

RECONHECIMENTO DE POSTES DA REDE ELÉTRICA A PARTIR DE IMAGENS DO GOOGLE STREET VIEW

ALLAN K. LOPES*, FABRIZIO A. A. M. N. SOARES*, LAIS C. R. S. LOPES*, GUSTAVO T. LAUREANO*, LEANDRO L. G. OLIVEIRA*, RONALDO M. COSTA*, ANDERSON S. SOARES*

* *Universidade Federal de Goiás - Instituto de Informática*
Alameda Palmeiras, Quadra D - Câmpus Samambaia - CEP 74690-900
Goiânia, GO, Brasil

Emails: allankardec.ti@gmail.com, fabrizio@inf.ufg.br, laiscandidolopes@gmail.com, gustavo@inf.ufg.br, leandroluis@inf.ufg.br, ronaldocosta@inf.ufg.br, anderson@inf.ufg.br

Abstract— Urban environments, such as streets, roads and buildings, always require maintenance and management to better use. In this sense, this paper presents an approach to recognition pole utility in street images mapped by Google Street View. Features such as color, texture and shape were investigated in order to find the best set of features to represent the objects of interest. The classification was performed using an Multilayer Perceptron trained with the Levenberg-Marquardt algorithm. Initial results show a higher accuracy by using RGB mode and texture properties as features to represent objects.

Keywords— poles of power grid, google street view, characteristics, color, texture, shape, neural network, multilayer perceptron.

Resumo— Ambientes urbanos, tais como ruas, estradas e construções, sempre demandam gerenciamento e manutenção para que sejam melhor utilizados. Nesse sentido, este artigo apresenta uma abordagem para o reconhecimento de postes da rede elétrica em imagens de ruas mapeadas pelo Google Street View. Características como cor, textura e forma foram pesquisadas com o objetivo de se encontrar o melhor conjunto de informações para representar os objetos de interesse. O reconhecimento foi realizado por uma rede neural do tipo *Multilayer Perceptron* treinada com o algoritmo *Levenberg-Marquardt*. Os resultados iniciais demonstram uma acurácia superior utilizando, de forma combinada, a moda RGB e propriedades de textura como características para representar as estruturas presentes nas imagens.

Palavras-chave— postes da rede elétrica, google street view, características, cor, textura, forma, multilayer perceptron.

1 Introdução

O Google Street View é parte integrante do Google Maps e permite aos seus usuários, através da geolocalização, a obtenção de imagens panorâmicas de ruas de vários lugares do mundo. Estas imagens possuem muitas informações úteis e constituem verdadeiro desafio para as áreas de Processamento Digital Imagens (PDI) e Visão Computacional (VC). Por exemplo, em (Hara et al., 2013) imagens do Street View foram utilizadas para identificar calçadas inacessíveis para pedestres usando técnicas de VC. O escopo era a identificação da presença ou ausência de rampas nas calçadas. Para isso, foi utilizado exemplos positivos e negativos de tais situações para treinar uma *Support Vector Machine* (SVM).

De acordo com (Yokoyama et al., 2013), a demanda por sistemas de VC que auxiliem no gerenciamento de ambientes urbanos tem aumentado bastante nos últimos anos. Estes ambientes geralmente possuem objetos de responsabilidade do poder público, tais como postes, estradas, sinais de trânsito, etc. Nesse sentido, pretende-se utilizar a API do Google Street View para demonstrar que é possível identificar postes da rede elétrica em um ambiente urbano mapeado por imagens. No trabalho de (Yokoyama et al., 2013) foi utilizado

um dispositivo de escaneamento *laser* acoplado a um carro para obter uma nuvem de pontos e assim realizar um mapeamento de ruas a fim de detectar vários objetos com características semelhantes às de um poste. O objetivo era identificar diferentes tipos de postes, como por exemplo, postes de distribuição de energia, postes de iluminação e postes de sustentação de placas de trânsito. O método de classificação utilizado era baseado na extração de características usando a classificação dos pontos por meio da transformada conhecida como *Principal Component Analysis* (PCA). Em (Barranco-Gutierrez et al., 2014) um sistema de visão foi projetado para detectar postes da rede elétrica. A ideia era integrar este sistema a um robô capaz de realizar reparos nas conexões dos cabos dos postes. Já em (Cetin et al., 2009), foi proposto um método para auxiliar um operador humano a identificar mais facilmente postes da rede elétrica em imagens aéreas. O método proposto baseava-se nas sombras projetadas por estas estruturas.

Pretende-se, neste trabalho, demonstrar e avaliar a utilização de algumas características para reconhecer postes da rede elétrica em imagens do Google Street View. As características estudadas foram utilizadas em uma rede neural do tipo *Multilayer Perceptron* (MLP) para classificar regiões,

obtidas na fase de segmenta  o, como pertencentes, ou n o, a postes da rede el trica. As caracter sticas investigadas foram cor (moda RGB da regi o), textura de Haralick (contraste, correla  o, energia e homogeneidade) e forma (Descritores El pticos de Fourier (DEF's)).

O restante deste artigo est  organizado da seguinte forma: os conceitos b sicos das t cnicas utilizadas s o apresentados na se  o 2, em seguida, exibi-se os materiais e m todos na se  o 3, os resultados s o apresentados e discutidos na se  o 4, na se  o 5 apresenta-se as conclus es e, por  ltimo, na se  o 6 comenta-se os trabalhos futuros.

2 T cnicas de Extra  o de Caracter sticas

2.1 Caracter sticas de Textura

A textura de uma imagem refere-se   distribui  o espacial e estat stica dos tons de cinza presentes nela. Haralick et al. propuseram a computa  o de um total de 28 caracter sticas a partir das matrizes de co-ocorr ncia dos n veis de cinza (Haralick et al., 1973). Essas matrizes s o uma fun  o do relacionamento  ngular entre um pixel e sua vizinhan a e, ao mesmo tempo, uma fun  o de dist ncia entre eles. Quatro  ngulos, a partir de 0 , quantificados em intervalos de 45  podem ser considerados. A vizinhan a de um pixel pode ser composta pelos pixels que est o numa dist ncia d na sua horizontal, vertical e nas suas diagonais. A Figura 1 ilustra este conceito.

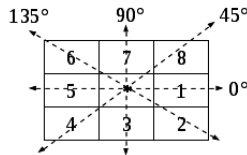


Figura 1: Representa  o da vizinhan a de um pixel *. Neste caso, $d = 1$ e todos os  ngulos s o considerados.

A partir da , uma matriz de co-ocorr ncia para cada  ngulo pode ser computada, conforme as Equa  es 1, 2, 3 e 4.

$$P(i, j, d, 0^\circ) = \# \{ ((k, l), (m, n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid k - m = 0, |l - n| = d, I(k, l) = i, I(m, n) = j \} \quad (1)$$

$$P(i, j, d, 45^\circ) = \# \{ ((k, l), (m, n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k - m = d, l - n = d) \text{ or } (k - m = -d, l - n = d), I(k, l) = i, I(m, n) = j \} \quad (2)$$

$$P(i, j, d, 90^\circ) = \# \{ ((k, l), (m, n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid |k - m| = d, l - n = 0, I(k, l) = i, I(m, n) = j \} \quad (3)$$

$$P(i, j, d, 135^\circ) = \# \{ ((k, l), (m, n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k - m = d, l - n = d), \text{ or } (k - m = -d, l - n = -d), I(k, l) = i, I(m, n) = j \} \quad (4)$$

Onde I   uma imagem, $(L_y \times L_x)$ a dimens o de I e $\#$ denota o n mero de elementos no conjunto. Cada posi  o destas matrizes armazena o n mero de vezes que dois tons de cinza (i, j) s o adjacentes. Com estas,   poss vel extrair as caracter sticas de textura, como por exemplo, contraste, correla  o, energia e homogeneidade que, por sua vez, podem ser utilizadas para caracterizar um objeto de interesse. Veja (Haralick et al., 1973) para uma lista completa e para como comput las.

2.2 Descritores El pticos de Fourier

Os descritores de Fourier (DF's) podem ser utilizados para modelar a forma de um objeto. Uma das vantagens da utiliza  o desta t cnica de representa  o   a possibilidade de se obter invari ncia a v rios problemas conhecidos como, transla  o, escala, rota  o, reflex o e ponto inicial de processamento.

Os coeficientes de Fourier possuem propriedades el pticas que foram utilizadas por (Kuhl and Giardina, 1982) para normalizar a representa  o de Fourier dos contornos dos objetos, da  o nome descritores el pticos de Fourier. Estes s o invariantes a rota  o, dilata  o e transla  o do contorno, ponto inicial de processamento, al m de n o apresentarem perda de informa  o sobre a forma do contorno como pode acontecer quando se aplica as opera  es matem ticas nos DF's a fim de se obter invari ncia aos problemas mencionados no par grafo anterior.

De acordo com (Kuhl and Giardina, 1982), os DEF's s o decomposi  es do contorno de um objeto atrav s de varia  es nas coordenadas x e y , separadamente Δx e Δy , como fun  es param tricas da dist ncia cumulativa t . Essa dist ncia v ria sempre de 0 a 2π . As equa  es 5 e 6 s o fun  es param tricas de $x(t)$ e $y(t)$.

$$x(t) = A_0 + \sum_{n=1}^N a_n \cos nt + \sum_{n=1}^N b_n \sin nt \quad (5)$$

$$y(t) = C_0 + \sum_{n=1}^N c_n \cos nt + \sum_{n=1}^N d_n \sin nt \quad (6)$$

Onde a_n, b_n, c_n e d_n representam os coeficientes da harmônica n e N é o número máximo de harmônicas. Os coeficientes para as projeções x e y são dados pelas Equações 7, 8, 9 e 10.

$$a_n = \frac{1}{n_2\pi} \sum_{p=1}^q \frac{\Delta x_p}{\Delta t_p} [\cos(nt_p) - \cos(nt_p - 1)] \quad (7)$$

$$b_n = \frac{1}{n_2\pi} \sum_{p=1}^q \frac{\Delta x_p}{\Delta t_p} [\sin(nt_p) - \sin(nt_p - 1)] \quad (8)$$

$$c_n = \frac{1}{n_2\pi} \sum_{p=1}^q \frac{\Delta y_p}{\Delta t_p} [\cos(nt_p) - \cos(nt_p - 1)] \quad (9)$$

$$d_n = \frac{1}{n_2\pi} \sum_{p=1}^q \frac{\Delta y_p}{\Delta t_p} [\sin(nt_p) - \sin(nt_p - 1)] \quad (10)$$

Onde q é número total de pontos do contorno, t_p é a distância entre o ponto p e o próximo ponto $p+1$, x_p e y_p são as projeções de p e $p+1$. Os descritores a_n, b_n, c_n e d_n podem ser utilizados para representar o contorno de um objeto dentro de um método de reconhecimento de padrões.

3 Materiais e Métodos

3.1 Conjunto de Dados

As imagens da base foram obtidas com a ajuda da API pública do Google Street View que permite, dentre outros, a obtenção de imagens panorâmicas de coordenadas geográficas. Ao todo 264 imagens foram obtidas no formato RGB com resolução de 600x600. Na Figura 2 são exibidas algumas imagens da base. Pode-se observar que as mesmas foram obtidas a partir de diferentes ângulos e distâncias. Há também a presença de bastante ruído, o que é muito comum em imagens obtidas em situações reais de operação.

Na fase de pré-processamento as imagens foram convertidas em 256 tons de cinza com o objetivo de diminuir os custos computacionais de processamento.

3.2 Segmentação

As imagens foram segmentadas utilizando a técnica de clusterização de propósito geral conhecida como *Mean Shift* em sua versão para segmentação apresentada em (Comaniciu and Meer, 2002).

Os melhores valores encontrados, após sucessivas execuções, para o parâmetro *band width* foi de 0.14. Mostra-se na Figura 3(a) a saída obtida com esta técnica. Os postes, objetos de interesse, não foram segmentados em uma única região. Isto pode ser explicado pelo fato do mesmo não ser

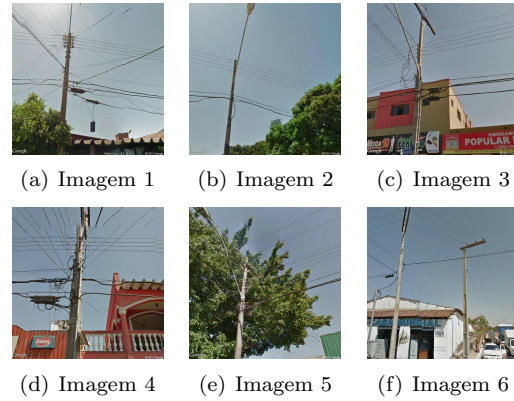
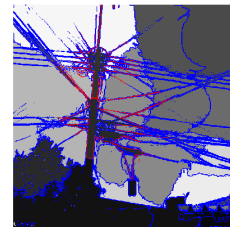


Figura 2: Imagens obtidas com ajuda da API do Google Street View.

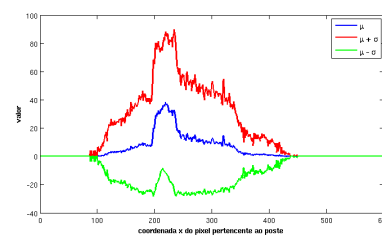
uma estrutura homogênea do ponto de vista da cor, como mostrado na Figura 3(b).

Por essa razão, considera-se que um mesmo poste pode fornecer n exemplos positivos de treinamento. No entanto, isto pode gerar mais exemplos negativos que positivos. Assim, no momento da separação das amostras para os conjuntos de treinamento e teste, a mesma quantidade de amostras de postes e não poste foi considerada.

Após a segmentação, as regiões foram manualmente rotuladas como pertencente ou não pertencente a poste. Isto é necessário porque o algoritmo de treinamento do classificador é do tipo supervisionado.



(a) Resultado da segmentação



(b) Desvio padrão dos pixels do poste

Figura 3: Na Figura 3(a) Exemplo de retorno do algoritmo *Mean Shift*. Delineado de vermelho, regiões classificadas manualmente como pertencentes ao poste. Na Figura 3(b) apresenta-se o desvio padrão dos pixels pertencentes ao poste. Onde, μ é a média dos pixels das colunas pertencente ao poste e σ é o desvio padrão. Observa-se que esta estrutura não possui coloração uniforme.

3.3 Representação

Para representar os objetos obtidos na fase de segmentação foram pesquisadas a moda RGB, as propriedades da textura e os DEF's.

Com relação às características de textura, as utilizadas nesta fase da pesquisa foram contraste, correlação, energia e homogeneidade computadas de acordo com as equações 11, 12, 13 e 14 respectivamente.

$$contraste = \sum_{i,j} |i - j|^2 p(i, j) \quad (11)$$

$$correlacao = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (12)$$

$$energia = \sum_{i,j} p(i, j)^2 \quad (13)$$

$$homogeneidade = \sum_{i,j} \frac{p(i, j)}{1 + |i - j|} \quad (14)$$

Onde $p(i, j)$ representa a matriz de co-ocorrência dos níveis de cinza. Para a sua obtenção foram considerados todos os ângulos (0° , 45° , 90° , e 135°) e distância $d = 30$.

Apenas os 8 DEF's mais representativos foram utilizados. Esses valores foram escolhidos após a aplicação do PCA. Cada descriptor é composto por 4 valores. Assim, 8 DEF's dá um total de 32 características.

Portanto, um vetor de características de um objeto pode ser composto por um total de até 43 valores divididos da seguinte forma, 3 valores que representam a moda RGB (um valor por plano), 8 valores das características de textura e as 32 características obtidas através dos 8 DEF's mais representativos.

3.4 Classificação

Uma RNA foi treinada com o algoritmo descrito em (Hagan and Menhaj, 1994) para reconhecer os objetos. Os objetos foram agrupados em duas classes, poste e não poste. 70% das amostras foram utilizadas na fase de treinamento, enquanto, os outros 30% restantes na fase de teste. De acordo com o explicado na subseção 3.2, foi considerado o mesmo número de amostras de poste e não poste nos dois conjuntos.

Em relação à arquitetura, a RNA possui 10 neurônios na camada oculta, quantidade obtida empiricamente a partir dos dados de erro médio quadrático, e um na camada de saída. Os neurônios da camada oculta usam a função de ativação sigmóide tangente hiperbólica, conhecida por acelerar a convergência do algoritmo de treinamento (Vogl et al., 1988). Já o neurônio da camada de saída usa a função de ativação linear.

Durante o treinamento, dois critérios de parada foram adotados: a quantidade de ciclos = 200 ou; erro médio quadrático atingir o valor mínimo de 0.0416. Na Figura 4 é apresentado um resumo de todas as etapas descritas até este ponto.

3.5 Experimentos

Objetivando encontrar o melhor subconjunto de características, o classificador foi treinado com várias combinações das características descritas na subseção 3.3.

Com o melhor subconjunto de características, conduziu-se um experimento para determinar o melhor valor de limiar da rede. Este valor é utilizado para dizer se uma saída pertence ou não a um poste.

Os modelos treinados com diferentes características foram avaliados a partir de suas respectivas matrizes de confusão.

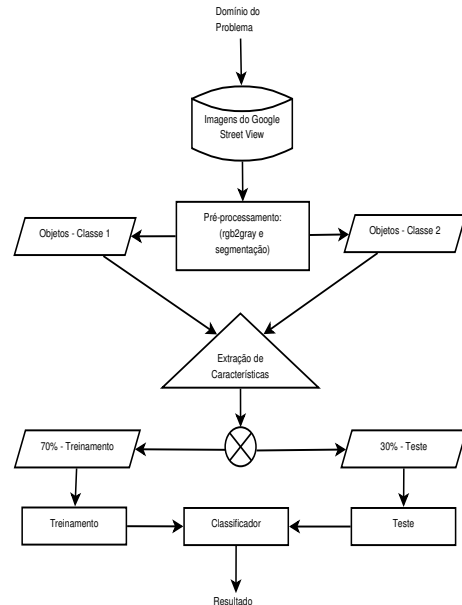


Figura 4: Resumo das etapas do modelo proposto.

4 Resultados

O processo de segmentação resultou em 7039 segmentos, destes, apenas 635 pertenciam a postes. Como explicado anteriormente, uma imagem pode fornecer vários exemplos de treinamento positivos e negativos. Isso acontece, no que diz respeito ao poste, porque ele não é uma estrutura de cor uniforme.

Pode-se observar, na Figura 5, que o valor para o *limiar* que maximiza o reconhecimento dos objetos pertencentes a poste, sem perdas significativas na taxa de acertos para os objetos não pertencentes a poste, é $\text{limiar} = -0,2$.

Na Tabela 1, os indicadores calculados a partir das matrizes de confusão montadas após o trei-

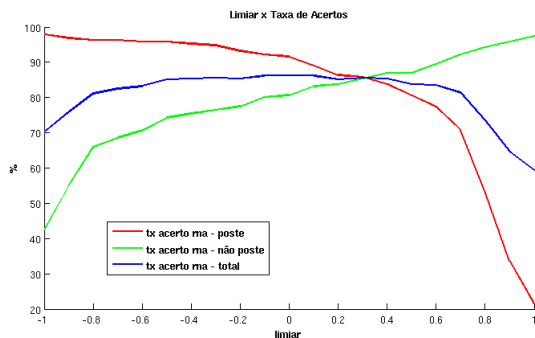


Figura 5: Resposta do classificador para diferentes valores de *limiar*.

namento com o conjunto de teste. Pode-se observar que a acurácia do classificador treinado com representações dos objetos compostas da moda RGB e das características de textura de Haralick é superior à acurácia dos demais e que a adição de outras características não converge para o aumento desta acurácia. No entanto, a taxa de verdadeiro positivo do classificador treinado somente com a moda RGB foi superior às dos demais.

Tabela 1: Indicadores calculados a partir das matrizes de confusão.

	Moda RGB	Ha-ralick	Moda RGB + Ha-ralick	DEF	To-das
Accuracy	0,8168	0,8586	0,8717	0,6257	0,7827
Error Rate	0,1832	0,1414	0,1283	0,3743	0,2173
True Positive Rate	0,9110	0,9058	0,8848	0,7277	0,8220
False Positive Rate	0,2775	0,1885	0,1414	0,4764	0,2565
Specificity	0,7225	0,8115	0,8586	0,5236	0,7435
Precision	0,7665	0,8278	0,8622	0,6043	0,7621
Prevalence	0,5000	0,5000	0,5000	0,5000	0,5000

De acordo com o que pode ser visualizado na Figura 6, o modelo 1 é o que foi proposto neste trabalho, já o modelo 2 foi proposto em (Yokoyama et al., 2013) e o modelo 3 em (Barranco-Gutierrez et al., 2014). Apesar da acurácia de 100% obtida

no modelo 3, vale ressaltar que estes resultados foram obtidos em cima do conjunto utilizado para treinar o classificador, ou seja, este já conhecia os padrões. Portanto, conforme pode ser observado, o modelo proposto neste trabalho apresentou, usando como características a moda RGB e textura de Haralick, acurácia superior à do modelo 2.

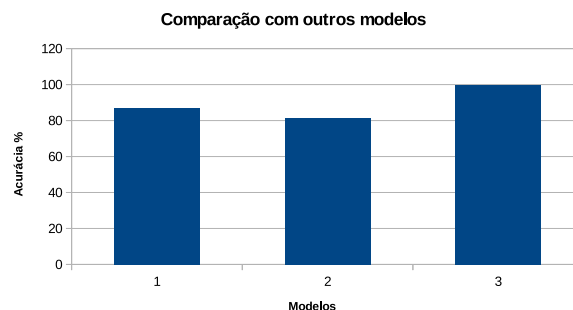


Figura 6: Comparação da acurácia obtida com a acurácia obtida em outros modelos.

Por último, na Figura 7, exemplos de resultados obtidos com um protótipo construído para exemplificar a utilização do classificador apresentado.

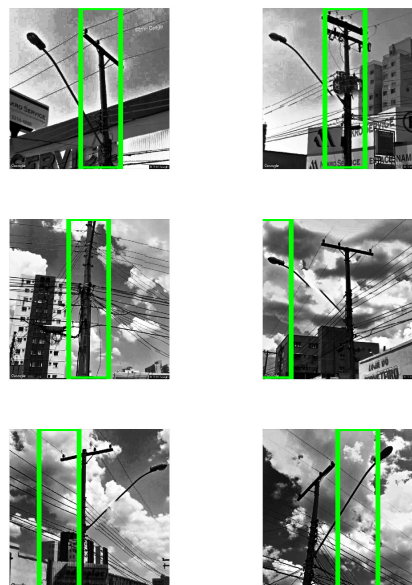


Figura 7: Exemplos de resultados obtidos com o protótipo desenvolvido a partir do classificador construído.

5 Conclusões

Neste trabalho, foi apresentado um estudo de algumas características extraídas de imagens do Google Street View e que podem vir a serem utilizadas em um sistema de reconhecimento de postes

da rede elétrica. Os postes, presentes nas imagens, se apresentaram como estruturas de cor não uniforme o que dificultou segmentá-los como objetos únicos durante a fase de segmentação. Desta forma, foi necessário considerar que uma mesma imagem pode fornecer vários exemplos positivos e negativos. Logo, é necessário aperfeiçoar a técnica de segmentação utilizada.

Por conta do grande número de estruturas segmentadas com formatos diferentes, a utilização de características da forma, aqui representada pelos DEF's, demonstrou-se não muito exitosa.

A informação da cor do objeto, representada pela moda RGB de suas cores, incrementou a acurácia do classificador quando utilizada em conjunto com as características de textura de Haralick se comparado com a utilização de ambas as informações em separado. A acurácia utilizando estas informações foi de 87%.

6 Trabalhos Futuros

Como trabalhos futuros, pretende-se:

- aperfeiçoar a fase de segmentação;
- aumentar o número de amostras da base de dados;
- utilizar os DEF's para ajudar na tomada de decisão de classificar um determinado objeto como poste ou não;
- treinar o classificador para reconhecer diferentes tipos de postes;
- criar um sistema que navegue automaticamente nas ruas do Google Street View em busca de postes, retornando suas coordenadas geográficas.

Agradecimentos

Este trabalho contou com o suporte da Fundação de Amparo à Pesquisa do Estado de Goiás (Fapeg), da Celg Distribuição S.A. (Celg-D), do Instituto de Informática da Universidade Federal de Goiás (UFG) e do Instituto Federal Goiano - Campus Posse.

for Utility Pole Recognition in Real Conditions, in F. Huang and A. Sugimoto (eds), *Image and Video Technology - PSIVT 2013 Workshops*, number 8334 in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 113–121.

Cetin, B., Bikdash, M. and McInerney, M. (2009). Automated electric utility pole detection from aerial images, *IEEE Southeastcon, 2009. SOUTHEASTCON '09*, pp. 44–49.

Comaniciu, D. and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5): 603–619.

Hagan, M. T. and Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks* **5**(6): 989–993.

Hara, K., Le, V., Sun, J., Jacobs, D. and Froehlich, J. (2013). Exploring Early Solutions for Automatically Identifying Inaccessible Sidewalks in the Physical World Using Google Street View, *HCIC 2013*.

Haralick, R. M., Shanmugam, K. and Dinstein, I. (1973). Textural Features for Image Classification, *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3**(6): 610–621.

Kuhl, F. P. and Giardina, C. R. (1982). Elliptic Fourier features of a closed contour, *Computer Graphics and Image Processing* **18**(3): 236–258.

Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T. and Alkon, D. L. (1988). Accelerating the convergence of the back-propagation method, *Biological Cybernetics* **59**(4-5): 257–263.

Yokoyama, H., Date, H., Kanai, S. and Takeda, H. (2013). Detection and Classification of Pole-like Objects from Mobile Laser Scanning Data of Urban Environments, *International Journal of CAD/CAM* **13**(2).

FAPEG

FUNDAÇÃO DE AMPARO
À PESQUISA
DO ESTADO DE GOIÁS



CELG
DISTRIBUIÇÃO



**INSTITUTO FEDERAL
GOIANO**
Câmpus Posse

Referências

Barranco-Gutierrez, A. I., Martinez-Diaz, S. and Gomez-Torres, J. L. (2014). An Approach