

Marcelo Gonella Fernandez

**Tratamento e Compressão baseada em Wavelets
para Dados Adquiridos por Sensores**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Departamento de Informática da PUC-Rio como parte dos requisitos parciais para obtenção do título de Mestre em Informática.

Orientador: Prof. Marco Antonio Casanova

Rio de Janeiro
Setembro de 2007



Marcelo Gonella Fernandez

**Tratamento e Compressão baseada em Wavelets
para Dados Adquiridos por Sensores**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Marco Antonio Casanova

Orientador

Departamento de Informática – PUC-Rio

Prof. Antonio Luz Furtado

Departamento de Informática – PUC-Rio

Profª. Karin Koogan Breitman

Departamento de Informática – PUC-Rio

Prof. José Eugênio Leal

Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 13 de Setembro de 2007

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Marcelo Gonella Fernandez

Graduou-se em Informática na PUC-Rio (Pontifícia Universidade Católica) em 2004. Atua como Analista de Sistemas e Programador no desenvolvimento de soluções para a área de Exploração e Produção (E&P) da Petrobras S.A. desde 2001. Possui interesse nas áreas de Engenharia de Software, Orientação a Objeto, Teoria da informação e Descoberta de conhecimento em Bancos de Dados (KDD).

Ficha Catalográfica

Fernandez, Marcelo Gonella

Tratamento e compressão baseada em wavelets para dados adquiridos por sensores. / Marcelo Gonella Fernandez; orientador: Marco Antonio Casanova. – 2007.

87 f. : il. ; 30 cm

Dissertação (Mestrado em Informática) – Pontifícia Universidade Católica do Rio de Janeiro, 2007.

Inclui bibliografia.

1. Informática – Teses. 2. Dados gerados por sensores 3. Compressão. 4. Wavelets. I. Casanova, Marco Antonio II. Pontifícia Universidade do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

À minha esposa Bárbara Kurpan por sempre estar ao meu lado e me apoiar nos momentos mais difíceis. Pelo trabalho de ler e revisar todo este texto.

Ao meu orientador professor Marco Antonio Casanova pela oportunidade de crescimento e todo conhecimento transmitido durante as aulas e na orientação deste trabalho.

À Petrobras, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado. Um agradecimento especial a Francisco Neves de Aquino pelo apoio.

Ao meu amigo Pedro Mário pela co-orientação informal e por me convencer que eu era “matematicamente” capaz.

À minha equipe de trabalho, Róbson Alves dos Santos, pelo tempo dedicado à discussão das idéias utilizadas neste trabalho, Dennis Lobato e Sakae Kagohara, por assumirem diversas tarefas extras para que eu pudesse me dedicar a este trabalho.

Aos meus pais Gisella Gonella e José Antônio por investirem em mim e sempre acreditarem que esse dia chegaria.

Resumo

Fernandez, Marcelo Gonella; Casanova, Marco Antonio; **Tratamento e Compressão baseada em Wavelets para Dados Adquiridos por Sensores**. Rio de Janeiro, 2007. 87p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Esta dissertação apresenta uma estratégia para desenvolver mecanismos de compressão de dados adquiridos por sensores, seguindo como inspiração o processo utilizado no formato JPG2000. A estratégia adota a abordagem das séries históricas dos dados sob o ponto de vista do processamento de sinais. Dada à natureza instável dos sensores é natural que ruídos sejam adicionados ao sinal original. Estes ruídos são detectados e tratados enquanto o sinal é suavizado e limpo, facilitando a análise, ao passo que em que componentes pouco relevantes são removidos ou aproximados, permitindo que o sinal seja comprimido com pouca perda de informação.

Palavras-chave

Compressão de dados, Processamento de Sinais, Remoção de ruídos, Redes de sensores, Transformadas Wavelet, Teoria da informação, Teoria da aproximação, Técnicas de quantização.

Abstract

Fernandez, Marcelo Gonella; Casanova, Marco Antonio; Treatment and Wavelet-based Compression of Sensor Data: Rio de Janeiro, 2006. 87p. Master Dissertation – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

This dissertation introduces a strategy to develop a compression method for sensor data inspired on the JPG2000 techniques. The strategy adopted processes data streams much in the same way as signal processing. Due to the unstable nature of sensor data, noise is added to the original signal. This noise is detected and treated while the signal is cleaned and smoothed, making it easier to analyze the data stream. Less relevant signal components are removed or approximated allowing the signal to be compressed with few information loss.

Keywords

Data Compression, Signal Processing, Denoising, Sensor data, Sensor Network, Wavelet Transform, Information Theory, Approximation Theory, Quantization Techniques.

Sumário

1	Introdução	13
1.1	Motivação	13
1.2	Abordagem proposta	15
2	Preliminares	17
2.1	Toolkit	17
2.2	Tecnologias de sensores	18
2.3	Processamento de sinais	20
2.3.1	Introdução	20
2.3.2	Ferramentas para análise de sinais	21
2.3.3	Análise Wavelet	22
2.4	Técnicas para compressão de dados	28
2.4.1	Introdução	28
2.4.2	Compressão sem perdas	29
2.4.3	Compressão com perdas	29
2.4.4	Codificação de Huffman	30
3	Arquitetura para tratamento e compressão de dados de sensores	34
3.1	Visão geral	34
3.2	Transformada Wavelet	36
3.3	Suavização por descarte de coeficientes	41
3.4	Tratamento dos coeficientes para remoção de ruídos	45
3.5	Quantização dos coeficientes	48
3.6	Codificação dos coeficientes quantizados	52
3.7	Codificação de datas e horas	55
3.8	Estrutura do arquivo binário	63
4	Apresentação dos resultados	65
4.1	Introdução	65
4.2	Análise do Conjunto 1	66

4.3	Análise do conjunto 2	79
5	Conclusões	84
5.1	Contribuições	84
5.2	Trabalhos futuros	85
6	Referências bibliográficas	86

Lista de figuras

Figura 1: Função base de Haar	25
Figura 2: Arquitetura da solução de compressão	34
Figura 3: Função base Daubechies D4	37
Figura 4: Exemplo de aproximação por Haar	44
Figura 5: Original X Usando 50% dos coeficientes com D4	44
Figura 6: Esquema de compressão de informações temporais	56
Figura 7: Exemplo de Coleta de estatísticas para compressão de tempo	57
Figura 8: Sinal original do conjunto 1	66
Figura 9: Sinal original do conjunto 1 sem detalhes de tempo	67
Figura 10: Sinal do conjunto 1 na resolução nível 3	67
Figura 11: Comparação Original X Tratado no trecho pontilhado da figura 10	68
Figura 12: Sinal do Conjunto 1 resolução nível 3 com fator denoise 0.75	69
Figura 13: Comparação Original X Tratado no trecho pontilhado da figura 12	69
Figura 14: Sinal do conjunto 1 com resolução nível 3 e fator denoise 2.0	71
Figura 15: Comparação Original X Tratado do trecho pontilhado da figura 14	71
Figura 16: Sinal do conjunto 1 com resolução nível 4	72
Figura 17: Comparação Original X Tratado do trecho pontilhado da figura 16	73
Figura 18: Sinal do conjunto 1 com resolução nível 5	74
Figura 19: Comparação Original X Tratado do trecho pontilhado da figura 18	74
Figura 20: Conjunto 1: Original X Resolução nível 5 e denoise fator 1.5	75
Figura 21: Comparação Original X Tratado do trecho pontilhado da figura 21	76
Figura 22: Conjunto 1: Original X Resolução nível 6	76
Figura 23: Comparação Original X Tratado do trecho pontilhado da figura 23	77
Figura 24: Gráfico comparativo do tamanho do arquivo por nível de resolução	78
Figura 25: Sinal original do conjunto 2	79
Figura 26: Sinal do conjunto 2 com resolução nível 3	80
Figura 27: Conjunto 2: Trecho aleatório do sinal original contaminado	80
Figura 28: Conjunto 2: Trecho Sinal Contaminado X Resolução Nível 3	80
Figura 29: Conjunto 2: Original X Resolução nível 3	81

Figura 30: Sinal do conjunto 2 com resolução nível 4 82

Figura 31: Sinal do conjunto 2 com resolução nível 4 e fator denoise 0.35 83

Lista de tabelas

Tabela 1 : Tabela de correspondência entre escalas e frequências	24
Tabela 2: Exemplo de codificação de um alfabeto	31
Tabela 3: Exemplo de frequências por intervalo de quantização	49
Tabela 4: Exemplo de atribuição de níveis de quantização por intervalo	50
Tabela 5: Exemplo de definição de valores representativos por intervalo	51
Tabela 6: Exemplo de tabela de codificação de Huffman	53
Tabela 7: Exemplo de decodificação das séries históricas comprimidas	62
Tabela 8: Conjuntos de teste	65
Tabela 9: Definição da quantidade de coeficientes por resolução	66
Tabela 10: Resultados da compactação para o caso 1	68
Tabela 11: Resultados da compactação para o caso 2	70
Tabela 12: Resultados da compactação para o caso 3	72
Tabela 13: Resultados da compactação para o caso 4	73
Tabela 14: Resultados da compactação para o caso 5	75
Tabela 15: Resultados da compactação para o caso 6	77
Tabela 16: Resultados da compactação para o caso 7	81

Lista de Siglas

DWT	Transformada Wavelet Discreta
SGBD	Sistema gerenciador de Banco de Dados
FFT	Transformada Rápida de Fourier
CLP	Controlador Lógico Programável