



Fernando Rimola da Cruz Mano

**Classificação e Segmentação de Áudio a partir de Fatores
de Escala MPEG**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para
obtenção do título de Mestre pelo Programa de Pós-
Graduação em Informática da PUC-Rio.

Orientador: Prof. Bruno Feijó

Rio de Janeiro,
setembro de 2007

Fernando Rimola da Cruz Mano

Classificação e Segmentação de Áudio a partir de Fatores de Escala MPEG

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Bruno Feijó

Orientador

Departamento de Informática - PUC-Rio

Prof. Lisandro Lovisolo

Departamento de Eletrônica e Telecomunicações - UERJ

Mônica Maria Ferreira da Costa

Midiarte

Prof. José Eugenio Leal

Coordenador(a) Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 03 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.

Fernando Rimola da Cruz Mano

Ficha Catalográfica

Mano, Fernando Rimola da Cruz

Classificação e segmentação de áudio a partir de fatores de escala MPEG / Fernando Rimola da Cruz Mano ; orientador: Bruno Feijó. – 2007.

62 f. : ; 30 cm

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

Inclui bibliografia

1. Informática – Teses. 2. MPEG. Fatores de escala. 3. Domínio comprimido. 4. Segmentação. 5. Classificação. 6. Análise do áudio. I. Feijó, Bruno. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

À minha família, por todo o apoio e incentivo.

Aos meus amigos, em especial à Cristina, eterna companheira de reclamações.

Ao Professor Bruno Feijó e aos membros da Midiarte Mônica, em especial, e Dilza, pela orientação e ajuda no desenvolvimento desse trabalho.

A todos que me ofereceram apoio e ajuda, por menor que fosse, durante todo o mestrado.

Resumo

Rimola da Cruz Mano, Fernando; Feijó, Bruno. **Classificação e Segmentação de Áudio MPEG a partir dos Fatores de Escala**. Rio de Janeiro, 2007. 62p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

As tarefas de segmentação e classificação automáticas de áudio vêm se tornando cada vez mais importantes com o crescimento da produção e armazenamento de mídia digital. Este trabalho se baseia em características do padrão MPEG, que é considerado o padrão para acervos digitais, para sugerir algoritmos de grande eficiência para realizar essas tarefas. Ao passo que há muitos estudos trabalhando a partir do vídeo, o áudio ainda é pouco utilizado de forma eficiente para auxiliar nessas tarefas. Os algoritmos sugeridos partem da leitura apenas dos fatores de escala presentes no Layer 2 do áudio MPEG para ambas as tarefas. Com isso, é necessária a leitura da menor quantidade possível de informações, o que diminui significativamente o volume de dados manipulado durante a análise e torna seu desempenho excelente em termos de tempo de processamento. O algoritmo proposto para a classificação divide o áudio em quatro possíveis tipos: silêncio, fala, música e aplausos. Já o algoritmo de segmentação encontra as mudanças significativas de áudio, que são indícios de segmentos e mudanças de cena. Foram realizados testes com diferentes tipos de vídeos, e ambos os algoritmos mostraram bons resultados.

Palavras-chave

MPEG; fatores de escala; domínio comprimido; segmentação; classificação; análise do áudio

Abstract

Rimola da Cruz Mano, Fernando; Feijó, Bruno. **Classification and Segmentation of MPEG Audio based on Scale; Factors.** Rio de Janeiro, 2007. 62p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

With the growth of production and storing of digital media, audio segmentation and classification are becoming increasingly important. This work is based on characteristics of the MPEG standard, considered to be the standard for digital media storage and retrieval, to propose efficient algorithms to perform these tasks. While there are many studies based on video analysis, the audio information is still not widely used in an efficient way. The suggested algorithms for both tasks are based only on the scale factors present on layer 2 MPEG audio. That allows them to read the smallest amount of information possible, significantly diminishing the amount of data manipulated during the analysis and making their performance excellent in terms of processing time. The algorithm proposed for audio classification divides audio in four possible types: silent, speech, music and applause. The segmentation algorithm finds significant changes on the audio signal that represent clues of audio segments and scene changes. Tests were made with a wide range of types of video, and both algorithms show good results.

Keywords

MPEG; scale factors; compressed domain; segmentation; classification; audio analysis

Sumário

| | |
|--|----|
| 1 Introdução | 11 |
| 1.1. Classificação e Segmentação de Vídeo | 11 |
| 1.2. MPEG-1 Layer 2 | 13 |
| 1.2.1. O padrão de áudio MPEG-1 Layer 2 | 13 |
| 1.2.2. Motivos da escolha do padrão | 16 |
| 1.3. Escopo da dissertação | 18 |
| 1.4. Organização da Dissertação | 19 |
| 2 Trabalhos relacionados | 20 |
| 2.1. Domínio não-comprimido | 20 |
| 2.2. Domínio comprimido | 22 |
| 2.3. Trabalhos a partir dos <i>scale factors</i> | 25 |
| 3 Classificação | 27 |
| 3.1. Resumo do algoritmo proposto | 27 |
| 3.2. Características do áudio extraídas para análise | 29 |
| 3.3. Funcionamento do algoritmo | 30 |
| 3.3.1. Etapa principal | 30 |
| 3.3.2. Pós-processamento | 32 |
| 3.4. Inovações apresentadas | 35 |
| 4 Segmentação | 38 |
| 4.1. Algoritmo proposto | 38 |
| 4.2. Inovações apresentadas | 42 |
| 5 Resultados e testes | 45 |
| 5.1. Acervo de vídeos | 45 |
| 5.2. Classificação | 46 |
| 5.3. Segmentação | 48 |
| 5.3.1. Métodos e termos | 48 |

| | |
|------------------------------|----|
| 5.3.2. Resultados | 49 |
| 6 Conclusão | 53 |
| 6.1. Resumo | 53 |
| 6.2. Trabalhos futuros | 55 |
| 7 Referências bibliográficas | 58 |

Lista de figuras

| | |
|---|----|
| Figura 1 – Sensibilidade do ouvido humano | 14 |
| Figura 2 – Mascaramento de frequências e temporal | 14 |
| Figura 3 – Formato dos quadros de áudio MPEG | 15 |
| Figura 4 – Algoritmo de classificação do áudio | 28 |

Lista de tabelas

| | |
|---|----|
| Tabela 1 – Média de acertos da classificação por tipo de vídeo | 46 |
| Tabela 2 – Média de acertos da classificação por tipo de áudio | 47 |
| Tabela 3 – Desempenho do algoritmo de segmentação a partir do tipo de vídeo | 50 |