



Gustavo Costa Gomes Moreira

**Reconhecedor de Objetos em Vídeos Digitais para
Aplicações Interativas**

Dissertação de Mestrado

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

Orientador: Prof. Bruno Feijó

Rio de Janeiro
Agosto de 2008



Gustavo Costa Gomes Moreira

Reconhecedor de Objetos em Vídeos Digitais para Aplicações Interativas

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. Bruno Feijó

Orientador

Departamento de Informática – PUC-Rio

Prof. Marcelo de Andrade Dreux

Departamento de Engenharia Mecânica – PUC-Rio

Prof. Raul Queiroz Feitosa

Departamento de Engenharia Elétrica – PUC-Rio

Prof. José Eugênio Leal

Coordenador Setorial do Centro

Técnico Científico – PUC-Rio

Rio de Janeiro, 22 de agosto de 2008

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

Gustavo Costa Gomes Moreira

Graduou-se no curso de Bacharelado em Informática pela PUC-Rio em 2005. Como pesquisador tem interesse nas áreas de Computação Gráfica e Inteligência Artificial. Sua experiência profissional inclui oito anos de consultoria e desenvolvimento de aplicações financeiras, comerciais e administrativas. Atualmente é Doutorando em Informática pela PUC-Rio e consultor da Superintendência Administrativa da mesma Universidade.

Ficha Catalográfica

Moreira, Gustavo Costa Gomes

Reconhecedor de objetos em vídeos digitais para aplicações interativas / Gustavo Costa Gomes Moreira ; orientador: Bruno Feijó. – 2009.

70 f. : il. ; 30 cm

Dissertação (Mestrado em Informática)–Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.

Inclui referências bibliográficas.

1. Informática – Teses. 2. Reconhecimento de objetos em tempo-real. 3. Aplicações de vídeo interativo. 4. TV Digital interativa. 5. Visão computacional. 6. Análise de imagens. I. Feijó, Bruno. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Para meus pais, esposa, avôs e irmão,
por todo o incentivo dado até hoje.

Agradecimentos

Ao meu orientador, Professor Bruno Feijó, que com sua mente brilhante me orientou de forma clara e precisa durante toda a realização desta pesquisa. Sou grato por todo o apoio e incentivo que tornaram esta dissertação possível.

Aos meus pais Edgar e Elza Maria que me ensinaram o valor do saber e do conhecimento, sempre me apoiando nos estudos.

A minha esposa Claudia, por todo o incentivo, paciência e companheirismo durante esta jornada.

Aos meus avôs José e Elza e ao meu irmão Edgar Jr, que me mostraram a importância da Informática.

À Professora Clarisse Sieckenius de Souza, a qual me incentivou a ingressar no Mestrado e mostrou meu futuro como pesquisador ainda durante a Graduação.

Ao Dr. Álvaro Rodrigues, grande amigo que me apóia desde o meu primeiro contato profissional na Informática.

Ao Superintendente Administrativo Floriano Saad Mazini, ao Coordenador de Sistemas Administrativos da PUC Gustavo Miranda, e a todos da equipe do SGU, que sempre me deram todo o apoio necessário para cursar o Mestrado.

A cada um dos meus amigos e familiares, agradeço a amizade e a alegria que contribuíram imensamente para que durante todo este trabalho a descontração e o bom humor estivessem sempre presentes.

Agradeço também a CAPES pelos auxílios financeiros concedidos durante o curso.

Resumo

Moreira, Gustavo Costa Gomes; Feijó, Bruno. **Reconhecedor de Objetos em Vídeos Digitais para Aplicações Interativas**. Rio de Janeiro, 2008. 70p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Detecção e reconhecimento de objetos são uma questão importante na área de Visão Computacional, onde a sua realização em tempo real e com taxas baixas de falsos positivos tem se tornado o objetivo principal de inúmeras pesquisas, inclusive daquelas relacionadas às novas formas de interatividade na TV Digital. Esta dissertação propõe um sistema de software baseado em aprendizado de máquina que permite um treinamento eficiente para novos objetos e realiza o subsequente reconhecimento destes objetos em tempo real, tanto para imagens estáticas como para vídeos digitais. O sistema é baseado no uso de características Haar do objeto, que requerem um baixo tempo de computação para o seu cálculo, e na utilização de classificadores em cascata, que permitem tanto um rápido descarte de áreas da imagem que não possuem o objeto de interesse, quanto uma baixa ocorrência de falsos positivos. Por meio do uso de técnicas de segmentação de imagem, o sistema torna a busca por objetos uma operação extremamente rápida em vídeos de alta resolução. Além disto, com a utilização de técnicas de paralelismo, pode-se detectar vários objetos simultaneamente sem perda de desempenho.

Palavras-chave

Reconhecimento de Objetos em Tempo-Real, Aplicações de Vídeo Interativo, TV Digital Interativa, Visão Computacional, Análise de Imagens.

Abstract

Moreira, Gustavo Costa Gomes; Feijó, Bruno (Advisor). **Object Recognition System in Digital Videos for Interactive Applications**. Rio de Janeiro, 2008. 70p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Object detection and recognition are an important issue in the field of Computer Vision, where its accomplishment in both real time and low false positives rates has become the main goal of various research works, including the ones related to new interactivity forms in Digital TV. This dissertation proposes a software system based on machine learning that allows an efficient training for new objects and performs their subsequent recognition in real time, for both static images and digital videos. The proposed system is based on the use of Haar features of the object, which require a low computation time for their calculation, and on the usage of a cascade of classifiers, which allows a quick discard of image areas that does not contain the desired object while having a low occurrence of false positives. Through the use of image segmentation techniques, the system turns the search for objects into an extremely fast operation in high-resolution videos. Furthermore, through the use of parallelism techniques, one can simultaneously detect various objects without losing performance.

Keywords

Real-Time Object Recognition, Interactive Video Applications, Interactive Digital TV, Computer Vision, Image Analysis.

Sumário

1	Introdução	12
1.1.	Motivação	12
1.2.	Proposta	14
1.3.	Estrutura	15
2	Trabalhos Relacionados	16
2.1.	Introdução	16
2.2.	<i>Template</i> de Objetos	16
2.3.	Sistema de Coordenadas	19
2.4.	Análise de Histogramas	22
2.5.	Aprendizado de Máquina e Classificadores	25
2.6.	Conclusão	27
3	Técnica Baseada em Treinamento e Cascata de Classificadores	29
3.1.	Introdução	29
3.2.	Formato das Características de um Objeto	29
3.3.	Funções de Classificação	32
3.4.	Cascata de Classificadores	34
3.5.	A Busca pelo Objeto	37
3.6.	Conclusão	38
4	Proposta de Aperfeiçoamento da Técnica	40
4.1.	Introdução	40
4.2.	Segmentação por Subtração de Fundo	40
4.3.	Redução das Áreas Analisadas da Imagem	43
4.4.	Paralelização	48
4.5.	Conclusão	50
5	Protótipo e Resultados Obtidos	51
5.1.	Introdução	51
5.2.	Arquitetura do Protótipo	51
5.3.	Treinamento	53

5.4. Detecção de Objetos	56
5.5. Resultados Obtidos	57
5.6. Conclusão	62
6 Conclusão e Trabalhos Futuros	64
6.1. Principais Contribuições	64
6.2. Trabalhos Futuros	65
7 Referências	67

Lista de figuras

Figura 1. Área ativa para interação com o telespectador (extraído de VisionLab(2007))	12
Figura 2. Combinação de Imagens usando Transformações de Distância	17
Figura 3. Comparação usando características de bordas (a) original (b) <i>template</i> (c) bordas (d) imagem TD (extraído de Gavrilin e Philomin (1999))	18
Figura 4. Hierarquia de <i>templates</i> de um pedestre (visão parcial) (extraído de Gavrilin e Philomin (1999))	19
Figura 5. Ilustração do princípio de detecção por sistema de coordenadas (extraído de Guanglin et al. (2007))	20
Figura 6. Detecção através do método “Pedestrian Detection Strip” (extraído de Guanglin et al. (2007))	21
Figura 7. Ilustração da comparação do pixel central com seus vizinhos	22
Figura 8. <i>CTs</i> para pixels vizinhos (bit 5 da $CT(x,y)$ e bit 4 da $CT(x,y+1)$ são complementares	23
Figura 9. Exemplo de imagem <i>CT</i> (Census Transform) (extraído de Wu e Rehg (2008))	23
Figura 10. Pirâmide espacial de uma imagem com 3 níveis	24
Figura 11. Exemplo de imagens no formato de Haar Wavelets (extraído de Papageorgiou e Poggio (2000))	26
Figura 12. Formas possíveis de uma característica retangular (extraído de Viola e Jones (2001))	30
Figura 13. Exemplo de área retangular a ser calculada na Imagem Integral (extraído de Viola e Jones (2001))	31
Figura 14. Cálculo da área de região retangular de uma Imagem Integral	31
Figura 15. Características Estendidas (extraído de Lienhart e Maydt (2002))	32
Figura 16. Características retangulares mais marcantes de uma face (extraído de Viola e Jones (2001))	33
Figura 17. Cascata de Classificadores	35
Figura 18. Exemplo de Subtração de Fundo por métodos básicos que usam a equação (4-2), para valores altos e baixos do parâmetro <i>Limiar</i> . (extraído de Piccardi (2004b))	43

Figura 19. Exemplo de extração de plano de frente utilizando $\alpha = 0.7$ e $\alpha = 0.03$ respectivamente	44
Figura 20. Definição de duas regiões retangulares sobre o plano de frente de um instante t do vídeo	45
Figura 21. Heurística para o descarte de áreas sem necessidade de processamento	47
Figura 22. Representação da arquitetura dos módulos que compõem o protótipo	52
Figura 23. Criação de conjunto finito de imagens positivas a partir de uma imagem	54
Figura 24. Marcação de áreas que contém o objeto para geração de conjunto de imagens positivas	55
Figura 25. Parâmetros para se realizar o treinamento de um objeto	55
Figura 26. Parâmetros para a detecção de objetos	56
Figura 27. Detecção em imagem com resolução 1024x768 (extraída de X Files 2 (2008))	57
Figura 28. Tempo para detecção em imagem estática de resolução 320x240	59
Figura 29. Tempo para detecção de imagem estática de resolução 640x480	59
Figura 30. Tempo para detecção de imagem estática de resolução 1024x768	59
Figura 31. Exemplo de alto dinamismo em uma cena	60
Figura 32. Exemplo de baixo dinamismo em uma cena	60
Figura 33. Tempo médio para detecção em cada <i>frame</i> de vídeo com resolução 320x240	61
Figura 34. Tempo médio para detecção em cada <i>frame</i> de vídeo com resolução 1280x768	61
Figura 35. Tempo médio para detecção de 4 objetos simultaneamente em cada <i>frame</i> do vídeo com resolução 1280x768	61