FEATURE EXTRACTION VIA COMPRESSED LOCAL MAHALANOBIS METRICS

Paolo Emilio Barbano & Ronald R. Coifman

Department of Mathematics
Yale University
New Haven CT

Index Terms— Pattern Recognition, Dimensionality Reduction, Compressed Sensing, Mahalanobis Distance

1. COMPRESSED SENSING IN HIGH NOISE

We propose a new energy-based ([1]) online learning approach to the problem of extracting small features from high dimensional, noisy datasets. We demonstrate the power of this new technique by applying it to the Detection of cellular nuclei in large hyperspectral images. Thus, we prove that only a minimum number of random, *wideband* measurements is necessary to achieve efficient feature extraction of nuclei from the highly structured background. The technology we introduce in our work rests on known results in *Compressed Sensing* (CS, [2]), a field of study dealing with the task of economically recording information about signals, viewed as elements of a high dimensional vector space $\mathcal{V} \simeq \mathbf{R}^N$, $N >> 1$. More precisely, CS is a framework developed to allow the reconstruction of discrete signals from few projections onto appropriate Vector Spaces. When attempting to solve a Detection/Estimation problem for a K-sparse [1] set in $V \simeq \mathbb{R}^N$, where measurments are taken place, we will need M such projections, where:

$$M = C \cdot K \cdot log\left(\frac{N}{K}\right) \quad (1)$$

The penalty for this reduction in sampling density is the *non-linearity* of the reconstruction algorithms needed. In fact, if $Y = \{y_1, \ldots, y_M\}$ denotes our measurements, obtained by applying the projection Φ to the data, condition (1) guarantees perfect recovery of the features with high probability. The solution s is defined as:

$$s = argmin\left(\|\hat{s}\|_1\right) \quad s.t.\ \Phi(\hat{s}) = Y \quad (2)$$

We also propose a new, iterative method for equation (2), designed to reject very high levels of distortion. These are often present in real biomedical imaging data. [2]

In the work summarized here, we have fully explored the robustness of CS reconstruction algorithms and developed a technique capable of dealing with a mixed distribution of Gaussian noise and Non-Gaussian clutter based on adaptive, online learning algorithms.

[1] This simply means that there will be a suitable basis $\Phi = \{\phi_k\}$ in which any $s \in \Sigma$ can be written as a linear combination on no more than K of its vectors. This is a condition verified by a plethora of objects of interest occurring in natural images

2. LOCAL MAHALANOBIS METRICS

In our model, a representative $R = \sum_{k=1}^{K} a_k \phi_k$ of the features we are interested in detecting is embedded in a mixture of noise and background objects:

$$R' = \Phi(R) + N + I \quad (4)$$

Where Φ represents the measurement matrix, N the wideband (e.g. Gaussian) component of noise and I represents a combination of narrowband, sparse interferences affecting the samples as well as background objects, or other objects present in the images. Now, since each of the multi-spectral images in the set $\mathcal{D} = \{D_j\}$ needs to contain information about all wavelengths of interest, complex signal processing tasks make the classifier unpractical.

A simple strategy can be used to get around this obstacle. Namely, projecting each image D on an appropriate low dimensional space. We generate a *randomized* dataset $\tilde{\mathcal{D}} = \{\tilde{D}_j\}$ by acquiring each image with M random, wideband waveforms $\{w_1, \ldots, w_M\}$, as in (1). This procedure achieves an almost *lossless compression* of the original information content ([4]). In the new, compressed data we construct a classifying map by utilizing a paradigm similar to the one

[2] A similar result fact has been successfully demonstrated by Candes and Tao in their work on error correction [3] There, the problem was formulated in terms of recovering a binary code-word after it had been encoded by a sparsifying transformation X and distorted by the interference active within a noisy, non-linear transmission channel. The decoder they solves the L1 minimization problem, by looking for the code-word w_0 such that:

$$w_0 = argmin_w\left(\|y' - X \cdot w\|_1\right) \quad (3)$$

Where X is a *known* reconstruction matrix, chosen randomly among those verifying some characteristic properties and y the received signal, distorted by a set of statistically defined deletions.

presented in [5]. We note here, that the basic dimensionality reduction step $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^M$ is theoretically motivated by (1).

The parameters of the classifier $\Delta : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$, used for detecting the nuclei are adjusted via online gradient learning from training data. The mapping Δ can be described mathematically as a *pseudo-metric* and is is defined as a one-layer, non-linear filterbank. In particular we write

$$\Delta(D_i) \;=\; \sum_{\alpha} w_\alpha \sigma_\alpha \left(\phi_\alpha \left(D_i\right)\right), \qquad (5)$$

where each component map ϕ_α is given by a the weighted version of a Local Mahalanobis Distance (LMD) generated by a small set of training points [3], weighted by scalars w_α and passed through sigmoid functions σ_α.

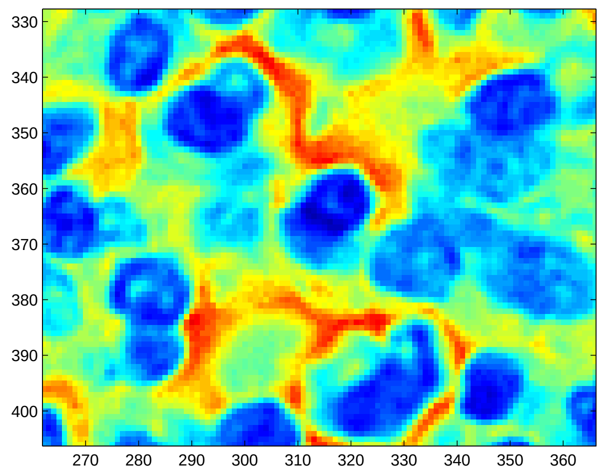

Fig. 1. Detail of the result of the nuclei extraction from one of the compressed wavelengths w_k. The Image $\tilde{D}$ contains both, local distortions as well as global features coming from other elements of the scene

3. REFERENCES

[1] Y. LeCun, S. Chopra, R. Hadsell, MA. Ranzato, and Huang F.J., “A tutorial on energy-based learning,” *MIT Press*, 2006.

[2] David L. Donoho, “Compressed sensing.,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] Fix a set $N = \{n_1, \ldots, n_M\}$ of points in a nucleus and determine the Mahalanobis Distance associated to a choice of N coefficients $\{c_{i(1)}, \ldots, c_{i(N)}\}$ and compute a point x_0 and all the other points $\{x_k\}$ in the nucleus.

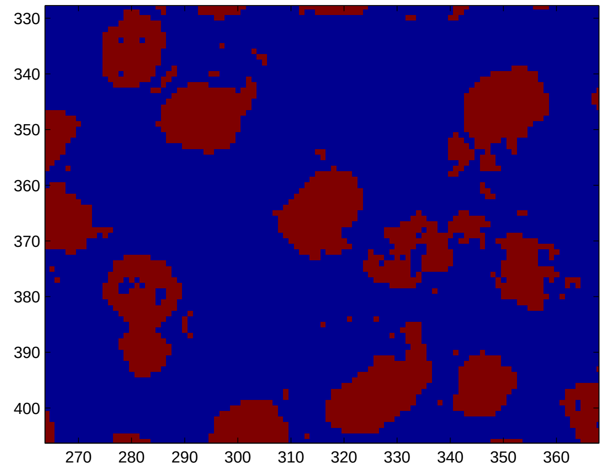

Fig. 2. Detail of the result of the nuclei extraction after processing with the Compressed Mahalanobis mapping Δ.

[3] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. on Information Theory*, vol. 51, pp. 4203 – 4215, 2005.

[4] M. Maggioni, G.L. Davis, F.J. Warner, F.B.Geshwind, A.C. Coppi, R.A. DeVerse, and R.R. Coifman, “Hyperspectral microscopic analysis of normal, benign and carcinoma microarray tissue sections,” 2006, vol. 6091, p. 60910I, SPIE.

[5] Hadsell, R., Chopra, S., and LeCun, Y., “Dimensionality reduction by learning and invariant mapping,” in *Proc. of Computer Vision and Pattern Recognition Conference*‘. 2006, IEEE Press.