Um estudo da detecção automática de campos de futebol de imagens aéreas e orbitais utilizando SVM e descritores HOG

Juliano E. C. Cruz[1], Lamartine N. F. Guimarães[2], Elcio H. Shiguemori[2]

[1]Programa de Mestrado em Computação Aplicada – CAP
Instituto Nacional de Pesquisas Espaciais – INPE

[2]Instituto de Estudos Avançados – IEAv

{juliano.cruz}@lac.inpe.br

{guimarae,elcio}@ieav.cta.br

Abstract. *Automatic object recognition in digital images is not an simple task due to diverse variations present within this process, consequently, different general purpose techniques have been proposed. In this paper, an approach combining HOG and SVM for automatic soccer field detection in airborne and satellite imagery is analyzed.*

Resumo. *A detecção automática de objetos presentes em imagens não é uma tarefa fácil devido às diversas variações presentes neste processo, consequentemente, diferentes técnicas têm sido propostas para fins gerais. Neste trabalho analisa-se a abordagem que combina HOG e SVM para detecção automática de campos de futebol presentes em imagens aéreas e orbitais.*

Palavras-chave: *VANT, processamento de imagens, reconhecimento de padrões, visão computacional*

1. Introdução

Imagens aéreas ou orbitais com uma alta resolução espacial, permitem que a maioria dos objetos possam ser identificados por especialistas. Alguns objetos são de fácil reconhecimento visual, outros somente com experiência adquirida ao longo do tempo. Em certas aplicações é possível dividir todos os objetos em solo em dois grandes grupos: os fixos e os móveis. Objetos em solo, quando fixos, podem ser utilizados como marcos referenciais[Rodrigues et al. 2009], como por exemplo, pista de pouso de aeroportos, campo de futebol, entre outros. Já os objetos móveis[Rebouças and Shiguemori 2012] são por exemplo pessoas, carros, embarcações, entre outros, e podem ser utilizados em aplicações de vigilância, fiscalização, resgate ou mesmo militar. Umas das possíveis utilizações dessa abordagem seria em aplicações que utilizam plataformas VANT(veículo aéreo não tripulado), onde há atualmente uma crescente demanda de seu uso por forças policiais, armadas e em aplicações civis.

A fim de realizar o reconhecimento de um determinado objeto de forma automática, utiliza-se, nesse trabalho, descritores HOG[Dalal and Triggs 2005] combinados com o classificador SVM[Boser et al. 1992]. Essa combinação de métodos tem seu uso mais comum em outras áreas, como será visto na Seção 2. A vantagem de uma abordagem de reconhecimento automática se deve ao fato da identificação através de um operador humano ser uma tarefa cansativa e altamente suscetível à erro, pois além da tarefa de

identificação ser entediante, há geralmente uma grande quantidade de informação a ser analisada, e ainda, em certos casos, há a necessidade de que o responsável pela tarefa tenha sido especialmente capacitado para o devido fim. As principais dificuldades encontradas na detecção automática de objetos na abordagem utilizada são que as imagens foram obtidas por diferentes tipos de sensores, os objetos podem estar em diferentes poses e podem também ter sofrido transformações geométricas, entre outros.

2. Trabalhos relacionados

A combinação entre descritores HOG e classificador SVM é conhecido na literatura principalmente na identificação do corpo humano em imagens. A primeira proposta de utilização de HOG com SVM foi em [Dalal and Triggs 2005], em imagens em solo obtidas por diferentes sensores e contendo pessoas diferentes. Afirma-se conseguir uma detecção com uma performance extremamente alta.

Em [Breckon et al. 2010], utiliza-se HOG e SVM como parte de um sistema de reconhecimento humano em imagens aéreas de baixa qualidade. O papel do HOG e SVM foi de identificar pessoas nas imagens em uma etapa preliminar. A escolha dessa abordagem para essa etapa ocorreu devido a sua performance já conhecida nesse tipo de aplicação. As outras etapas do sistema se encarregavam, então, de verificar se aquele indivíduo identificado já estava, ou não, presente em imagens anteriores, para que assim, a rotulação pudesse ser realizada de forma individualizada.

Também utilizando HOG e SVM, [Felzenszwalb et al. 2010] propõe um sistema de detecção de objetos que ao invés de procurar pelo objeto inteiro, procura-se por partes desse objetos. Quando se encontrar todas essas partes juntas e em uma determinada configuração, o objeto que se procura é detectado. Afirma-se ter uma alta performance da detecção de objetos com essa abordagem.

3. Métodos

3.1. *Histogram of Oriented Gradients*

Histogram of Oriented Gradients, ou simplesmente HOG, foi primeiramente descrito em 2005 por Navneet Dalal e Bill Triggs[Dalal and Triggs 2005] e é uma técnica que utiliza as orientações dos gradientes de uma determinada imagem para obter seus descritores.

A ideia principal deste descritor é que a aparência e forma de objetos em uma imagem podem ser descritos através da distribuição dos gradientes de intensidade dos pixels ou pelas direções das bordas[Gritti et al. 2008]. O processo para gerar o descritor pode ser dividido em quatro etapas: cálculo do gradiente em cada pixel, agrupamento dos pixels em células, agrupamento das células em blocos e obtenção do descritor.

Na primeira etapa, calcula-se o gradiente de cada pixel da imagem, como pode-se ver na Figura 2(a). No método canônico usa-se uma máscara unidimensional de derivada discreta pontual tanto no eixo vertical como horizontal, como mostrado abaixo na Equação 1. Outros tipos de filtros que também fazem o cálculo de gradiente já foram testados nessa etapa[Dalal and Triggs 2005][Gritti et al. 2008].

$$[-1, 0, 1] \text{ e } [-1, 0, 1]^T \tag{1}$$

Filtro horizontal e vertical para o cálculo de gradiente.

O passo seguinte é responsável por agrupar os pixels de uma determinada região, criando-se o que se chama de célula, como pode-se ver na Figura 2(b) e 3(a). Todas as células criadas na imagem possuem mesmo formato e tamanho. Cria-se um histograma com orientação do vetor gradiente dos pixels que compõe essa célula, onde são computados os valores de magnitude de acordo com o ângulo do vetor. O histograma possui uma quantidade finita de divisões. No modelo canônico utiliza-se o histograma com nove divisões.

Após a segunda etapa, os blocos são criados através do agrupamento de células de uma certa região, como pode-se ver na Figura 3(b). Assim como as células, os blocos também sempre possuem o mesmo formato e tamanho em toda a imagem. Existem áreas dos blocos em que há uma sobreposição proposital com o bloco vizinho, o que torna o método mais eficiente em relação a uma abordagem sem essas sobreposições[Dalal and Triggs 2005].

Na etapa final, cria-se o descritor. O descritor nada mais é do que uma lista dos histogramas de todas as células de todos os blocos. A atenuação do problema das variações locais de iluminação ou de contraste entre o primeiro plano e o plano de fundo, se dá através da normalização de cada histograma de acordo com seus próprios valores[Dalal and Triggs 2005]. No método canônico, o método de normalização do vetor utilizado é o *L2-hys*. *L2-hys* consiste em em aplicar *L2-norm*, descrito na Equação 2, e limitar os resultados em um teto padrão, em seguida calcula-se *L2-norm* novamente.

$$\text{L2-norm: } \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \tag{2}$$

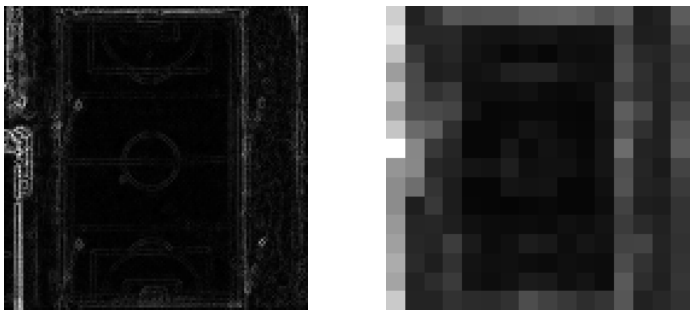
onde v é o vetor descritor, $\|v\|_k$ a sua *k*-norma para $k = 1, 2$ e e uma constante muito pequena[Dalal and Triggs 2005].

Mudanças não-lineares de iluminação podem ocorrer devido à saturação causada pela câmera ou devido a mudanças de iluminação em superfícies tridimensionais vindo de diferentes ângulos e com diferentes intensidades. Esses tipos de mudanças podem afetar diretamente na magnitude relativa de alguns gradientes. Assim, reduz-se a influência de grandes magnitudes de gradientes estipulando um valor teto para a magnitude. O valor de teto de 0,2 foi encontrado depois da execução de testes com imagens com diferente iluminação para os mesmos objetos tridimensionais[Lowe 2004].

3.1.1. *Support Vector Machine*

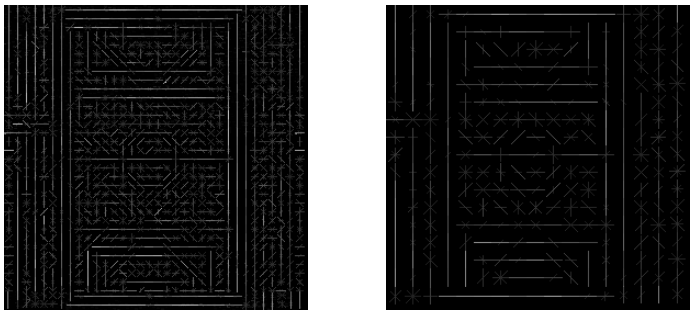
SVM, acrônimo para *Support Vector Machine*, foi descrito em 1992 por Vladmir Vapnik[Boser et al. 1992] e é um método de aprendizado supervisionado que analisa dados e reconhece padrões usado para classificação e análise de regressão. O SVM é um classificador linear binário, mas existem abordagens que o tornam capaz de classificar um conjunto de dados com classes não-linearmente separáveis ou mesmo com mais de uma classe[Theodoridis and Koutroumbas 2006].

Figura 1. Imagem de entrada[Google]



(a) (b)

Figura 2. Magnitude do vetor gradiente dos pixels (a) e das células (b)



(a) (b)

Figura 3. Histograma de orientação das células (a) e dos blocos (b)

Seja $\mathbf{X}$ um conjunto de treinamento, onde $\mathbf{x}_i$, $i = 1, 2, ..., N$, são vetores de atributos. Estes vetores pertencem a somente a duas classes ω_1 ou ω_2 e assumi-se que são linearmente separáveis. O objetivo é então encontrar um hiperplano(Equação 3) que classifique corretamente os vetores de treinamento.

$$g(\mathbf{x}) = \mathbf{w^T x} + w_0 \tag{3}$$

onde $\mathbf{w}$ é uma matriz unidimensional contendo os vetores de suporte, w_0 é o *bias* e $\mathbf{x}$ o vetor de atributos.

Portanto, tal hiperplano não é único, como pode-se ver na Figura 4. Mas há um conceito que deve-se sempre levar em consideração: o poder de generalização do classificador, ou seja, a capacidade do classificador de operar satisfatoriamente com dados

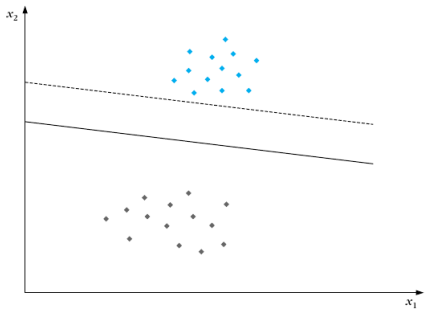


Figura 4. Exemplo de duas classes linearmente separáveis com dois classificadores lineares possíveis[Theodoridis and Koutroumbas 2006].

de fora do conjunto de treinamento sendo somente projetado com os dados de treinamento. O que o SVM faz a respeito dessa questão, é durante o treinamento escolher o hiperplano que possui maior margem entre as classes, tendo como exemplo a Figura 5 [Theodoridis and Koutroumbas 2006].

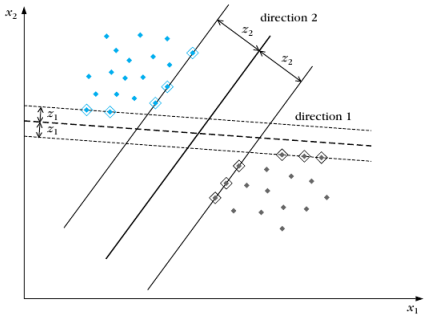


Figura 5. Exemplo de duas classes linearmente separáveis com dois classificadores lineares possíveis e suas margens em relação às classes[Theodoridis and Koutroumbas 2006].

Para se lidar com classes que não são linearmente separáveis, utiliza-se as funções *kernels* que modificam o espaço de atributos transformando o problema em linearmente separável. Além do linear, pode-se encontrar na literatura *kernels* do tipo polinomial, RBF(*Radial Basis Function*) e sigmoidal[Hsu et al. 2010].

4. Metodologia

SVM é um método de classificação supervisionada. Portanto, o processo completo é composto de duas fases: o treinamento e a classificação propriamente dita. O treinamento é uma etapa que é necessário ser repetida até que se consiga encontrar um hiperplano que separe a classe positiva da classe negativa de maneira satisfatória para o problema. Na implementação desse trabalho foram utilizados a biblioteca OpenCV [OpenCV] e o SVMlight [SVMlight], onde a primeira foi utilizada tanto na fase de treinamento, para a extração dos descritores das imagens dos conjuntos de treinamento, quanto na fase de classificação, utilizada para a extração dos descritores e para a classificação. Já o SVMlight foi utilizado somente na fase de treinamento para se obter os vetores de suporte do hiperplano.

Na extração do descritor das imagens, utilizou-se um descritor com uma janela de 128 por 128 pixels, com blocos de 16 por 16 pixels e células de 8 por 8 pixels, ou seja, cada bloco possui 4 células. Os blocos são sobrepostos em 8 pixels. Devido a essa sobreposição, em uma janela obtêm-se 225 blocos. O histograma de orientação utilizado possuía 9 divisões. O *kernel* utilizado pelo SVM é do tipo linear, pois a *classe* do descritor HOG já possui implementado um classificador SVM que suporta somente *kernel* do tipo linear. O parâmetro de penalidade *C* utilizado no treinamento foi de 0,01 como proposta em [Dalal and Triggs 2005]. Na classificação utilizou um passo variação de escala na imagem de análise de 5%, onde o processo inteiro possui 64 níveis de variação.

O conjunto de treinamento utilizado é composto de 1064 amostras positivas e 6325 negativas. As amostras positivas são compostas somente por imagens quadradas contendo campos de futebol rotacionados de 9 em 9 graus, como pode-se ver na Figura 6. A escolha em identificar campos de futebol foi feita, pois a montagem de um *dataset* desse alvo é de maior facilidade em relação à outros alvos, devido a popularidade do esporte ao redor do mundo e de ser de fácil visualização a olho nu em imagens aéreas ou orbitais. A maioria das imagens foram obtidas através do aplicativo Google Maps [Google], pois ele é de fácil uso, acesso e disponibiliza imagens orbitais em alta resolução de grande parte do mundo, no entanto as imagens possuem marcas d'água com o logotipo do Google e ano de captura. Mas também foram utilizadas imagens aéreas e Ikonos. Antes da etapa de extração do descritor na fase de treinamento todas as imagens foram primeiramente reduzidas para o tamanho de 128x128. O conjunto de imagens negativas é compostas por imagens aéreas, Ikonos, do Google Maps e de câmeras fotográficas portáteis. Na etapa de extração do descritor na fase de treinamento, as imagens negativas foram mantidas com seus tamanhos originais, onde uma janela de extração de 128x128 percorria a imagem.

Figura 6. Exemplo de amostras do conjunto de treinamento positivo[Google]

5. Resultados

Utilizou-se duas métricas[Fawcett 2006] para medir o desempenho da classificação de um conjunto de imagens obtidas por diversos sensores. As duas métricas são: a precisão, Equação 4, e a sensibilidade, Equação 5.

$$P = \frac{VP}{VP + FP} \times 100 \tag{4}$$

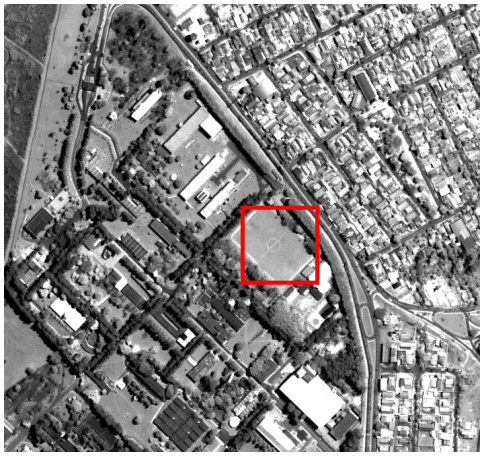


Figura 7. Resultado do processo de reconhecimento automático de campo de futebol em uma imagem Ikonos contendo a instalação do INPE de São José dos Campos

$$S = \frac{VP}{VP + FN} \times 100 \tag{5}$$

onde verdadeiros positivos(VP) são campos de futebol reconhecidos corretamente, falsos positivos(FP) são regiões da imagem em que foram classificadas erroneamente como campo de futebol e falsos negativos(FN) são campos de futebol não reconhecidos. Foram processadas 10 imagens. O resultados e os índices de performance podem ser vistos na Tabela 1.

Campos de Futebol	VP	FP	FN	Precisão	Sensibilidade
22	17	7	5	70,8%	77,3%

Tabela 1. Resultados

6. Conclusão

Pelas imagens das etapas intermediárias(Figuras 2 e 3) da obtenção do descritor HOG nota-se que as marcas d'água presentes nas imagens do Google Maps(Figura 1) tem pouquíssima ou nenhuma influência no descritor HOG, permitindo assim, utilizar as imagens desse tipo aplicativo tanto para a criação dos conjuntos de treinamento quanto no processo de classificação.

A detecção automática de objetos é um processo complexo que para ter um funcionamento satisfatório depende que algumas variáveis sejam ajustadas até que se encontre um ponto ideal na detecção de objetos. As variáveis que podem ser ajustadas nessa abordagem são: conjunto de treinamento, penalidade da margem do SVM e passo e nível máximo de variação de escala. Quando não se sabe quais são os ajustes necessários para se obter um classificador com boa performance, é necessário então, repetir o processo de treinamento e classificação inúmeras vezes variando as varáveis até que se obtenha um classificador com a performance desejável.

Notou-se que a maior parte dos falsos positivos eram de objetos com formato retangular, mas em seu interior não havia qualquer semelhança com um campo de futebol. Esse trabalho ainda está em desenvolvimento e como os resultados na Seção 5 mostram, há a necessidade em repetir incessantemente todo o processo de treinamento e classificação identificando os erros e ajustando as variáveis até que se consiga uma melhora significativa no desempenho da detecção de campos de futebol, alcançando-se, assim, uma diminuição considerável nos números de falsos positivos e falsos negativos.

Referências

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, COLT '92, pages 144–152, New York, NY, USA. ACM.

Breckon, T., Barnes, S., Eichner, M., and Wahren, K. (2010). Human identity recognition in aerial images. *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*.

Dalal, N. and Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1 of *CVPR '05*, pages 886–893, Washington, DC, USA. IEEE.

Fawcett, T. (2006). An introduction to ROC analysis. In *ROC Analysis in Pattern Recognition*, volume 27, pages 861–874. Elsevier Science Inc., New York, NY, USA.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object Detection with Discriminatively Trained Part-Based Models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645.

Google. Google Maps. `http://maps.google.com`.

Gritti, T., Shan, C., Jeanne, V., and Braspenning, R. (2008). Local Features based Facial Expression Recognition with Face Registration Errors. *IEEE International Conference on Automatic Face and Gesture Recognition*.

Hsu, C. W., Chang, C. C., and Lin, C. J. (2010). A Practical Guide to Support Vector Classification.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110.

OpenCV. `http://opencv.willowgarage.com/wiki/`.

Rebouças, R. and Shiguemori, H. (2012). Acompanhamento de objetos móveis em imagens aéreas. *I Simpósio de Ciência e Tecnologia do IEAv*.

Rodrigues, R., Shiguemori, H., Forster, C., and Pellegrino, S. (2009). Color and Texture Features for Landmarks Recognition on UAV Navigation. *Anais do XIV Simpósio Brasileiro de Sensoriamento Remoto*.

SVMlight. `http://svmlight.joachims.org/`.

Theodoridis, S. and Koutroumbas, K. (2006). *Pattern Recognition, Third Edition*. Academic Press, Inc., Orlando, FL, USA.