \{\mathrm{d}(\hat{\boldsymbol{y}}_i, \boldsymbol{q}_j)\}, i = 1, \ldots, n;\ \ j = 1, \ldots, k$, $\mathrm{d}(\cdot,\cdot)$ is the euclidian distance,
8. $\hat{\boldsymbol{B}} \in \mathbb{R}^{h\times m} = \mathrm{SFFS}\{\boldsymbol{B}\}$, m is the number of relevant variables, $m < k$
9. Test: Classifiers k-nn, LDC and SVM (70 % for training and 30 % for test).

$\hat{\boldsymbol{B}} = \{$ effective feature set$\}$

4. Results and discussions

The acquisition of recordings was carried out in the auditory and visual modalities through electrodes placed at positions Fz, Cz and Pz, as explained in Section 3.1. In this work, it is reported only the results of the Pz position in the auditory modality, since it is the region where the ERP signal generators yield event-related potentials with component more defined and greater amplitude.

Data matrix $\mathbf{X}$ has been made up for 16 morphological and 3 spectral features, and 32 wavelet coefficients. To calculate the wavelet features, the records must be resampled to 1024 Hz and discrete wavelet transform was used with a *biorthogonal spline* as a wavelet function, and 3 vanishing moments. For this work, a decomposition of 7 levels was applies, in order to approximately adjust the frequency band levels into the brain rhythms such as delta (0,2 to 3,5 Hz), theta (3,5 to 7,5 Hz) , alpha (7,5 to 13 Hz) and beta (13 to 28 Hz). From 7 obtained decomposition levels, approximation coefficients of level 7 and details coefficients of levels 7, 6 and 5 were selected as characteristic wavelet. To justify the selection of these coefficients was used a criterion of informativeness based on accumulated *Shannon* entropy [13] with a threshold greater than 60 %.

To carry out the classification tasks, it was used three different classifiers: a k-NN, a linear discriminant (LDC), and a support vector machine (SVM), in order to compare the performance of them and select the one that offers higher performance classification. In validation step was used a partition of 70 % for the training group and 30 % for the test group. The testings produced the following result:

Figure 4 displays the performance of a k-NN classifier in continuous repetition to show the stability of the proposed methodology. It can be observed, that all values of the classification performance is above 80 % and maintain an acceptable standard deviation.

Figure 4 shows the performance obtained by the feature subsets obtained after selection algorithm SFFS, namely: 1. Performance for the first selected characteristic, 2. Performance for the subset formed by the first and second selected features, and so on.

Table 1 shows the accuracy, specificity and sensitivity of each group of features. In table can be seen that from original set of features $\mathbf{X}$, the morphological characteristics are the major contributors in the performance of the classifier. This same condition is also evident in percentages achieved by this subset of features for the sensitivity and specificity.

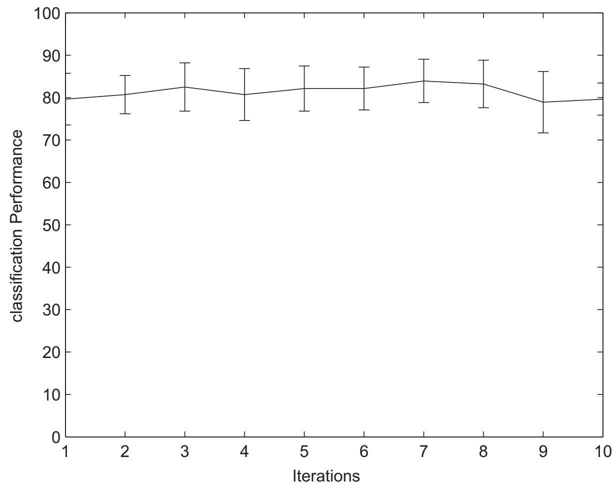

Figura 2: Stability of methodology with respect to iterations

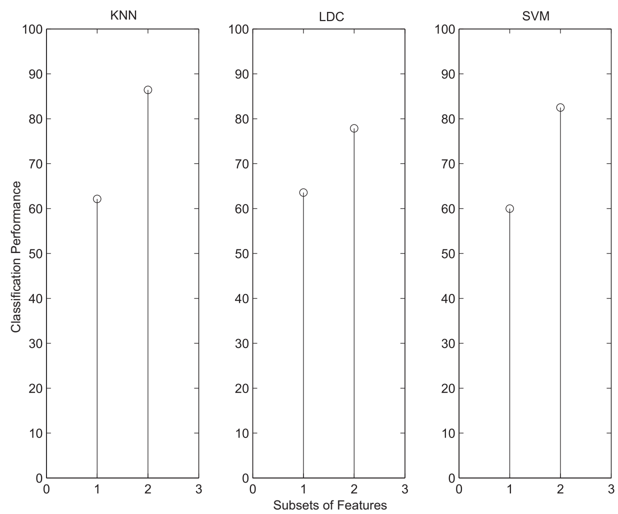

Figura 3: Classification performance for each feature subset

Feature	Accuracy (%)	Specificity (%)	Sensitivity(%)
Morphological	85,35 ± 3,9	85,00	85,83
Spectral	63,92 ± 8,6	73,12	51,66
Wavelet	67,85 ± 8,5	73,12	60,83

Cuadro 1: Classification performance for each group of features

Table 2 relates the classification accuracy of three classifiers mentioned above. It is noted that a simple parametric classifier as the k-NN can achieve an optimum performance. Moreover, it is observed that the highest rates of specificity and sensitivity were achieved also with the k-NN.

Classifier	Accuracy (%)	Specificity (%)	Sensitivity(%)
k-NN	86,07 $\pm$ 3,5	85,00	87,50
LDC	73,92 $\pm$ 7,1	82,5	62,50
SVM	78,57 $\pm$ 4,7	81,25	75,00

Cuadro 2: Classification performance for each group of features

Table 3 shows the classification rates achieved without preclustering over $\mathbf{X}$ applied in preprocessing stage. In comparison with table 2, there is a considerable drop in the percentage of classification, which can be attributed to the presence of outliers or heterogeneous data.

Classifier	Accuracy (%)	Specificity (%)	Sensitivity(%)
k-NN	56,04 $\pm$ 5,6	62,91	49,16
LDC	48,54 $\pm$ 6,5	42,08	55,00
SVM	48,33 $\pm$ 4,2	24,58	72,08

Cuadro 3: Classification performance for the proposed methodology

Given the low classification accuracy obtained in this test, it has been proved the need to make a preclusting in preprocessing. These results also show the low reliability of the labels given by medical specialists.

5. Conclusion

Because of the nature of ERP signals and the low reliability of labeling given by specialists, the identification of ADHD represents a difficult task for both medicine and pattern recognition. The design of classification systems for discrimination between ADHD and normal signals require a new data representation since signal samples cannot be enough to obtain a good classes separability.

In this work, to try overcoming this problem, a methodology of data re-characterization is proposed. This methodology is mainly composed by two stages: truncated liner projection and clustering-based characterization. They are done to achieve a good data representation in terms of compactness and the distance-based representation from obtained clusters that improves the classification performance, respectively. These two stages are complementary and coherent at each other.

In addition, a pre-clustering stage was introduced, which allows that the classifiers perform better because supposed outlier observations are discarded.

Acknowledgment

Authors would like to thank COLCIENCIAS and Universidad Nacional de Colombia for the financial support on the projects *"Identificacin Automtica del Trastorno por Dficit de Atencin y /o Hiperactividad sobre registros de Potenciales Evocados Cognitivos"* and *"Sistema de Diagnstico Asistido para la Identificacin de TDAH sobre Registros de Potenciales Evocados Cognitivos"*, respectively.

Referencias

[1] M. A. Idiazbal, A. Palencia-Taboada, J. Sangorrn, J. Espadaler-Gamissans, Potenciales evocados cognitivos en el trastorno por dficit de atencin con hiperactividad, Rev Neurol 34 (2002) 301–305.

[2] R. A. Barkley, Attention deficit hyperactivity disorder: a handbook for diagnosis and treatment, Guilford Press, New York, 3rd Edition (2005).

[3] V. Bostanov, Data sets Ib and IIb: Feature extraction from eventrelated brain potentials with the continuous wavelet transform and the t-value scalogram, IEEE Transactions on Biomedical Engineering 51 (6) (2004) 10571061.

[4] S. H. Patel, P. N. Azzam, Characterization of n200 and p300: Selected studies of the event-related potentials, International Journal of Medical Sciences 2 (4) (2005) 147–154.

[5] I. Kalatzis, N. Piliouras, E. Ventouras, C. Papageorgiou, A. Rabavilas, D. Cavouras, Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals, Computer Methods and Programs in Biomedicine 75 (2004) 11–22.

[6] Y. S. X., S. Jianbo, Multiclass spectral clustering, in: ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision, IEEE Computer Society, Washington, DC, USA, 2003, p. 313.

[7] L. Zelnik-manor, P. Perona, Self-tuning spectral clustering, in: Advances in Neural Information Processing Systems 17, MIT Press, 2004, pp. 1601–1608.

[8] P. Hansen, N. Mladenovic, E. Commerciales, J-means: A new local search heuristic for minimum sum-of-squares clustering, Pattern Recognition 34 (2) (2001) 405–413.

[9] E. Delgado-Trejos, A. Perera-Lluna, M. Vallverd-Ferrer, P. Caminal-Magrans, G. Castellanos-Domnguez, Dimensionality reduction oriented toward the feature visualization for ischemia detection, in: IEEE Transactions on Information Technology in Biomedicine VOL. 13, NO. 4, JULY 2009, Vol. 13, 2009.

[10] V. Abootalebi, M. H. Moradi, M. A. Khalilzadeh, A new approach for EEG feature extraction in P300-based lie detection, Computer methods and programs in biomedicine 94 (2009) 4857.

[11] S. Murillo-Rendn, G. castellanos Domnguez, Construccin, limpieza y depuracin previa al anlisis estadstico de bases de datos., in: XV SIMPOSIO DE TRATAMIENTO DE SENALES, IMAGENES Y VISI ON ARTIFICIAL - STSIVA 2010, 2010.

[12] D. Ververidis, C. Kotropoulos, Fast and accurate sequential floating forward feature selection with the bayes classifier applied to speech emotion recognition, Signal Processing 88 (12) (2008) 2956–2970.

[13] R. Coifman, M. Wickerhauser, Entropy-based algorithms for best basis selection, Information Theory, IEEE Transactions on 38 (2) (1992) 713 –718. doi:10.1109/18.119732.