

Reconhecimento de Padrões Numéricos Através da Transformada Karhunen-Loève

JOÃO FERNANDO MARAR^{1,2}
EDSON COSTA DE BARROS CARVALHO FILHO²

¹UNESP—Universidade Estadual Paulista
Faculdade de Ciências - Depto de Computação
Bauru - São Paulo - Brasil

²UFPE—Universidade Federal de Pernambuco
Departamento de Informática - CCEN
Cx 7851, 50.732-970, Recife, PE, Brasil
{jfm,ecdbcf}@di.ufpe.br

Abstract. The merit of the Karhunen-Loève transform is well known. Since its basis is the eigenvector set of the covariance matrix, a statistical, not functional, representation of the variance in pattern ensembles is generated. By using the Karhunen-Loève transform coefficients as a natural feature representation of a character image, the eigenvector set can be regarded as a feature extractor for a classifier.

1 Introdução

A visão constitui um poderoso sistema de comunicação humana. Uma simples imagem engloba uma enorme quantidade de informações, onde a grande maioria dessas informações são difíceis de serem expressadas com igual concisão, precisão e flexibilidade através de mecanismos artificiais. O cérebro humano é capaz de extrair características essenciais de uma imagem em um simples relance, registrando o específico padrão, bem como, ele consegue reconhecer padrões mesmo na presença de ruídos, transformações geométricas, movimento, entre outros.

A ciência por sua vez, procura meios de desvendar os mistérios cerebrais associados a visão e outras habilidades humanas, com a finalidade de construir dispositivos artificiais que realizem tais tarefas com tamanha perfeição. Durante as últimas décadas, as áreas de *reconhecimento de padrões e processamento de imagens* têm dedicado muita pesquisa na elaboração de modelos matemáticos para o processamento, descrição e classificação de imagens [COO 65],[OPP 75], [PRO 88],[GON 92].

Embora os estudos sobre reconhecimento de padrões e processamento de imagens tenham se desenvolvido em áreas distintas, existe uma forte relação entre elas. A área de processamento de imagens consiste não somente em codificar, filtrar, ampliar e restaurar, mas como também em realizar análise e reconhecimento de imagens. Da mesma forma, a área de reconhecimento de padrões incluem não somente a extração de características e classificação, mas

também o pré-processamento e a descrição de padrões.

O objetivo deste artigo é apresentar um método para reconhecimento de caracteres numéricos em uma retina de pontos através de uma técnica associada as áreas de reconhecimento de padrões e processamento de imagens. Na realidade, o método apresentado trata-se de uma técnica matemática que possibilita a extração de um conjunto de características mensuráveis das imagens (padrões) que serão tratadas através de um classificador linear, onde um conjunto de padrões será apresentado para o reconhecimento. A base desta técnica reside na utilização da transformada Karhunen-Loève [CAM 71],[STA 88],[CHE 91], que corresponderá a parte de extração das principais componentes, onde cada imagem (padrão) é representado por um vetor de características e o reconhecimento do padrão é realizado através de medidas realizadas no espaço de características das classes existentes.

Nas seções que se seguem, trataremos de maneira mais geral sobre transformações do espaço de características, transformadas ortogonais, conjunto de dados para a realização do experimento, descrição do método e as conclusões.

2 Transformações do espaço de características

A idéia central desta seção é a transformação do espaço original dos padrões para outro com dimensão menor. De maneira que, esta nova representação dos

padrões fique livre de informações redundantes, não necessárias para o reconhecimento. Em geral, para se obter uma boa transformação, adota-se um critério de otimização, entre eles destacam-se o *mínimo erro quadrado* e a *entropia*. Estes critérios associados as transformações tem por objetivo auxiliar a discriminação das classes existentes, de maneira a maximizar as distâncias inter-classes e a minimização das distâncias intra-classes.

Dentre as transformações mais utilizadas, destacamos a transformada de *Fourier* [COO 65],[OPP 75],[BLA 87] a transformada *Walsh-Hadamard* [HUT 75],[GON 92], a transformada *Haar*, a transformada *Karhunen-Loève* e mais recentemente as *Wavelets* [DAU 88], [MAL 89],[DAU 90].

A transformada Karhunen-Loève (KLT) é ótima em relação ao critério de mínimo erro quadrado. Este método conhecido como análise das principais componentes tem sido muito utilizado em aplicações onde a redução de dimensionalidade, torna-se necessária.

As transformações mencionadas acima são denominadas transformadas ortogonais, na seção seguinte, as descreveremos de maneira mais técnica.

3 Transformadas Ortogonais

As transformadas ortogonais constituem uma poderosa ferramenta matemática para representar funções extremamente complexas através de outras mais simples. Imagens podem ser vistas como funções. Uma importante aplicação desta técnica é a redução de dimensionalidade ou compressão de dados [AHM 75],[WOM 77]. Estas transformadas podem ser divididas em duas classes, segundo as funções básicas utilizadas. Desta forma, temos a classe baseada em funções não senoidais e a de funções senoidais, tendo esta última como única representante a transformada de Fourier.

Em geral, aplica-se esta técnica a um conjunto de dados obtidos experimentalmente, cujo objetivo é encontrar uma lei de formação para a função. Detalhes sobre a representação de funções através desta técnica serão expostas a seguir, bem como apresentaremos a transformada Karhunen-Loève, como uma das inúmeras representante da classe das transformadas ortogonais que são definidas através funções básicas não senoidais.

3.1 Extração de Características via Transformadas Ortogonais

Seja $\Psi = \{\psi_0(t), \psi_1(t), \dots, \psi_n(t), \dots\}$ um conjunto de funções reais e contínuas, (utilizamos funções reais por conveniência), será dito ortogonal no intervalo

$(t_0, t_0 + T)$ se:

$$\int_{t_0}^{t_0+T} \psi_m(t)\psi_n(t)dt = \begin{cases} c_n & \text{se } m = n \\ 0 & \text{se } m \neq n \end{cases} \quad (1)$$

Para o caso onde $c_n = 1$, o conjunto Ψ é chamado ortonormal. Seja $x(t)$ uma função de valores reais, definida em um intervalo (t_0, t_0+T) , e suponha que $x(t)$ possa ser escrita na forma :

$$x(t) = \sum_{n=0}^{\infty} a_n \psi_n(t) \quad (2)$$

então os coeficientes a_n podem ser obtidos da seguinte forma : multiplicamos ambos os lados da equação 3 por ψ_m e integramos o resultado no intervalo $(t_0, t_0 + T)$, onde obteremos :

$$\int_{t_0}^{t_0+T} x(t)\psi_m(t)dt = \int_{t_0}^{t_0+T} \sum_{n=0}^{\infty} a_n \psi_n(t)\psi_m(t)dt \quad (3)$$

Como ψ_m e ψ_n são ortogonais, temos :

$$a_m = \frac{1}{c_n} \int_{t_0}^{t_0+T} x(t)\psi_m(t)dt, \quad m = 0, 1, 2, \dots$$

Um conjunto de funções ortogonais é chamado fechado ou completo se for verificada a seguinte condição:

Para qualquer parte contínua de $x(t)$ com :

$$\int_{t_0}^{t_0+T} x^2(t)dt < \infty$$

Qualquer que seja $\epsilon > 0$, existe N tal que seja possível representar uma aproximação de $x(t)$ por uma expansão finita :

$$\int_{t_0}^{t_0+T} |x(t) - \hat{x}(t)|^2 dt < \epsilon$$

onde :

$$\hat{x}(t) = \sum_{n=0}^{N-1} a_n \psi_n(t)$$

Pelo desenvolvimento acima, é visível que por uma expansão em funções ortogonais, sempre será possível representar $x(t)$ por um conjunto infinito, mas enumerável $\{a_0, a_1, a_2, \dots\}$. Entretanto, quando Ψ for completo torna-se possível uma aproximação de $x(t)$ através de um conjunto finito $\{a_0, a_1, \dots, a_{N-1}\}$.

3.2 Extração de Características via Transformadas Karhunen-Loève

Seja $\{X\}$ um conjunto de vetores, obtidos por amostragem, de uma classe de sinais aleatórios, podendo

ser um conjunto de imagens. Um representante de $\{X\}$ é dado por $x_j = (x_{j,1}, x_{j,2}, \dots, x_{j,N})$. A amostra x_j pode ser aproximada por 4:

$$x_j = y_{j,1}\psi_1 + y_{j,2}\psi_2 + \dots + y_{j,K}\psi_K = \sum_{i=1}^N y_{j,i}\psi_i \quad K < N \quad (4)$$

$$y_{j,i} = x_j^t \psi_i \quad i = 1, 2, \dots, K \quad (5)$$

onde N é o número total de componentes da amostra e K é o número de componentes utilizadas na aproximação.

Por definição, o mínimo erro quadrado, ϵ , é dado pela expressão 6:

$$\epsilon = \left(\sum_{i=1}^N y_{j,i}\psi_i - \sum_{i=1}^K y_{j,i}\psi_i \right)^2 = \sum_{i=K+1}^N \psi_i^t R_X \psi_i \quad (6)$$

onde R_X corresponde a matriz de covariância do conjunto $\{X\}$. Dada por $R_x = \frac{1}{V} \sum_{j=1}^V (x_j - \bar{X})(x_j - \bar{X})^t$, onde V representa o número total de elementos do conjunto $\{X\}$ e \bar{X} é o vetor médio do referido conjunto.

Quando $\{\psi_i\}$ constituem a base ortogonal de Karhunen-Loève, os elementos ψ_i são determinados a partir dos autovetores de R_X , de acordo com a equação 7:

$$R_X \psi_i = \lambda_i \psi_i \quad (7)$$

onde λ_i são os autovalores da matriz de covariância.

De maneira que, o erro de truncamento da equação 6 é minimizado pela equação 8

$$\text{Min}_{\{\psi_i\}} \epsilon = \sum_{i=K+1}^N \lambda_i \quad (8)$$

Isto significa que, se utilizarmos apenas K autovetores para a representação de funções, o erro de truncamento será a soma dos auto-valores da matriz de covariância, sendo dado pela equação 8. A equação 4, escrita em termos dos auto-vetores da matriz de covariância, é denominada expansão Karhunen-Loève. A correspondente transformação ortogonal inversa, na equação 5, é chamada transformada Karhunen-Loève (K-L) [CHE 91].

4 Conjunto de Dados Experimentais

Para investigarmos as propriedades fundamentais e a qualidade do método proposto, utilizamos um conjunto de dados contendo informações retiradas do mundo real. A base de dados do *British Post Office* possui um conjunto de dados alfanuméricos extraídos de envelopes de cartas. Este conjunto constitui-se de 34 classes de padrões, *i.e.*, caracteres alfabéticos de A-Z e caracteres numéricos de 0-9.

Cada classe contém 300 amostras de padrões digitalizados e normalizados através de 384 pontos em preto e branco, sobre uma matriz de 24 linhas e 16 colunas. Nosso experimento baseou-se nas classes constituídas dos caracteres numéricos de 0-9, totalizando 3.000 padrões divididos em 10 classes. A figura 1 ilustra alguns dos melhores padrões existentes na base de dados.

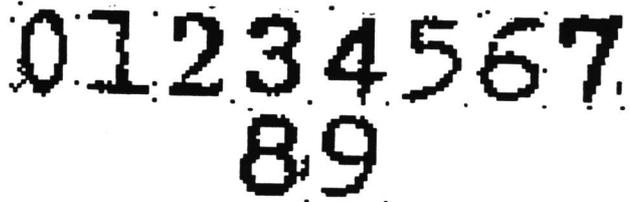


Figura 1: Exemplos de numerais utilizados para o processo de reconhecimento de padrões.

Em nosso experimento, utilizamos 128 padrões de cada uma das 10 classes para o treinamento do classificador, perfazendo um total de 1280 padrões para treinamento.

Por outro lado, selecionamos 100 padrões de cada uma das 10 classes (1000 padrões) para testarmos a eficiência do método proposto. Vale ressaltar, os padrões utilizados para o treinamento e teste são todos distintos entre si, e foram selecionados de maneira aleatória.

5 Descrição do Método

Como sabemos, o estudo de reconhecimento de padrões pode ser dividido em duas áreas principais: *extração de características* e *projeto do classificador*, como podemos observar na figura 2.

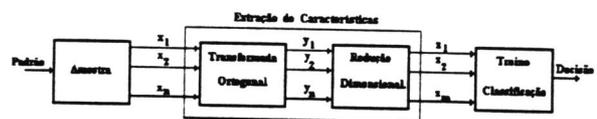


Figura 2: Representação gráfica do processo de Reconhecimento de Padrões.

5.1 Processo de Amostragem

O processo de amostragem, basicamente, fornece uma sequência de valores que definem o padrão dado. Após este processo, o padrão de entrada é tratado como um vetor N -dimensional.

$$\text{padrao} = (x_1, x_2, \dots, x_n)$$

5.2 Processo de Extração de Características e Redução de Dimensionalidade

Obtido o vetor, correspondente ao padrão de entrada, inicia-se o processo de extração de características. O primeiro estágio, consiste em uma transformação ortogonal dos valores de entrada (o padrão amostrado). Neste trabalho, utilizamos a transformada K-L, esta provê um mapeamento um para um, transformando a sequência de entrada $X, (x_1, x_2, \dots, x_n)$, em outra $Y, (y_1, y_2, \dots, y_n)$, de mesma dimensão.

Na prática, a dimensão dos padrões, N , é muito grande, sendo necessário sua redução. Por exemplo, neste trabalho, a dimensão dos padrões é $N = 384$. Utilizando-se das propriedades inerentes da transformada K-L, reduziremos a dimensionalidade. Desta forma, o segundo estágio do processo de extração de características é a escolha das componentes mais significativas do padrão de entrada. A saída deste módulo será um vetor $Z, (z_1, z_2, \dots, z_K)$, com K componentes de Y , tal que $K \ll N$. Esta redução de dimensionalidade é alcançada segundo os critérios que norteiam a transformada Karhunen-Loève, de maneira que o erro seja tão pequeno quanto se deseja e facilmente calculado através da equação 8.

Neste experimento, a matriz de covariância, é dada por :

$$R_x = P_1 R_{x_1} + P_2 R_{x_2} + \dots + P_{10} R_{x_{10}} \quad (9)$$

onde R_{x_k} é a matriz de covariância da k -ésima classe; P_k denota a probabilidade de ocorrência da classe C_k , assumimos $\frac{1}{10}$ para cada classe. A distribuição de variância é calculada em R_x . A transformada Karhunen-Loève é dada pelo conjunto dos autovetores extraídos da matriz de covariância, R_x , a qual denotamos por T .

A transformação de um padrão amostrado é dado por:

$$Y = TX$$

Fixando as K primeiras componentes de Y , estas estão associadas as maiores variâncias, obteremos o vetor Z denotado por (z_1, z_2, \dots, z_K) .

O passo seguinte, consiste em classificar o padrão de entrada. Na realidade, o processo de classificação consiste inicialmente de um treinamento e posteriormente a classificação. A classificação é somente uma tomada de decisão, que nos informa qual a classe a que pertence o padrão dado.

5.3 Projeto do Classificador

Nosso desejo é construir e treinar um classificador que seja capaz de localizar automaticamente a classe a que pertence um determinado padrão Z , escrito

através das funções básicas K-L e com dimensionalidade reduzida.

Projetaremos um classificador baseado no critério de mínima distância quadrada, este é definido a partir da seguinte regra:

Um padrão Z é dito pertencente a classe C_i se Z é mais próximo a \bar{Z}_i . onde C_i corresponde a classe i e \bar{Z}_i é o vetor médio da classe C_i .

A figura 3 ilustra o processo de classificação para três classes (C_1, C_2, C_3). Neste caso, o padrão Z será reconhecido como elemento pertencente a classe C_2 .

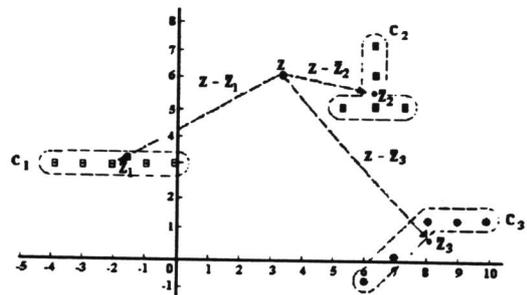


Figura 3: Representação geométrica de um classificador de mínima distância quadrada para três classes

Denotaremos por D_i^2 a distância quadrada entre Z e os \bar{Z}_i s. A fórmula para o cálculo da distância é dada por :

$$D_i^2 = |Z - \bar{Z}_i|^2 = (Z - \bar{Z}_i)(Z - \bar{Z}_i)^t \quad i = 1, \dots, r \quad (10)$$

onde Z é o padrão a ser classificado e \bar{Z}_i é o vetor médio da classe i , r é a quantidade de classes.

Reescrevendo a equação 10, temos :

$$D_i^2 = |Z|^2 - 2(\bar{Z}_i Z^t - \frac{1}{2} |\bar{Z}_i|^2) \quad i = 1, \dots, r \quad (11)$$

Observando a equação acima, D_i^2 é mínimo quando o termo $(\bar{Z}_i Z^t - \frac{1}{2} |\bar{Z}_i|^2)$ é o máximo. Assim, o classificador baseado na mínima distância deve calcular as distâncias do padrão para cada uma das classes, isto é, calcular D_i^2 da equação 11. Note que não é necessário calcular toda a expressão; basta apenas calcular o termo $(\bar{Z}_i Z^t - \frac{1}{2} |\bar{Z}_i|^2)$. Desta maneira, podemos descrever o classificador pela seguinte função de discriminação:

$$g_i(Z) = \bar{Z}_i Z^t - \frac{1}{2} |\bar{Z}_i|^2 \quad i = 1, \dots, r \quad (12)$$

onde o termo $\frac{1}{2} |\bar{Z}_i|^2$ é conhecido por limiar.

O classificador calculará "r" valores reais $g_1(Z)$, $g_2(Z), \dots, g_r(Z)$ como mostra a figura 4, e os compara. De maneira que, Z pertence a classe C_i quando $g_i(Z)$ for maior valor encontrado entre os $g_i(Z)$ para $i = 1, 2, \dots, r$. A implementação geral para este tipo de classificador é ilustrado na figura 4. Mostrando para r -classes.

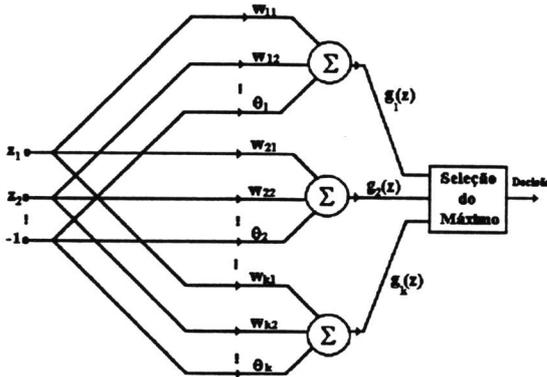


Figura 4: Classificador de mínima distância quadrada para k classes

Em nosso experimento, utilizamos para este classificador de 10-classes a seguinte função discriminadora :

$$g_i(Z) = \bar{Z}_i Z^t - \frac{1}{2} | \bar{Z}_i |^2 \quad i = 1, 2, \dots, 10 \quad (13)$$

onde \bar{Z}_i é o vetor padrão médio da classe C_i . Dado que $Z = (z_1, z_2, \dots, z_k)$, a função discriminação assume a seguinte forma :

$$g_i(Z) = w_{i,1}z_1 + w_{i,2}z_2 + \dots + w_{i,k}z_k - \theta_i, \quad i = 1, 2, \dots, 10$$

onde $\theta_i = \frac{1}{2} | \bar{Z}_i |^2$, $(w_{i,1}, w_{i,2}, \dots, w_{i,k})$ é o padrão de entrada. A implementação desta técnica esta baseado na figura 4

O classificador acima foi treinado várias vezes com números diferentes de padrões por vez. Realizamos o treinamento para 1, 2, 4, 8, 16, 32, 64, 128 padrões e testamos o classificador para cada treinamento com outros 100 padrões. Ambos os conjuntos, o de treino e o de teste foram escolhidos aleatoriamente da base de numerais. Obtendo os seguintes resultados ilustrados nas figuras 5.

6 Resultados Obtidos

Como podemos observar na figura 5, o processo de reconhecimento, inicia-se com uma pequena margem

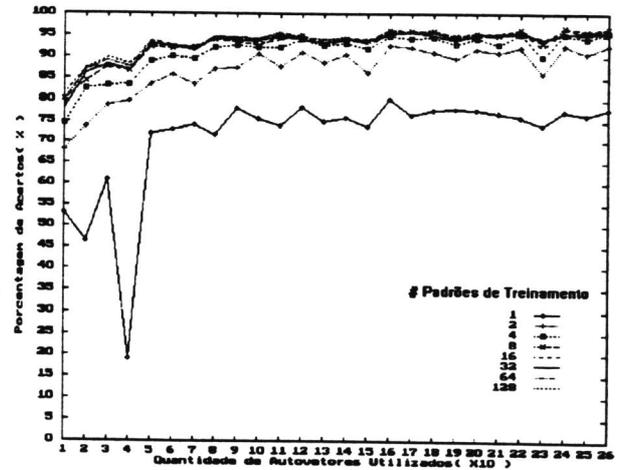


Figura 5: Gráfico do reconhecimento : Classificação correta versus o número de autovetores utilizados, as curvas correspondem a quantidade de padrões utilizados no treinamento.

de classificações corretas, quando o classificador é treinado com um único padrão.

Incrementando-se o número de padrões de treino para 2 e 4, os resultados associados a estes começam a melhorar. A partir de 8 até 128 padrões de treinamento, a performance do classificador não apresenta resultados significativos (em termos de melhorias de classificações corretas), aumentando apenas o custo computacional. O melhor resultado conseguido, foi justamente o treinamento com 8 padrões por classe e utilizando-se 250 autovetores, obtendo-se 96.80% de classificação correta. Por outro lado, podemos observar na figura 5, que a partir da utilização do centésimo autovetor o classificador começa a estabilizar num patamar acima de 94,78% de reconhecimentos corretos, para os treinamentos realizados com 8 até 128 padrões.

7 Conclusões

O trabalho demonstrou a utilização da transformada Karhunen-Loève como uma boa ferramenta para a extração e redução de características de padrões numéricas. A metodologia empregada, pode ser generalizada para padrões quaisquer em uma retina de pontos.

A transformada Karhunen-Loève foi aplicada a um conjunto de dados, com o objetivo de investigar sua adequabilidade e desempenho. Nos experimentos realizados, a técnica apresentou níveis de reconhecimento da ordem de 96.80%, demonstrando assim, sua viabilidade no processo de extração de

características, redução de dimensionalidade em sistemas de reconhecimento de padrões.

Agradecimentos

Os autores agradecem o suporte financeiro da CAPES-PICD, CNPQ e FACEPE.

Ao grupo de Redes Neurais -DI/UFPE-, em especial : Fred, Herman e Patricia pelo companheirismo, não trivial, nas horas de sufoco.

A grande Najla Rox Uzeizi Marar, minha mãe, pelo grande incentivo ao meu trabalho.

8 Referências

- [AHM 75] N. Ahmed, P.J. Milne, S.G. Harris, *Electrocardiographic Data Compression Via Orthogonal Transforms*, IEEE Transactions on Biomedical Engineering, vol.bme-22,Nro 6, November 1.975, p 484-487.
- [BLA 87] R.E. Blahut, *Fast Algorithms for Digital Signal Processing*, Addison-Wesley Publishing Company, Reading, MA, 1.987.
- [CAM 71] S.J. Campanella, G.S. Robinson, *A Comparison of Orthogonal Transformations for Digital Speech Processing*, IEEE Transactions on Communication Technology, Vol.Com-19, Nro 6, December 1.971, p 1045-1049.
- [CHE 91] C.S. Chen, K.S. Huo, *Karhunen-Loève Method for Data Compression and Speech Synthesis*, IEE Proceedings-I, Vol. 138, Nro. 5, October 1.991, p 377-380.
- [COO 65] J.W. Cooley, J.W. Tukey, *An Algorithm for the Machine Calculation of Complex Fourier Series*, Mathematics of Computation, Vol 19, Nro 90, Publication National Academy of Sciences - National Reserch Center by The Americam Mathema tical Society, April 1.965, p 297-301.
- [DAU 88] I. Daubechies, *Orthonormal Basis of Compactly Supported Wavelets*, Comm. Pure Applied Math., 1.988, Vol. 41, p 909-996.
- [DAU 90] I. Daubechies, *The Wavelets Transform, Time-Frequency Localization and Signal analysis*, IEEE Transactions on Information Theory, September 1.990, Vol. 36, Nro. 5, p 961-1005.
- [GON 92] R.C. Gonzales, R. E. Woods, *Digital Image Processing*, Addison Wesley Publishing Company, Inc, New York, 1.992.
- [HUT 75] B.A. Hutchins, *Applications of a real-time Hadamard Transform Network to sound synthesis*, Journal of Audio Engineering Society, Vol 23, Nro 7, September 1975, p 558-562.
- [MAL 89] S.G. Mallat, *A Theory for Multiresolution Signal Decomposition : The Wavelet Representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1.989, Vol. 11, Nro. 7, p 674-693.
- [OPP 75] A.V. Oppenheim, R.W. Schafer, *Digital Signal Processing*, Prentice-Hall international, Inc., London, 1.975.
- [PRO 88] J.G. Proakis, D.G. Manolakis, *Introduction to Digital Signal Processing*, Macmillan Publishing Company, New York, 1.988.
- [STA 88] J.C. Stapleton, S.C. Bass, *Synthesis of Musical Tones Based on the Karhunen-Loève Transform*, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 36, Nro. 3, March 1.988, p 305-319.
- [WOM 77] M.E. Womble, J.S. Halliday, S.K. Mitter, M.C. Lancaster, J.H. Triebwasser, *Data Compression for Storing and Transmitting*, Proceedings of the IEEE, Vol 65, Nro 5, May 1.977, p 702-706.