

Automatização do Ajuste de Segmentadores Neurais de Imagens Coloridas*

Fernando H. B. Cardoso[†] e Herman M. Gomes

Universidade Federal de Campina Grande, Departamento de Sistemas e Computação

Laboratório de Visão Computacional

Av. Aprígio Veloso, S/N, Bloco CN, Sala 202

{fernando, hmg}@dsc.ufcg.edu.br

Resumo

O ajuste de segmentadores de imagens que detectam materiais é executado manualmente, com pouca automatização. Uma vez que esta segmentação é geralmente um passo intermediário de uma aplicação maior da Visão Computacional, a falta de automatização faz com que esforço seja investido em uma tarefa secundária. Neste trabalho é proposta uma técnica de automatização do ajuste de Redes Neurais Artificiais que segmentam imagens ao detectar materiais baseando-se em informações de cor. Esta técnica não necessita de supervisão humana, permitindo uma redução no esforço de obtenção dos segmentadores. As redes neurais obtidas detectam 2,56% a mais da área da superfície que se deseja segmentar – em relação a outros segmentadores também testados – com taxa de falsas aceitação de 6,89%.

Abstract

Tuning material-detecting-capable image segmenters is performed manually, with little automation. Once the segmentation is usually an intermediary step to a myriad of applications within Computer Vision, the lack of automation leads to effort wasting in secondary tasks. In this work, we propose a technique to automatically tune neural networks that segment images based on color and texture information. This technique does not require human supervision, reducing the effort in obtaining segmenters. The automatically tuned neural networks detect 2.56% more of the material's surface – related to other segmenters also tested – with false acceptance rate of 6.89%.

* Resumo estendido de dissertação de mestrado.

† Este trabalho é parcialmente financiado pelo CNPq (831590/1999-8).

1. Introdução

A detecção de superfície de materiais em imagens é um passo importante para diversas aplicações em Visão Computacional. A detecção de pele humana, por exemplo, pode restringir o espaço de busca de algoritmos de localização de faces [4, 7, 11] ou fornecer informação semântica das imagens que indica a presença de pornografia [8]. Outro exemplo é a detecção de grama em quadros de vídeos esportivos, para facilitar a sua sumarização [3, 15].

A obtenção de segmentadores de imagens capazes de detectar materiais envolve (i) construir um conjunto de imagens segmentadas manualmente – ou, eventualmente, utilizar uma base pública de imagens –; (ii) analisar os melhores parâmetros – espaço de cores, quantidade de níveis de quantização, tamanho da janela de *pixels* a ser considerada, dentre outras – para detectar o material em questão; (iii) ajuste de um classificador que decide se cada *pixel* faz parte ou não da superfície do material e (iv) testar o segmentador para verificar a sua qualidade.

Estas tarefas são feitas com pouca ou nenhuma automatização, uma vez que, após extensiva revisão da literatura, não foi encontrado um método automático para obtenção e ajuste de segmentadores deste tipo. Assim, é necessário investir tempo e esforço para uma atividade secundária durante um projeto em Visão Computacional, que poderiam ser investidos em outras atividades que contribuem mais diretamente com o seu objetivo geral.

Desta maneira, a automatização dos processos de obtenção e ajuste de segmentadores introduz uma maior agilidade para alcançar resultados nos projetos de pesquisa em Visão Computacional que necessitam da detecção de materiais como passo intermediário, não sendo mais necessário desviar esforços para resolver um problema periférico.

A principal contribuição deste trabalho é a proposição de um método automatizado de obtenção e ajuste de segmentadores neurais capazes de detectar materiais (Seção 2), cujo protótipo preliminar foi testado para detectar dois tipos de material: grama e pele (Seção 3).

São apresentados também uma visão comparativa da técnica proposta com trabalhos relacionados (Seção 4) e as conclusões finais, juntamente com sugestões de trabalhos futuros (Seção 5).

2. Método Proposto

O método proposto neste trabalho tem por objetivo automatizar a obtenção e o ajuste de segmentadores neurais que são capazes de detectar materiais em imagens digitais: cada *pixel* da imagem é submetido a um classificador neural baseado em redes *perceptron* de múltiplas camadas – MLP – [5] para que seja determinada a sua pertinência ou não à superfície do material em questão.

2.1. Avaliação dos Classificadores

As métricas utilizadas para verificar a qualidade de um segmentador ζ capaz de detectar materiais são a Taxa de Detecção – $D_r(\zeta)$ – e a Taxa de Falsas Aceitações – $P_r^f(\zeta)$. A Taxa de Detecção é a razão entre a área da superfície que foi corretamente detectada – $P^t(\zeta)$ – e a área total da superfície do material em questão – S^P –, conforme a Equação 1. Já a Taxa de Falsas Aceitações é a razão entre a área da superfície que foi erroneamente considerada como pertencente ao material em questão – $P^f(\zeta)$ – e a área total da superfície que não faz parte do mesmo – S^N –, conforme a Equação 2.

$$D_r(\zeta) = \frac{P^t(\zeta)}{S^P} \quad (1)$$

$$P_r^f(\zeta) = \frac{P^f(\zeta)}{S^N} \quad (2)$$

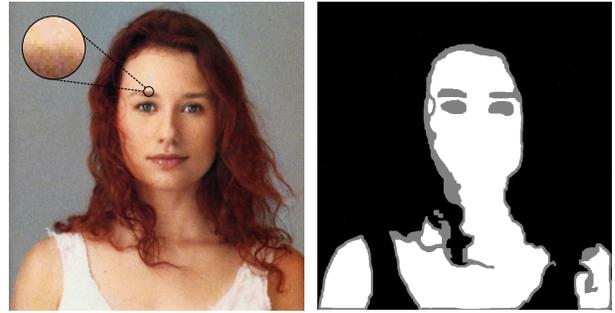
Para unificar estes dois valores em apenas um, utiliza-se como critério de decisão entre dois segmentadores a diferença entre as taxas de detecção e falsas aceitações $\delta(\zeta)$, conforme a Equação 3.

$$\delta(\zeta) = D_r(\zeta) - P_r^f(\zeta) \quad (3)$$

Com base nesta Equação, pode-se considerar o segmentador ζ_a melhor que um concorrente ζ_b quando $\delta(\zeta_a) > \delta(\zeta_b)$.

2.2. Composição dos Conjuntos de Imagens de Treinamento

O único insumo para obtenção de um segmentador utilizando a técnica deste trabalho é um conjunto de imagens acompanhadas do resultado de sua segmentação manual, da qual será extraído o conjunto de treinamento para a rede neural artificial. O rotulamento manual deve estar em conformidade com as seguintes diretrizes (Figura 1):



(a) Imagem original.

(b) Imagem segmentada.

Figura 1. Exemplo de imagem e sua respectiva segmentação manual para composição da base de treinamento – material: pele. A área em destaque exibe um exemplo de região de difícil segmentação.

- Os *pixels* correspondentes ao material que se quer detectar – superfície P – devem ser marcados de branco;
- Os que não fizerem parte de nenhuma superfície do material em questão – superfície N – são marcados de preto;
- Há ainda os *pixels* que são de difícil segmentação por estarem em regiões de transição entre superfícies ou que não se pode determinar com precisão em que superfície se encontram. Estes *pixels* são marcados de cinza (50% de brilho).

A extração do conjunto de treinamento é feita respeitando a igualdade entre as amostras de *pixels* pertencentes à superfície do material que se deseja detectar: em cada imagem da base segmentada manualmente são extraídos números iguais de amostras das duas classes. Isto garante também a igual distribuição no conjunto de treinamento como um todo e desta forma é possível treinar a rede neural sem que haja efeitos colaterais devido à maior presença de uma das classes.

2.3. Parâmetros dos Segmentadores

A técnica de automatização da obtenção e do ajuste dos segmentadores se baseia na determinação dos valores que seus parâmetros devem assumir para maximizar a taxa de detecção e minimizar a taxa de falsas aceitações. Os parâmetros dos segmentadores no estado atual do trabalho são: Espaço de Cor, Quantidade de Níveis de Quantização, Tamanho da Janela de *Pixels* e Quantidade de Neurônios na Camada Oculta.

2.3.1. Espaços de Cor O espaço de cor “padrão” no qual as imagens digitais são codificadas é o RGB. No entanto,

nem sempre este é o espaço que provê maior separabilidade entre as superfícies P e N . Desta maneira, são explorados também os espaços de cor HSV, YCbCr, XYZ, HLS, CIE $L^*a^*b^*$ e $L^*u^*v^*$.

2.3.2. Níveis de Quantização A quantidade de níveis de quantização depende do número de *bits* utilizados para representar as intensidades de cada canal. Apesar de, normalmente, as imagens digitais utilizarem 256 níveis de quantização em cada canal, a separabilidade entre P e N pode ser aumentada caso a quantidade de níveis de quantização seja reduzida. Para tanto, a quantidade de *bits* em cada canal é reduzida e são explorados também 128, 64, 32, 16, 8, 4 e 2 níveis de quantização em cada canal.

2.3.3. Tamanho da Janela de Pixels Considerar uma vizinhança em redor do *pixel* que está sendo classificado dá a rede maior informação acerca da superfície em que este se encontra, uma vez que considera a sua localização em uma região maior em vez de considerar sua cor isoladamente. O tamanho desta vizinhança é um dos parâmetros a serem determinados pela técnica proposta, que além de janelas de tamanho 1 – o *pixel* isolado – considera tamanhos ímpares de janela variando de 3 a 9.

2.3.4. Neurônios na Camada Oculta A qualidade dos resultados de uma rede neural artificial do tipo MLP depende da quantidade de neurônios em sua camada oculta [5]. Desta forma, é necessário explorar a quantidade ideal para cada tipo de material que se deseja detectar. No estágio atual do trabalho os valores explorados para a quantidade de neurônios na camada escondida vai de 3 a 12.

2.4. Obtenção e Ajuste dos Segmentadores

A obtenção e o ajuste do segmentador neural capaz de detectar o material desejado ocorre variando-se automaticamente parâmetros especificados na Seção 2.3 um a um para que se encontre a combinação mais adequada dos mesmos (Figura 2). Primeiramente fixa-se a quantidade de níveis de quantização em 256, o tamanho da janela de *pixels* em 1 e a quantidade de neurônios na camada escondida em 6 para que se possa testar todos os espaços de cor a fim de se determinar o espaço de cor C_{best} cujo segmentador apresenta menor δ .

Em seguida, para que seja obtida a quantidade de níveis de quantização adequada Q_{best} , testam-se todos os valores que este parâmetro pode assumir, utilizando C_{best} como espaço de cor e fixando-se o tamanho da janela em 1 e a quantidade de neurônios na camada escondida ainda em 6 e adota-se a que apresentar δ mais alto.

O terceiro parâmetro a ser ajustado é o tamanho da janela de *pixels*: para obter o tamanho W_{best} ideal são testados segmentadores obtidos com cada um dos tamanhos possíveis, no espaço de cores C_{best} com Q_{best} níveis de quanti-

zação em cada canal e é escolhido o que apresentar melhor δ .

Por fim, a quantidade de neurônios na camada escondida é ajustada da mesma maneira: são obtidos segmentadores para cada uma das quantidades de neurônios na camada escondida e a quantidade H_{best} ideal é a que apresentar maior δ .

Vale ressaltar que, para cada configuração de rede neural MLP testada (combinação de espaço de cor, quantidade de níveis de quantização, tamanho de janela e quantidade de neurônios na camada oculta) são treinadas 10 redes diferentes, a fim de se reduzir os efeitos da aleatoriedade inerente à inicialização das MLPs [5]. Dentre estas 10, é adotada a que apresentar maior D_r .

3. Resultados Obtidos

Com o objetivo de verificar a aplicabilidade da técnica proposta, ela foi utilizada para obter automaticamente segmentadores para dois tipos de materiais: pele e grama. Os resultados fornecidos por estes segmentadores foram então confrontados com os resultados de outros trabalhos, e de classificadores genéricos, a fim de se estabelecer uma análise comparativa da técnica abordada.

3.1. Detecção de Grama em Quadros de Vídeos Desportivos

A detecção de grama em quadros de vídeos desportivos é útil para prover informações que subsidiam a sua posterior sumarização (geração automática de resumos das partidas, contendo apenas os melhores momentos), através da distinção entre tomadas de longa distância - lances que provavelmente serão interessantes para o espectador - e de curta distância (*close*s e tomadas da torcida, por exemplo) - que provavelmente não são interessantes para o entendimento da partida -, da determinação da posição da bola com relação ao campo de jogo, entre outros [10].

O conjunto de imagens que compõe o conjunto de treinamento para obtenção e ajuste do segmentador foi composta de 25 imagens extraídas da base de dados utilizada por Firmino Jr. e Gomes [3] para parametrizar seu detector de grama baseado em casamento de padrões. Devido ao fato de a base ter sido preparada por terceiros, não há regiões de indecisão (áreas marcadas de cinza) neste conjunto de treinamento. A Figura 3 exibe um exemplo de imagem contida neste conjunto.

Os parâmetros ideais encontrados para o segmentador de grama foram:

- Espaço de Cor: RGB;
- Níveis de Quantização: 128;
- Tamanho da Janela: 1 *pixel* e

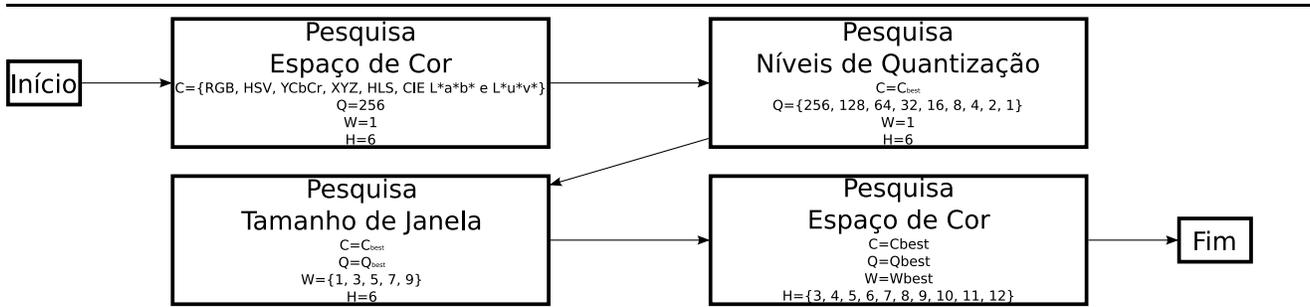
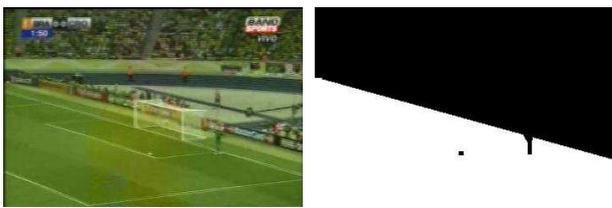


Figura 2. Esquema do ajuste dos segmentadores utilizando a técnica proposta.



(a) Imagem original.

(b) Imagem segmentada.

Figura 3. Exemplo de imagem e sua respectiva segmentação manual para composição da base de treinamento para o segmentador de grama.



(a) Imagem Original.

(b) Imagem Segmentada.

Figura 4. Exemplo de imagem segmentada com a técnica proposta.

Segmentador (ζ)	$D_r(\zeta)$	$P_r^f(\zeta)$	$\delta(\zeta)$
MLP Auto-treinada	88,42%	5,93%	82,50%
<i>Naïve Bayes</i>	86,25%	6,42%	79,83%
Firmino Jr. e Gomes	98,12%	20,79%	77,33%

Tabela 1. Resultados Obtidos para Segmentação de Grama

- Neurônios na Camada Oculta: 3.

O segmentador obtido com estes parâmetros foi testado, juntamente com um classificador *bayesiano* – gerado através da plataforma de mineração de dados Weka [16] – e com o segmentador de grama proposto por Firmino Jr. e Gomes [3], em um conjunto de imagens de teste composto pelas demais imagens da base de dados da qual foi extraído o conjunto de treinamento. A Tabela 3.1 exibe os resultados obtidos e a Figura 4 exibe dois exemplos de imagens segmentadas utilizando a rede neural MLP automaticamente obtida.

A Tabela de mostra que os resultados obtidos pela rede neural MLP treinada automaticamente utilizando o método proposto neste trabalho foram os melhores para detecção de grama em quadros de vídeos desportivos, uma vez que ob-

teve o maior δ (ver Equação 3).

3.2. Detecção de Pele em Fotografias

A detecção de pele em fotografias é utilizada para reduzir o espaço de busca em outras aplicações da Visão Computacional, como a detecção de faces humanas, ou a detecção de nudez, uma vez que em ambas aplicações só faz sentido executar buscas em regiões compostas de pele humana.

Para que se pudesse utilizar a técnica proposta e obter um segmentador de pele, foi utilizado um subconjunto da base de imagens fornecida por Jones e Regh [8]. Este subconjunto foi re-segmentado manualmente devido a alguns erros encontrados na segmentação semi-manual executada pelos autores. O subconjunto utilizado contava com 120 imagens. A Figura 1 (página 2) exibe um exemplo de imagem contida neste conjunto.

Os parâmetros ideais encontrados para o segmentador de pele foram:

- Espaço de Cor: YCbCr;
- Níveis de Quantização: 128;
- Tamanho da Janela: 9 *pixel* e
- Neurônios na Camada Oculta: 8.

O segmentador obtido com estes parâmetros foi testado, juntamente com um classificador *bayesiano* – também gerado através da plataforma de mineração de dados Weka –,

Segmentador (ζ)	$D_r(\zeta)$	$P_r^f(\zeta)$	$\delta(\zeta)$
HSV Cluster	88,03%	12,01%	76,02%
MLP Auto-treinada	81,40%	6,89%	74,51%
RGB-H-CbCr	61,17%	7,52%	53,63%
Naïve Bayes	79,60%	31,90%	47,70%
Cardoso e Gomes	72,77%	34,87%	37,90%

Tabela 2. Resultados Obtidos para Segmentação de Pele



(a) Imagem Original.

(b) Imagem Segmentada.

Figura 5. Exemplo de imagem segmentada com a técnica proposta.

com um segmentador de pele probabilístico de nossa autoria [2], outro que utiliza o modelo de pele RGB-H-CbCr proposta por Nusirwan *et al.* [12] e por fim, com um segmentador baseado na técnica de *cluster* no espaço de cores HSV, proposto por Herodotou *et al.* [6].

O conjunto de imagens de teste composto por 118 imagens também retiradas da base. A Tabela 3.2 exhibe os resultados obtidos e a Figura 5 exhibe dois exemplos de imagens segmentadas utilizando a rede neural MLP automaticamente obtida.

A Tabela de mostra que os resultados obtidos pela rede neural MLP treinada automaticamente utilizando o método proposto neste trabalho não foram os melhores para detecção de pele humana. No entanto, se forem consideradas a automatização da técnica proposta (não é uma técnica específica para detecção de pele) e a proximidade dos valores de δ da MLP Auto-treinada e do segmentador por *cluster* no espaço HSV, os resultados da técnica proposta são muito satisfatórios.

4. Trabalhos Relacionados

Para melhor organizar esta Seção, ela foi dividida em duas subseções: uma dedicada à detecção de materiais e outra à utilização das redes neurais artificiais na segmentação de imagens.

4.1. Detecção de Materiais

Após exaustiva revisão da bibliografia, foi encontrado apenas um grupo de pesquisa na área de detecção de materiais em imagens: o projeto PERMM [9], apesar de seu objetivo principal ser a disponibilização de imagens para recuperação baseada em conteúdo (RIBC).

Para alcançar este objetivo, as imagens são primeiramente segmentadas – utilizando uma técnica de segmentação baseada em diagramas de Voronoi [14] – e posteriormente fornecidas a extratores de características que analisam informação de cor, textura, forma e conteúdo. Por fim, as características extraídas são utilizadas como entrada de redes neurais artificiais MLP que classificam semanticamente o conteúdo da imagem, ou seja, determinam que tipo de material está presente nas mesmas.

De acordo com a descrição do projeto, as suas redes neurais são capazes de detectar os seguintes materiais: pele, céu, grama, árvores, água, areia, tijolos, tecidos, paredes interiores, nuvens, neve, asfalto e madeira. Cada classificador utiliza diferentes combinações das características extraídas para classificar as imagens: os detectores de paredes interiores e tecidos, por exemplo, não se utilizam de informação de matiz quando analisando a informação de cor. Todos os classificadores têm acurácia entre 85% e 95%, mas não foram localizados valores exatos.

A principal diferença entre a técnica proposta neste trabalho e o projeto PERMM é que este utiliza vários segmentadores dedicados e previamente ajustados manualmente para cada classe de material. Na técnica deste trabalho, cada classificador é obtido automaticamente após a aplicação de um método sistemático.

4.2. Segmentação com Redes Neurais Artificiais

Redes Neurais Artificiais do tipo MLP são modelos matemáticos do cérebro humano, dedicadas à resolução de problemas de classificação [5]. Elas podem, portanto, ser aplicadas a problemas de detecção de materiais: cada *pixel* da imagem é classificado como sendo pertencente ou não à superfície de um dado material. Elas têm também a vantagem de serem classificadores que necessitam de pouca memória, o que as fazem boas candidatas para execução em dispositivos de recursos restritos, como PDAs, Máquinas Fotográficas, dentre outros [13].

Especialmente no domínio da detecção de pele, as redes neurais MLP já foram utilizadas com relativo sucesso em alguns trabalhos. Phung *et al.* [13] demonstraram, em seu *survey* de técnicas para detecção de pele, que a rede MLP foi o método mais preciso entre os demais – foram investigados também a Mistura Gaussiana (2D e 3D) e classificadores Bayesianos, dentre outros. Neste trabalho, a rede MLP foi treinada a partir de informação de cor extraída de

30.000 *pixels* de cada classe (pele e não-pele), no espaço de cores RGB. A acurácia máxima obtida foi de 89,49% (em uma base de teste diferente da usada neste trabalho).

Akbari e Nakajima [1] também adotaram redes neurais para detectar pele em fotografias a partir de informação de cor. Em seu estudo, eles propuseram uma técnica de homogeneização para ser utilizada após a classificação das redes neurais. O modelo neural adotado foi o *Learning Vector Quantization* (LVQ), e também considerava como entrada os valores de cada *pixel* no espaço de cores RGB. A camada de saída era composta de dois neurônios – um para determinar os *pixels* de pele e outro para os de não-pele. Seu classificador foi capaz de detectar corretamente 92,75% da área de pele (em uma base de testes diferente da utilizada neste trabalho).

5. Conclusão e Trabalhos Futuros

Este trabalho propõe uma técnica automática de obtenção e ajuste de segmentadores neurais de imagens, baseados em informação de cor. O objetivo desta segmentação é detectar determinados materiais – grama, pele, dentre outros – em imagens.

De acordo com os resultados obtidos, e comparando-os com outras técnicas mais tradicionais, a segmentação produzida com a utilização das redes neurais automaticamente treinadas tem alta qualidade. Desta forma a técnica apresentada contempla o seu objetivo de fornecer ao usuário segmentadores de alta qualidade, sem a necessidade do emprego de esforço na supervisão e ajuste dos mesmos.

Futuramente, técnicas de otimização global serão incorporadas no protótipo apresentado neste trabalho, com o objetivo de facilitar a descoberta da arquitetura ideal da rede neural artificial a ser utilizada para cada um dos materiais para os quais se deseja obter segmentadores. Assim, a exploração da quantidade de neurônios da camada escondida será restrita a menos valores, além de serem ampliadas as possibilidades no que diz respeito aos outros aspectos arquiteturais das redes neurais: número de camadas escondidas, conectividade entre neurônios, dentre outros.

Além disto, extratores de características de textura serão adicionados ao protótipo, fornecendo mais pistas sobre a pertinência de um *pixel* a determinado material. Com isto, surgirão oportunidades de automatização tanto para descoberta dos parâmetros do extrator de textura, quanto para descobrir a melhor combinação entre os classificadores de cor e textura para compor o resultado final da segmentação.

Referências

- [1] M. A. Akbari and M. Nakajima. A novel color region homogenization and its application in improving skin detection accuracy. In *GRAPHITE '05: Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 269–272, New York, NY, USA, 2005. ACM.
- [2] F. H. Cardoso and H. M. Gomes. A probabilistic approach to skin detection. In *Anais do Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens 2007*, 2007.
- [3] Firmino and H. M. Gomes. Detecção de lances no campo de ataque em uma partida de futebol. In *Anais do Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens 2007*, 2007.
- [4] C. Garcia and G. Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis. *Multimedia, IEEE Transactions on*, 1(3):264–277, 1999.
- [5] S. Haykin. *Redes Neurais, Princípios e Práticas*. Editora Bookman, 2a edition, 2001.
- [6] N. Herodotou, K. N. Plataniotis, and A. N. Venetsanopoulos. *Image Processing Techniques for Multimedia Processing*. 2000.
- [7] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):696–706, 2002.
- [8] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002.
- [9] A. Laboratories. Permm: Image analysis. Project Home Page, 2007.
- [10] E. O. Lima. Sumarização de vídeos de futebol com baixo custo computacional utilizando ritmo visual: descritores de cor e movimento de câmera. Master's thesis, UNIFOR, 2006.
- [11] R. Montufar-Chaveznaza, F. H. Gallardo, and S. P. Hernandez. Face detection by polling. In *Intelligent Signal Processing, 2005 IEEE International Workshop on*, pages 292–297, 2005.
- [12] Nusirwan, K. C. Wei, and J. See. Rgb-h-cbcr skin colour model for human face detection. In *Proceedings of The MMU International Symposium on Information & Communications Technologies (M2USIC 2006)*, 2006.
- [13] S. L. Phung, A. Bouzerdoum, and D. Chai. Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(1):148–154, January 2005.
- [14] D. Sinclair. Voronoi seeded colour image segmentation. Technical report, AT&T Laboratories, 1999.
- [15] J. Wang, C. Xu, E. Chng, X. Yu, and Q. Tian. Event detection based on non-broadcast sports video. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 3, pages 1637–1640 Vol. 3, 2004.
- [16] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann, June 2005.