AN ABSTRACT VIEW ON PATTERN RECONGNITION BASED ON CORRELATION

Octavian-Florian GEORGESCU[1], Florica MOLDOVEANU[2]

In this paper we propose an abstract view on pattern recognition, pointing out some fundamental relations around the central concept of correlation between the main representation domains of the signals, both analogical and digital. These domains are: the temporal or spatial domain, the frequency domain and the probabilistic domain. In a unitary, but also reflexive vision upon these domains we will refer to the principle of orthogonal decomposition in all the three domains, but also to the Heisenberg's uncertainty principle applied to the signal time-frequency analysis. This way, we specify a framework for the pattern recognition problem, that is, in essence, a problem of correlation / decorrelation.

Keywords: correlation, spatial domain, frequency domain, unitary transform, Fourier transform, uncertainty principle, probabilistic domain, Karhunen–Loève transform

1. Introduction

It is said that humans think and get the most quality information from images and concepts. It is given big picture and the central concept of our discourse in Fig. 1.

Observation: We refer to "spatial domain", considering we are working with digital images, but we could also refer to 'temporal domain' or 'time domain'. Generally, they used the terms 'spatial filtering' and 'frequency domain' [10]. In fact, the frequency could be spatial, temporal or of other kind, depending on the domain in which we apply the Fourier Transform (direct or inverse transform, this is the meaning of the bidirectional arrow).

[1] Eng., Siveco România S.A., PhD student, University POLITEHNICA of Bucharest, România, e-mail: goctav16@yahoo.com

[2] Professor, Department of Computer Science and Engineering, University POLITEHNICA of Bucharest, e-mail:florica.moldoveanu@cs.pub.ro

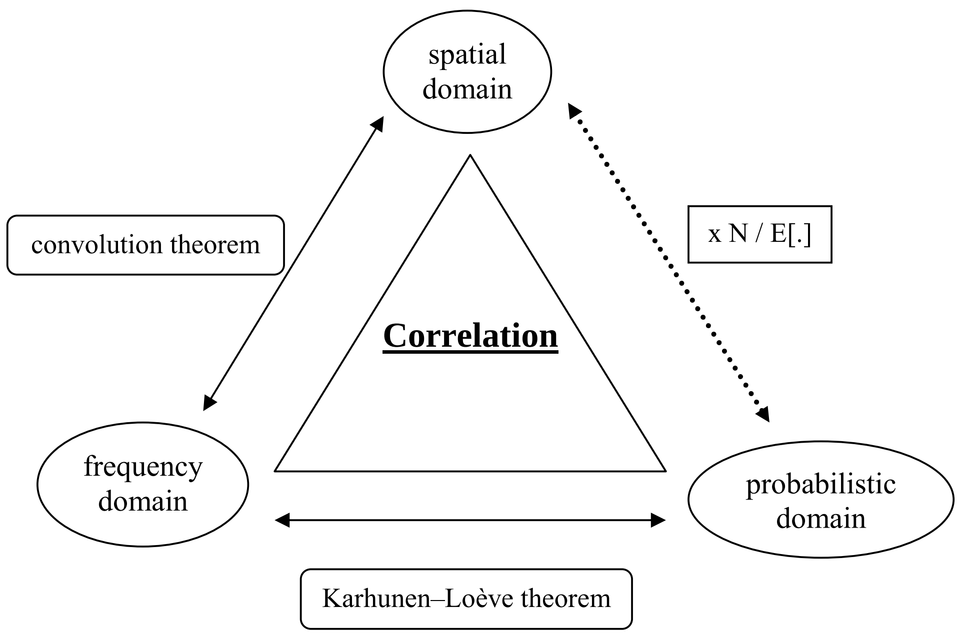

Fig. 1. Our «big picture»

Therefore:

- our central concept is the correlation; we will see how this concept is applied in all the mentioned domains
- there are some relations between the domains:

1. The convolution theorem establishes a relation between the spatial domain and the frequency domain. This theorem is equivalent with the correlation theorem from Fourier analysis. In fact, for an image-signal, by definition, the convolution is a correlation with the mirrored mask.
2. The Karhunen–Loève theorem by which a stochastic process could be represented as an infinite linear combination of orthogonal functions (forming an orthonormal basis), analogous to a Fourier series representation of a function on a bounded interval. The importance of the Karhunen–Loève theorem is that it yields the best such basis in the sense that it minimizes the total mean squared error.
3. The relation between spatial domain and probabilistic domain is that between what we see on an actual scene and the history of the events on the same scene. While ‘x N’ means the accumulation of spatial «images» which could generate a reasonable statistic, E[.] means the expectation or media

operator which estimates the most probable realization conform to the probabilistic data.

In the "Conclusions" section we also mention, without details, a main relationship between these domains and the «fuzzy» domain.

Someone could ask: What good is this synthesis? When a researcher has to analyze some signals in order to extract meaning from them, his ability to analyze signals in the main signals representation domains is almost a condition for the success of his investigation. This synthesis is also interesting because it catch in a clear picture, in an "orthophoto" view, the main approaches of signal processing and the intrinsic relations between these approaches. The author needed this view in front of various problems of signal processing and analysis.

2. Fourier transform - a prototype tool for frequency domain representations

First we have to state that our discussion takes place inside Linear time-invariant (LTI) system theory – see [1]. The fundamental result in LTI system theory is that any LTI system can be characterized entirely by a single function called the system's *impulse response*. The output of the system is simply the convolution of the system's input with the system's impulse response. This method of analysis is often called the "time domain" point-of-view. The same result is true for discrete-time linear shift-invariant systems in which signals are discrete-time samples, and convolution is defined on sequences.

Equivalently, any LTI system can be characterized in the *frequency domain* by the system's transfer function, which is the Laplace transform of the system's impulse response (or Z transform in the case of discrete-time systems). As a result of the properties of these transforms, the output of the system in the frequency domain is the *dot product* of the transfer function and the transform of the input. In other words, convolution in the time domain is equivalent to multiplication in the frequency domain: this is the general meaning of the convolution theorem.

For all LTI systems, the eigenfunctions, and the basis functions of the transforms, are complex exponentials. Thus, if the input to a system is the complex waveform Ae^{st} for some "complex" amplitude A and complex frequency s, the output will be some complex constant times the input, say Be^{st} for some new complex amplitude B. The ratio B / A is the transfer function at frequency s.

LTI system theory is suitable for describing many important systems - any system that can be modeled as a linear homogeneous differential equation with constant coefficients is an LTI system. In DSP, filtering operations could be seen as LTI systems - the filtering mask corresponds to the impulse response. Through the complex exponentials we could represent frequency components => LTI systems cannot produce frequency components that are not in the input. Most LTI systems are considered «easy» to analyze, at least compared to the time-varying and/or nonlinear case.

The Fourier transform could be seen as a special case of the Laplace transform - when the exponential have a purely imaginary argument -, giving the eigenvalues for pure complex sinusoids. The Fourier Transform (FT) for a real signal f(x) with finite energy is:

$$F(u) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ixu}dx \qquad (1)$$

This transform, compared with the Laplace transform, has the major advantage to be invertible: under suitable conditions, *f* can be reconstructed from F by the inverse transform:

$$F(u) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ixu}dx \qquad (2)$$

F(u) are the Fourier coefficients, i.e. the coefficients of the representation of the signal in the Fourier domain, the domain of the pure (spatial or temporal or any other parameter) frequencies u. FT is an *integral transform* with a *separable and symmetric kernel* – it operates on the entire image-signal and, for a 2-D input, in order to reduce the time of computation, we could apply the only a 1-D transform: first on rows, then on columns or inversely - see [3] for example. Regarding our central concept of correlation we point out that for a symmetric kernel the spatial convolution is equivalent with a spatial correlation.

In [2] he said: «Very broadly speaking, the Fourier transform is a systematic way to decompose «generic» functions into a superposition of «symmetric» functions. These symmetric functions are usually quite explicit (such as a trigonometric function sin(nx) or cos(nx)), and are often associated with physical concepts such as frequency or energy. A suggestive example of such a unique decomposition is:we can always write a function f as the superposition f = fe +fo of an even function fe and an odd function fo, by the formulae:

fe(x) := [f(x) + f(−x)] / 2 and (3)
fo(x) := [f(x) - f(−x)] / 2 (4)

fe(x) and fo(x) are decorrelated components, which form an orthogonal decomposition. *The correlation and orthogonality are perfect antonyms*.

This concept of symmetry is intimately related with the concept of correlation, which is defined by the inner product over a vector space.

3. The inner product and the orthogonal decomposition. Unitary transforms

The *inner product* over a vector space permits the rigorous introduction of the notion of length of a vector. An inner product naturally induces an associated norm, thus an inner product space is also a normed vector space. A *complete space* (by definition, a *metric space* in which every Cauchy sequence is convergent) with an inner product is called a *Hilbert space*. The inner product also permits the rigorous introduction of the notion of the angle between two vectors. In particular, it also provides a high level abstract definition of the orthogonality between vectors: two orthogonal vectors have a zero inner product.

A *sequence space* is a vector space whose elements are infinite sequences of real or complex numbers. For us this is the space of sampled signals. The most important sequences spaces in functional / signal analysis are the ℓ^p spaces, consisting of the *p*-power summable sequences, with the *p*-norm. The ℓ^2 space, the space of the signals with finite energy – a sufficient condition for the existence of the Fourier transform -, has found to be the only sequence Hilbert space of this class. Theorem: any Hilbert space has an *orthonormal basis*: a basis in which all the elements are orthogonal and have unit norm. Why is so important to have an orthonormal basis? Because for a general inner product space *V*, an orthonormal basis can be used to define normalized *orthogonal coordinates* on *V*. Under these coordinates, the inner product becomes simple *dot product* of vectors. Thus the presence of an orthonormal basis reduces the study of a finite-dimensional inner product space to the study of $\mathbf{R}^n$ under dot product. This is a fundamental tool from the computational and also representational point of view. The big advantage is that through such an orthogonal decomposition, we decorrelate the components of the signal, having the freedom to analyze separately every such a component.

So, we know that for our space of signals exists at least one orthonormal / orthogonal basis. How to build it? Using the *Gram-Schmidt Process* we may start with an arbitrary basis and transform it into an orthonormal basis. But some orthogonal decompositions have exceptional properties and meanings: these are the *unitary transforms*, which preserve the energy or information contained in the signal. A unitary transform is defined by a *unitary matrix* A, i.e. with

$$A^{-1} = A^{*T} \qquad (5)$$

the inverse of a *unitary matrix* is equal to its conjugate transpose. The forward 1-D transform for an input vector v, representing the samples of the input signal is:

$$u = Av \qquad (6)$$

, and the inverse transform, i.e. the reconstruction of v, is:

$$v = A^{*T}u \qquad (7)$$

Conform [3], a deep meaning of a unitary transform could be extracted if we point out the columns of A^{*T} :

$$A^{*T} \equiv [\mathrm{A}_0, \mathrm{A}_1, \ldots \mathrm{A}_{\text{N-1}}] \tag{8}$$

Relation (4) becomes:

$$u(k) = < A_k, v >, \text{with k} = 0 \ldots \text{N - 1} \tag{9}$$

, where <.,.> is the dot product, and relation (5) becomes:

$$\mathrm{v} = \sum_{k=0}^{N-1} u(k) A_k \tag{10}$$

Therefore: the input vector (v) is decomposed into the basis formed by the columns (A_k) of A^{*T} - the coefficients in this expansion is even (u(k)) the components of the transformed vector; the coefficients are the dot products ($< A_k$, v>) between the mentioned columns and the input vector (v). This basis is orthonormal, conform (5).

The set of functions $f_n(x) = \exp(2\pi i n x)$, where $n \in \mathbf{Z}$ is an orthonormal basis for the Hilbertian space $L^2([0,1])$ (L^2 generalizes ℓ^2 from sequences to functions): here is the foundation of the Fourier analysis.

Similar with the 1-D case, for the 2-D case, the unitary transform of an image-signal represents the expression of this signal as a linear combination over some «basic» images.

The inner product expresses the resemblance between a (sampled) signal and a (sampled) pattern (see also the 'image matching' described in [5] and the explanation given in [6]).

Thus, until now, we sketch the big picture of the paper. Further we will use these "mise-en-scène", to fill our big picture.

4. The uncertainty principle in time–frequency representations

Observation: 'time–frequency domain' is the term used in the most of reference materials. For this reason, in this chapter, instead of 'spatial domain' we will use the term 'time domain'.

Let us briefly review the theory. A signal, as a function of time, may be considered as a representation with perfect time resolution.

In contrast, the FT of the signal may be considered as a representation with perfect frequency resolution, but with no time information because it does not convey when, in time, different events occur in the signal. Generally speaking, the more concentrated f(x) is, the more spread out its FT, F(u) must be. In particular, the scaling property of the Fourier transform may be interpreted as: if we «squeeze» a function in x, its FT «stretches out» in u. It is not possible to arbitrarily concentrate both a function and its Fourier transform.

The trade-off between the compaction of a function and its Fourier transform can be formalized in the form of an *uncertainty principle* which states that, if $f(x)$ is absolutely continuous and the functions $x \cdot f(x)$ and $f'(x)$ are square integrable, then the product between the (normalized) dispersion about zero in the time domain and the (normalized) dispersion about zero in the frequency domain is limited inferior by a constant:

$$f(t) \leftrightarrow F(\omega) \Rightarrow \sigma_t \sigma_\omega \geq const \,, \quad \text{with:} \tag{11}$$

$$\sigma_t^2 = \frac{\int t^2 |f(t)|^2 dt}{\int |f(t)|^2 dt}, \quad \text{and} \quad \sigma_\omega^2 = \frac{\int \omega^2 |F(\omega)|^2 dt}{\int |F(\omega)|^2 dt} \tag{12}$$

The equality is attained only f(t) is a Gaussian function.

Conform [4], in quantum mechanics, the momentum and position wave functions are like Fourier transform pairs: the above inequality becoming the statement of the Heisenberg uncertainty principle.

The *Short-Time FT* localizes the signal with a window function, before performing the FT [7]. A drawback of the STFT is that it has a fixed resolution – given by the width of the windowing function. A wide window gives better frequency resolution, but poorer time resolution. A narrower window gives better time resolution, but poor frequency resolution. To illustrate this we mention an experiment regarding STFT: observing the spectrograms of a signal consisting of two sinusoidal (first with a frequency of 10 Hz, second with a frequency of 25 Hz), for different width (T) of the windowing function:

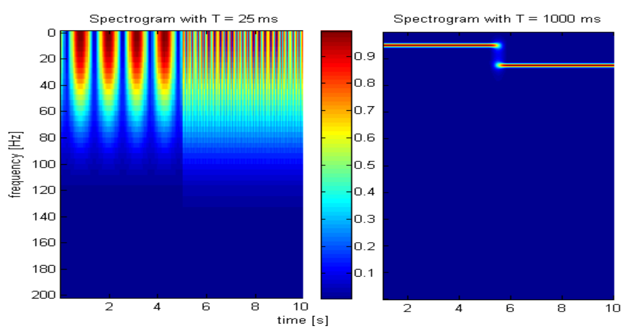

Fig. 2. Observing the spectrograms of a signal that could be described as a concatenation of two sinusoidal [8]

The intelligent compromise between the spatial resolution and frequency resolution is the wavelet transform (WT) [9], or, generalizing, the multiresolution analysis. The uncertainty principle was illustrated graphically, assigning to each basis function / signal used in the representation of a signal a tile in the time-frequency plane. The tile, also called *Heisenberg box or cell*, shows the frequency content of the basis function that it represents and where this function resides in time:

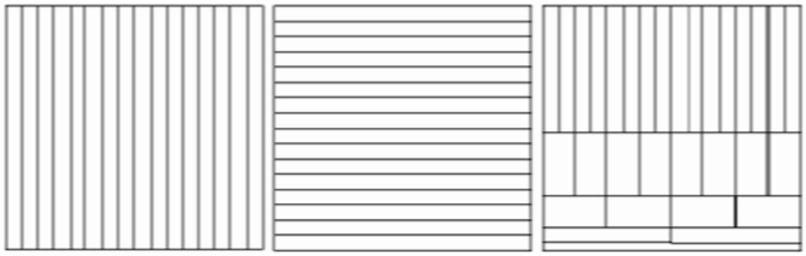

Fig. 3. The coverage [10] of the main types of basis: impulses (time domain), sinusoids (frequency domain) and WT (approximating / scaling + detail) functions

The area of each tile in the last representation is the same.WT is the good choice in the most practical cases: nonstationary signals – with frequencies varying in time. In plus, it is a more flexible representation, having the capability to obtain various shapes inside Heisenberg's uncertainty principle:

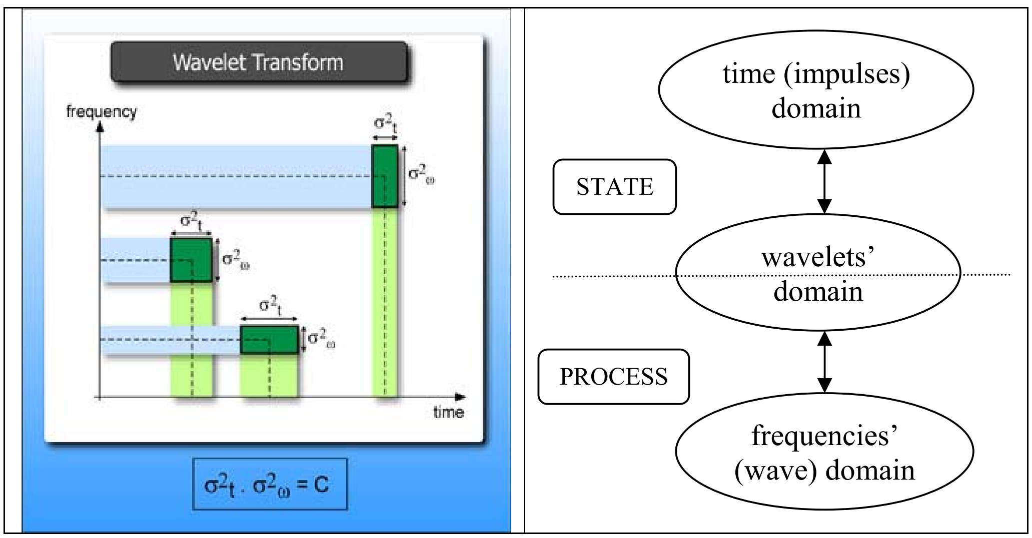

Fig. 4. a.The more flexible coverage of a signal [11] and b. an original view of the Wavelet representation as a common son of the time / spatial representation and frequency representation

5. The bi-orthogonal decomposition of a stochastic process: the Karhunen–Loève Theorem

There exist many orthonormal decompositions of a stochastic process: if the process is indexed over [a,b], any orthonormal basis of $L^2[a, b]$ yields such an expansion. As we have already mentioned the importance of the Karhunen–Loève theorem or transform (KLT) is that it provides the best basis in the sense that it minimizes the total mean square error.

If in a Fourier series the coefficients are real numbers and the expansion basis consists of sinusoidal functions, the coefficients in the KLT are random variables and the expansion basis depends on the process. In fact, the orthogonal basis functions used in this representation are determined by the *covariance* function of the process – it is like the KLT adapts to the process in order to produce the best possible basis for its expansion.

After Wikipedia [12], Karhunen–Loève Theorem: "A stochastic process $\{X_t\}t \in [a, b]$ with zero media / expectation ($E[X_t]$ (if not, we consider the process (X_t - $E[X_t]$)), satisfying some technical continuity conditions (easy to meet in reality), admits a decomposition

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t) \tag{13}$$

, where Z_k are pairwise uncorrelated random variables and the functions e_k are continuous real-valued functions on [a, b] that are pairwise orthogonal on $L^2[a, b]$. We could think that the expansion is bi-orthogonal since the random coefficients Z_k are orthogonal in the probability space, i.e. decorrelated, while the basis functions e_k are orthogonal in the time domain. Moreover, if the process is Gaussian, then the random variables Z_k are Gaussian and stochastically independent."

KLT is also called the principal component analysis, the Proper Orthogonal Decomposition (POD), or the Hotelling Transform (HT). It is associated with Principal Component Analysis (PCA) approach. Further we will present the effective process of the decomposition for a typical sampled signal [10].

Let a X (N x 1) vector be a multidimensional random variable representing some related characteristics of an object, for example the X (3 x 1) vector could be the representation (on 3 channels) of a certain pixel's color. We know that the autocovariance *matrix* of X, the same as its *autocorrelation matrix*, S_x (N x N), is real and symmetric, diagonalizable to:

$$S_X = Q^T S_Y Q \tag{14}$$

with Q(N x N) built with the eigen(and orthonormal) vectors of S_X,

$S_Y = diag(v_1, v_2, \dots v_N)$, i.e. having non-zero elements only on the principal diagonal – these are the S_Y's eigenvalues, the same as S_X's eigenvalues. The HT is:

$$Y = Q(X - M_X) \tag{15}$$

where M_X is the average vector for X and Q is the above matrix, in which eigenvectors were sorted after the decreasing of the corresponding eigenvalues and inserted in Q (on the first row is the eigenvector corresponding to the highest eigenvalue). This transform produces a random variable with remarkable properties:

1. zero mean value: $M_Y = 0$
2. decorrelated components: the covariance matrix is S_Y
3. reconstruction:

$$X = Q^T Y + M_X \tag{16}$$

If we do not use the entire S_X to form the HT matrix – suppose that from N we use K lines, the reconstruction will be approximate:

$$\boxed{\hat{X} = {Q_K}^T Y + M_X} \tag{17}$$

They have shown that the 'mean square error' is:

$$e_{ms} = \sum_{i=1}^{N} v_i - \sum_{i=1}^{K} v_i = \sum_{i=K+1}^{N} v_i \tag{18}$$

Thus, we could have a better approximation if we select the top lines. For this reason we sorted the eigenvalues.The KLT has the optimum Energy concentration property: no other unitary transform packs as much energy into the first *J* coefficients, for any *J*! For any unitary transform, the inverse transform can be interpreted in terms of the superposition of «basis images». With KLT transform, the basis images, which are the eigenvectors of the autocorrelation matrix of the stochastic process, are called *eigenimages*. If energy concentration works well, only a limited number of eigenimages is needed to approximate well the input set of images.

KLT is rarely used in practice in its original form, because it is not separable transform, therefore with an $O(N^4)$ complexity. In practice we could use suboptimal approximations for KLT, with efficient implementation – the most used is Discrete Cosinus Transform that works well for high correlated signals.

We could refine KLT in order to find a more general solution for a powerful decomposition and decorrelation of a signal, knowing his history (probability). To recognize complex patterns (e.g., faces), large portions of an image (say of size *MN*) might have to be considered. High dimensionality of «image space» means high computational burden. Example: *nearest-neighbor search* requires pairwise comparison with every image in a data base. The solution is to reduce the dimensionality from *MN* to J, by tailoring KLT to the *specific* set of images, so that the recognition task to preserve the salient features.

After [13], let us suppose a concrete pattern recognition task: detecting in a passport photo if the person is male or female. For J = 8, MIT Media Lab obtained:

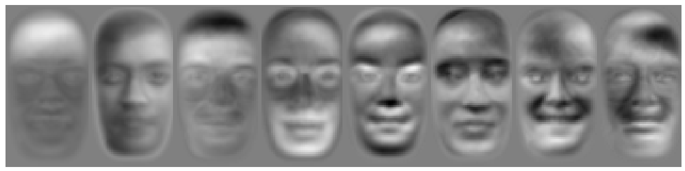

Fig. 5. The eigenfaces (eigenimages) obtained from a training set of 500 frontal views conform [13]

Can be used the above basis for face recognition by nearest neighbor search in 8-D «image face»? Or, can we use it to generate «sufficient» faces by adjusting 8 coefficients?

Eigenimage method maximizes the «scatter» within the linear subspace over the entire image set, regardless of classification task. Linear discriminant analysis (LDA) and the related Fisher's linear discriminant (FLD) are methods that maximize between-class scatter, while minimizing within-class scatter, taking in account the *generalized eigenvectors*, i.e. corresponding to the largest eigenvalues. Minimizing within-class scatter address also the problem when differences due to varying illumination can be much larger than differences between faces. In [13] they illustrated comparatively the two approaches:

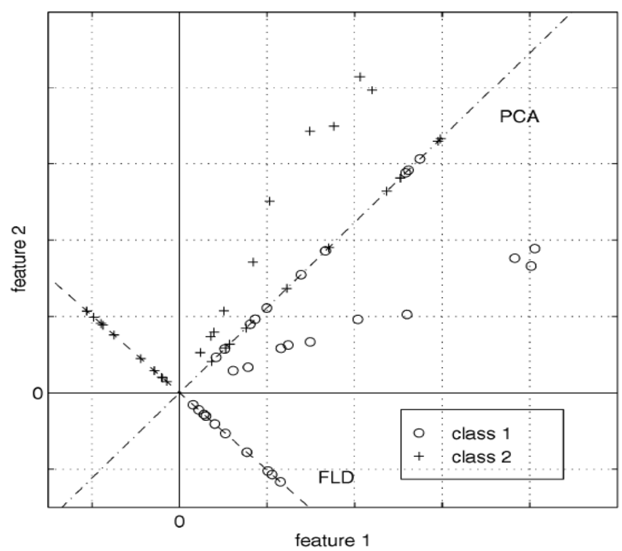

Fig. 6. 2-D example conform [13]: Samples for 2 classes are projected onto 1-D subspace using the PCA (KLT) or Fisher FLD (LDA). PCA preserves maximum energy, but the 2 classes are no longer distinguishable. FLD separates the classes by choosing a better 1-D subspace.

6. Conclusions

We mention some intimately related terms grouped from different points of view:

- linear algebra: inner product over a vector space, symmetric and separable kernel, orthogonal basis, projection and decomposition, factorization, eigenvalues & eigenvectors, unitary transforms;
- Heisenberg's uncertainty principle: state-process, position (or time)-frequency, *incompatibility principle* in fuzzy theory;
- probability theory: variance, covariance, correlation (= normalized covariance),
- image processing and analysis: spatial correlation (= mirrored convolution), dot product in frequency domain, template matching, pattern recognition, principal component analysis and so on.

Fig.1, Fig.4.b. and Fig.8. are original synthetical images, together sketching an abstract view on signal analysis. This paper could be seen as a scientific-mathematical view to the pattern recognition problem, in the classical representation domains of a signal. We address «the essence» of the general pattern recognition problem: recognition means correlation. A general observation, also from experimental results of a large category of researchers, is that the performance of a method comes with the adaptability of the method to the concrete problem space. The KLT is a prototype from this point of view. Given the geometrical interpretation of a unitary transform as a rotation of the coordinate axes, we could think KLT as the unitary transform by which we represent the input form / signal in his eigenvectors' coordinate system. Working in this system is very useful in the pattern recognition problem.

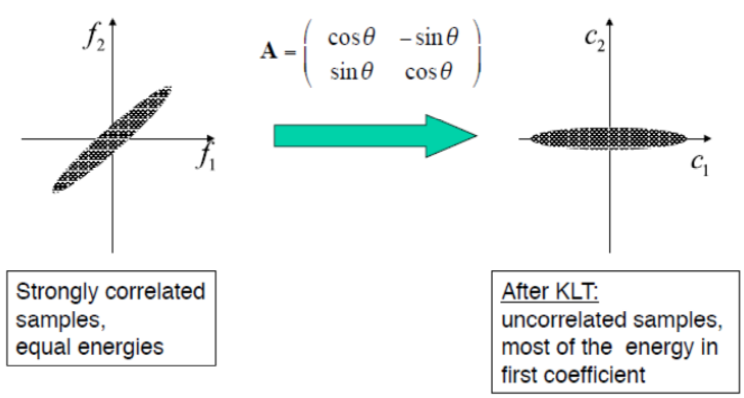

Fig. 7. KLT as a transform in the eigenvectors' coordinate system [13]

For almost two decades the term “adaptive research” has coined. It refers a research where one assesses the adaptivity of something. The practical problems from which emerged this term were related especially with clinical research and land use projects. From the first point of view, they used Bayesian methods: <<One of the most exciting benefits of Bayesian data analysis is being able to evaluate data ”on the fly”, as they are being collected, and decide whether or not to continue data collection and how to optimize the experimental treatment for the next observation. Bayesian adaptive research design can be especially helpful in clinical applications, when experimental treatments with null or detrimental effects should be discontinued as quickly as possible, and treatments with clearly beneficial effects should be disseminated as quickly as possible>> ([14] and [15]). From the second point of view "adaptive research programs must be directed to investigate the actual and real problems associated with the planning, design, implementation and management of land use projects. It is important that the resulting methodology to be technically feasible, environmentally and economically viable and socially acceptable” [16].

Most of all, the contemporary pattern recognition problems are complex classification problems, with uncertain border between classes (they could overlay one another). The theoretic framework for this kind of problems is the fuzzy theory. The theory of possibility was introduced by Zadeh [17] as a generalization of the theory of probability; further the Bart Kosko studies [18] proved this using as membership function the proportion in which a set could be considered a subset of another set.

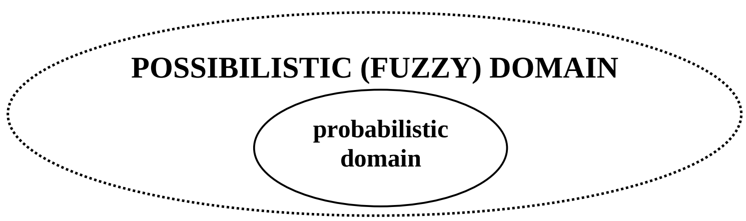

Fig. 8. KLT as a transform in the eigenvectors’ coordinate system [13]

The “fuzzy” domain has become classic since his logic entered successfully not only in the official scientific language, but also in industry. If for the previous domains we talked about the orthogonal decomposition, in the “fuzzy” domain we talk about graduation and granulation – here, the partitions / classes / primitives could overlap each other in a certain degree. We could say that, if the probabilistic domain express a “statistic adaptation” to the problems solution space, the “fuzzy” approach express a “possibilistic or intuitionistic adaptation” to the same space. Further, we consider it would be interesting the analysis of the concept of correlation in fuzzy domain.

R E F E R E N C E S

[1] *http://en.wikipedia.org/wiki/LTI_system_theory*
[2] http://www.math.ucla.edu/~tao/preprints/fourier.pdf
[3] *Constantin Vertan and Mihai Ciuc,* "Tehnici fundamentale de prelucrarea şi analiza imaginilor", ed. MatrixRom, 2007
[4] *Stein Elias and Rami Shakarchi*, Fourier Analysis: An introduction, ed. Princeton University Press, 2003
[5] *Mihai Bulea*, Prelucrarea imaginilor şi recunoaşterea formelor (teorie şi aplicații), ed. Academiei Române, 2003, pg. 161-162
[6] *Beatrice Pesquet-Popescu and Sorin Popescu*, Analiza cu undine (wavelets). Fractali – modele stocastice, ed. Politehnica Press, 2010, pg. 12-13
[7] *Beatrice Pesquet-Popescu and Sorin Popescu*, Analiza cu undine (wavelets). Fractali – modele stocastice, ed. Politehnica Press, 2010, pg. 13-15
[8] http://en.wikipedia.org/wiki/Short-time_Fourier_transform
[9] *Beatrice Pesquet-Popescu and Sorin Popescu*, Analiza cu undine (wavelets). Fractali – modele stocastice, ed. Politehnica Press, 2010, pg. 15-23
[10] *Rafael C. Gonzalez and Richard E. Woods*, ed. Digital Image Processing (3rd Edition) (Hardcover), ed. Prentice Hall, 2008, pg. 501
[11] http://lumenosys.com/wavelet-transforms.html
[12] http://en.wikipedia.org/wiki/Karhunen%E2%80%93Lo%C3%A8ve_theorem
[13] *http://www.stanford.edu/class/ee368/Handouts/10-EigenImagesMay25_27.pdf*
[14] *Michael J. Rosenberg*, “The Agile Approach to Adaptive Research: Optimizing Efficiency in Clinical Development”, ed. Wiley, 2010
[15] *John K. Kruschke*, "Doing Bayesian Data Analysis. A Tutorial with R and BUGS", ed. Academic Press / Elsevier, 2011
[16] *Hubert N. van Lier*, "Land Use Planning: A Key to Sustainable Development", XXX International Symposium Actual Tasks on Agricultural Engineering, Croatia, 2002
[17] *Lotfi A.of. Zadeh*, “Fuzzy sets as a basis for a theory of possibility”, Fuzzy Sets and Systems, **Vol. 1**, pp.3-28.
[18] *Bart Kosko*, “Neural Networks and Fuzzy Systems”, ed. Prentice Hall, Englewood Cliffs, NJ