

Reconhecimento preliminar de espécies de cigarrinhas (Cicadellidae: Cicadellinae) com o uso de algoritmo de aprendizado de máquina e representação esparsa

Preliminary identification of leafhopper species (Cicadellidae: Cicadellinae) using a machine learning algorithm and sparse representation

Leonardo Cechet Moro

Cícero Zanoni

Wilson Sampaio de Azevedo Filho

Universidade de Caxias do Sul - UCS - Caxias do Sul - Rio Grande do Sul - Brasil

Resumo

Este artigo descreve as etapas para o desenvolvimento de um algoritmo de reconhecimento de cigarrinhas (Cicadellidae: Cicadellinae), através do uso de um computador e uma câmera fotográfica digital. São demonstrados os resultados com o uso de extração de características SURF e o desenvolvimento teórico para o uso de representação esparsa e de algoritmos de aprendizado de máquina.

Palavras-chave

Identificação. Cigarrinhas. Algoritmo de aprendizado de máquina. Representação esparsa.

Abstract

This paper describes the steps for the development of an algorithm to perform the identification of leafhoppers (Cicadellidae: Cicadellinae) using of a computer and a digital camera. Additionally, it demonstrates the preliminary results that were obtained with the use of SURF features extraction, and the theoretical development for the use of sparse representation and machine learning algorithms.

Keywords

Identification. Leafhoppers. Machine learning algorithm. Sparse representation.

1. Introdução

Os Cicadellidae constituem um grupo de cigarrinhas numeroso com mais de 50 subfamílias o que dificulta a caracterização dos seus integrantes (NIELSON, 1985). As espécies apresentam um comprimento variado, desde 3,4 mm até 22 mm, com cores vistosas e contrastantes (YOUNG, 1968; MEJDALANI, 1998).

O número de espécies de Cicadellidae descritas atinge em torno de 21.000 (MEJDALANI, 1998), com estimativas de que essa quantidade possa superar os 45.000 (NIELSON, 1985). Contudo, as cigarrinhas incluídas nessa família para a Região Neotropical têm permanecido pouco conhecidas, se comparada com outras regiões.

O grupo Cicadellidae inclui espécies de reconhecida importância agrícola no território brasileiro e ainda são poucas as informações sobre a distribuição, plantas hospedeiras e danos causados por essas cigarrinhas (ZANOL e MENEZES, 1982). Esses insetos se estabeleceram em plantas não relacionadas às hospedeiras originais e tornaram-se pragas de importância econômica (NIELSON, 1985). Os danos produzidos pelos cicadélídeos podem ser divididos em dois grupos: danos diretos (sucção da seiva, ação tóxica da saliva, lesões provocadas pelas posturas e outros) e danos indiretos (doenças causadas por fitopatógenos, onde esses insetos atuam como vetores) (LIMA, 1942; NIELSON; 1985; GALLO *et al.*, 1988).

No Brasil, houve um aumento significativo no número de cicadélíneos vetores, trazendo preocupações aos produtores, extensionistas e pesquisadores. Dessa forma, doze espécies se destacam como comprovadamente capazes de transmitir um importante fitopatógeno, a bactéria *Xylella fastidiosa*, que tem causado graves prejuízos em diferentes culturas: *Acrogonia citrina* Marucci & Cavichioli, 2002; *Acrogonia virescens* (Metcalf, 1949); *Bucephalagonia xanthophis* (Berg, 1879); *Dilobopterus costalimai* Young, 1977; *Ferrariana trivittata* (Signoret, 1854); *Fingeriana dubia* Cavichioli, 2003; *Homalodisca ignorata* Melichar, 1924; *Macugonalia leucomelas* (Walker, 1851); *Oncometopia facialis* (Signoret, 1854); *Parathona gratiosa* (Blanchard, 1840); *Plesiommata corniculata* Young, 1977 e *Sonesimia grossa* (Signoret, 1854) (LOPES *et al.*, 1996; ROBERTO, 1996a; ROBERTO, 1996b; ROBERTO *et al.*, 1996; YAMAMOTO e ROBERTO, 1997; FUNDECITROS, 1999; LOPES, 1999; YAMAMOTO *et al.*, 2000). Esta bactéria é responsável por doenças como a Escaldadura das Folhas da Ameixeira (HICKEL *et al.*, 2001), Clorose Variegada dos Citros (DE NEGRI, 1990) e Requeimadas-Folhas ou Atrofia-dos-Ramos do Cafeeiro (LEITE JÚNIOR, 2002; PAIÃO *et al.*, 2002).

É fundamental o desenvolvimento de estratégias que permitam o reconhecimento das espécies potenciais vetoras da *X. fastidiosa* nas diferentes culturas para viabilizar a implementação de técnicas de manejo e controle das cigarrinhas, diminuindo as perdas econômicas associadas às doenças.

O reconhecimento de imagens é uma técnica que pode ser utilizada para solução de diversos tipos de problemas. Assim, devido à demanda no reconhecimento de insetos, alguns estudos já foram desenvolvidos na área (LARIOS *et al.* 2008; LU *et al.* 2010). O reconhecimento de insetos pode ser amplamente aplicado na área de Ciências Biológicas e Agrárias. Uma vez que essa é uma atividade que requer alta *expertise*, sem experiência profissional, é muito difícil realizar o reconhecimento das espécies de insetos, mesmo que de forma preliminar. Dessa forma, há uma alta demanda por algoritmos automáticos que resolvam esse problema.

Essa aplicação é desafiadora por uma série de motivos. Apesar de cada espécie apresentar características próprias, que podem ser identificadas por taxonomistas, existe a possibilidade da ocorrência de dúvidas entre pessoas não treinadas examinando diferentes táxons. Para um leigo, por exemplo, algumas espécies são relativamente fáceis de identificar, enquanto outras são extremamente difíceis, podendo acarretar em possíveis erros (LARIOS *et al.*, 2008). Além disso, há uma variabilidade na posição, tamanho e qualidade das imagens obtidas dos espécimes para uma possível identificação. Portanto, desenvolver uma aplicação que seja capaz de prever e integrar essas variações é um desafio de visão computacional.

O reconhecimento de insetos através de algoritmos de visão computacional tem sido alvo de pesquisas em todo o mundo (LARIOS *et al.*, 2008). O DAISY (*Digital Automated Identification System*) é um sistema de identificação de diversos grupos de insetos, baseado na teoria de reconhecimento facial *eigen-images* (TURK e PENTLAND, 1991). Uma das desvantagens do sistema é o fato de que o DAISY requer interação do usuário para captura de imagem e segmentação (LARIOS *et al.*, 2008). Um segundo sistema conhecido é o ABIS (*Automated Bee Identification System*), cuja função é realizar a identificação de abelhas baseado em características extraídas de suas asas (LARIOS *et al.*, 2008). O SPIDA-web (*Species Identification, Automated and web accessible*) é um sistema de identificação automatizado e acessível à web, que utiliza redes para identificação de espécies baseado em imagens codificadas através de *wavelets* (LU *et al.*, 2010).

Através do uso de histograma de cores e Matriz de Co-ocorrência de Níveis de Cinza (GLCM, *Gray Level Co-Occurrence Matrix*), pesquisadores da Universidade de Zhejiang Gongshang atingiram 71,1% de eficiência no reconhecimento de insetos do grupo Lepidoptera (ZHU e ZHANG, 2010). A técnica consiste inicialmente em realizar a segmentação da imagem, converter a imagem de RGB (*Red-Green-Blue*) para HSV (*Hue-Saturation Level*) e então gerar os histogramas.

Com base em técnicas de reconhecimento de padrão, Lim *et al.* (2006) demonstraram um sistema capaz de identificar lepidópteros e coleópteros. As imagens obtidas sofrem um processamento para rejeição de “ruídos” através do método de informações da cor. É realizada uma detecção de borda e a identificação do inseto atingiu eficiência na ordem de 90%. Em uma linha semelhante de pesquisa, Yang *et al.* (2010) obteve até 100% de eficiência na identificação de oito diferentes espécies de insetos. O sistema consiste na extração de 14 diferentes características e um algoritmo de aprendizado de máquina denominado *Random Trees*.

O trabalho teve como objetivo criar uma ferramenta computacional para identificar, de forma preliminar, espécies de cigarrinhas (Cicadellidae: Cicadellinae) vetoras da *X. fastidiosa*.

2. Material e métodos

O método proposto para este trabalho foi baseado em um algoritmo de aprendizado de máquina, através da extração de características de um conjunto de imagens de treino, para posterior identificação dos táxons. O fluxograma do sistema é apresentado na figura 1.

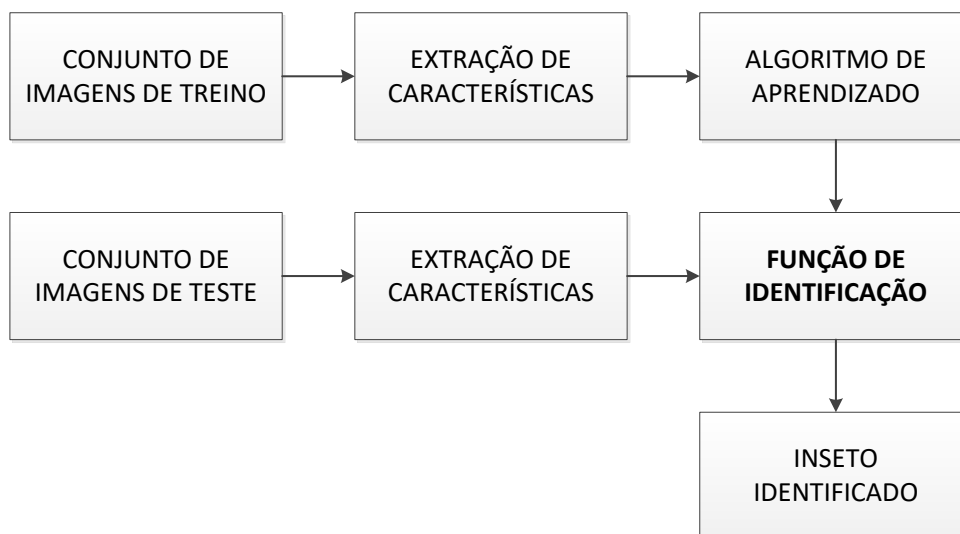


Figura 1. Fluxograma do sistema de reconhecimento dos espécimes.

Assim como qualquer sistema de aprendizado de máquina, a extração das características é a etapa mais importante de todo o processo. É necessário que as características sejam devidamente extraídas e analisadas, para que o sistema funcione como esperado. Esse estudo teve como base o método de representação esparsa, características SIFT e *bag of features*.

Esse método foi apresentado por Lu *et al.* (2010) e demonstrou resultados positivos para o reconhecimento de insetos, assunto que ainda é pouco explorado na literatura.

A representação esparsa é utilizada para representar um vetor de entrada como uma combinação linear de um número pequeno de vetores base. Wright *et al.* (2009) demonstrou resultados utilizando a técnica para fins de reconhecimento facial.

De acordo com Lu *et al.* (2010), para cada vetor de entrada x , pode-se obter C vetores de coeficiente s , para cada matriz base através da seguinte equação (Figura 2):

$$\min_{s^{(i)}} \|x - B_i s^{(i)}\|_2^2 + \lambda \|s^{(i)}\|_1, \text{ para } i=1,2,\dots,C$$

Figura 2. Equação segundo Lu *et al.* (2010).

O fluxograma do sistema funciona basicamente da seguinte forma, segundo o algoritmo proposto por Lu *et al.* (2010): um banco de dados é utilizado para treino e teste. Em seguida, as imagens são transformadas para escala de cinza e as características SIFT são extraídas de patches (conjuntos de pixels da imagem). Matrizes bases B_i são extraídas do conjunto de treino através das equações (Figuras 3 e 4):

$$\min_{s_{n_i}^{(i)}} \|x_{n_i}^{(i)} - B_i s_{n_i}^{(i)}\|_2^2 + \lambda \|s_{n_i}^{(i)}\|_1$$

Figura 3. Equação segundo Lu *et al.* (2010).

$$\min_{B_i} \sum_{n_i=1}^{N_i} \|x_{n_i}^{(i)} - B_i s_{n_i}^{(i)}\|_2^2 \text{ s.t. } \|b_{i,k_i}\|_2 \leq c, \forall k_i=1,2,\dots,K_i$$

Figura 4. Equação segundo Lu *et al.* (2010).

Conforme a equação (Figura 2), cada característica SIFT pode ser traduzida em vetores esparsos, que serão concatenados em um único vetor. Duas técnicas podem então ser utilizadas para concatenar os vetores de coeficiente C , a fim de representar o vetor original de entrada. A primeira delas é chamada *Minimal Residual Class Specific Sparse Representation* (MRCSSR). A segunda é a *Sparsest Class Specific Sparse Representation* (SCSSR).

Por fim, essas características podem ser utilizadas por algoritmos de aprendizado de máquina. Esse classificador pode ser do tipo SVM (*Support Vector Machine*) com kernel gaussiano. Essa técnica consiste em encontrar o vetor (ou os vetores) que mais se adequa à hipótese, de modo que o algoritmo é capaz de aprender a reconhecer os padrões.

Os testes do presente trabalho foram realizados no software MATLAB 2013a, utilizando um algoritmo que extrai características SURF (*Speeded Up Robust Features*) - parcialmente inspirado em características SIFT - de uma imagem e compara com um banco de imagens.

Para criação e organização do banco de imagens digitais utilizado na comparação dos táxons foram confeccionadas fotos das cigarrinhas com o auxílio de câmera fotográfica digital acoplada a um estereomicroscópio trinocular modelo SMZ800 (NIKON).

3. Resultados e discussões

Como os cicadelídeos (cigarrinhas com variado padrão de cores e tamanhos) representam um grupo grande, diversificado e de difícil caracterização (YOUNG, 1968; TRIPLEHORN e JOHNSON, 2011), a metodologia desenvolvida no estudo possibilitou realizar uma comparação direta entre as características extraídas dos espécimes. O algoritmo extrai características de um banco de dados e da imagem que se deseja identificar (Figura 5).

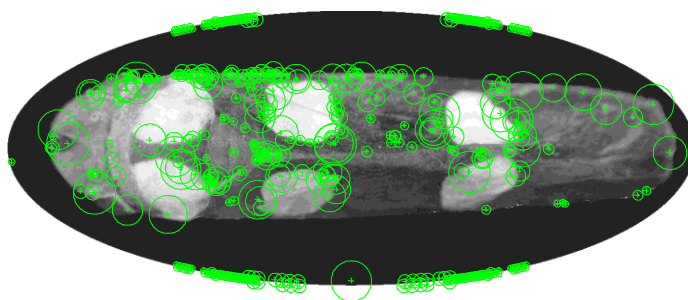


Figura 5. Características SURF do espécime isolado utilizado para comparação.

A figura 6 permite visualizar o processo de comparação e reconhecimento do espécime isolado, à esquerda, com a utilização do banco de dados à direita. É possível observar também que há um conjunto de linhas que compara as características entre as imagens permitindo realizar a identificação da cigarrinha. Assim, o algoritmo é capaz de identificar o táxon correspondente.

O sistema de reconhecimento já está estruturado e a próxima etapa será o desenvolvimento do algoritmo que extrai as características SIFT e utiliza o método de representação esparsa. Em seguida, o uso de algoritmos estado-da-arte em aprendizado de máquina, para a identificação dos insetos. O trabalho foi desenvolvido em MATLAB, utilizando-se os métodos já disponíveis no software, como é o caso do SVM (*Support Vector Machine*) e da extração de características SIFT. Contudo, o estudo já possibilitou desenvolver uma ferramenta computacional inicial capaz de reconhecer de forma preliminar cigarrinhas (Cicadellidae: Cicadellinae), com potencial para transmitir a *X. fastidiosa* em diferentes culturas no Brasil, através de imagem obtida a partir de uma câmera fotográfica digital de alta resolução acoplada a um estereomicroscópio.

Dessa forma, será possível desenvolver um sistema completo que possa realizar a identificação das espécies de cigarrinhas vetoras atendendo as necessidades de extencionistas, produtores e profissionais das diferentes áreas.

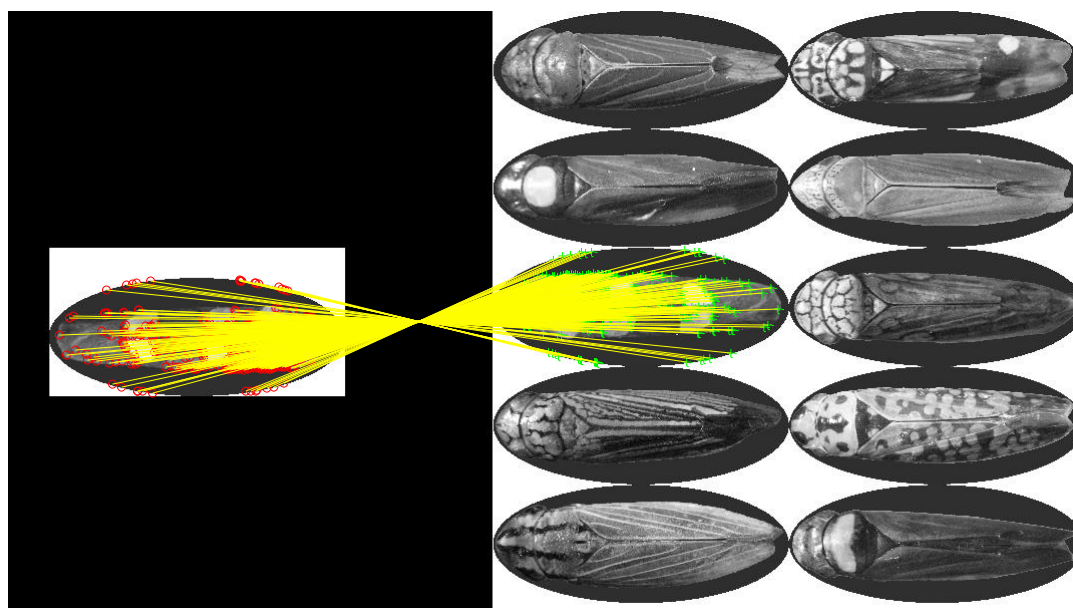


Figura 6. Processo de comparação e reconhecimento do espécime isolado com a utilização do banco de dados.

4. Conclusões

A ferramenta computacional proposta permitiu reconhecer de forma preliminar espécies de cigarrinhas (Cicadellidae: Cicadellinae) vetoras da *X. fastidiosa*.

As características SURF podem ser utilizadas para a comparação entre um espécime isolado e um banco de dados com diferentes espécies.

As características SIFT, ou semelhantemente - SURF, em conjunto com um algoritmo de aprendizado de máquina, podem ser utilizadas para treinar um banco de dados que seja variado e extenso o suficiente.

Referências

1. DE NEGRI, J.D. **Clorose variegada dos citros**: nova anomalia afetando pomares em São Paulo e Minas Gerais. Campinas: CATI, 1990.
2. FUNDECITRUS. Descobertos mais 6 vetores de CVC. **Revista do Fundecitrus**, Araraquara, n.94, p.8-9, 1999.
3. GALLO, D., NAKANO, O.; SILVEIRA-NETO, S.; CARVALHO, R.P.L.; BATISTA, G.C.; BERTI-FILHO, E.; PARRA, J.R.P.; ZUCCHI, R.A.; ALVES, S.B.; VENDRAMIN, J.D. **Manual de Entomologia Agrícola**. São Paulo: Ed. Agronômica Ceres, 1988. 649p.
4. HICKEL, E. R.; DUCROQUET, J. P. H. J.; LEITE JUNIOR, R. P.; LEITE, R. M. V. B. C. Fauna de Homoptera: Auchenorrhyncha em pomares de ameixeira de Santa Catarina. **Neotropical Entomology**, Londrina, v. 30, n. 4, p. 725-729, 2001.
5. LEITE JÚNIOR, R.P. Ocorrência de *Xylella fastidiosa* em café no Brasil. **Summa Phytopathologica**, Botucatu, v. 28, p. 122-124, 2002.

6. LARIOS, N.; DENG, H.; ZHANG W.; SARPOLA, M; YUEN, J.; PAASCH, R.; MOLDENKE, A.; LYTLE, D. A.; CORREA, S.R; MORTENSEN, E.N.; SHAPIRO, L. G.; DIETTERICH, T.G. Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. **Machine Vision and Applications**, New York, v.19, n.2, p.105-123, 2008.
7. LIM, J.; CHO, J.; NAM, T.; KIM, S. Development of a classification algorithm for butterflies and ladybugs. In: **TENCON 2006. 2006 IEEE Region 10 Conference**, IEEE, p. 1-3, 2006.
8. LIMA, A.C. **Insetos do Brasil**. 3º tomo: Homopteros. Rio de Janeiro: Escola Nacional de Agronomia, 1942. 327p.
9. LOPES, J.R.S. Estudos com vetores de *Xylella fastidiosa* e implicações no manejo da clorose variegada dos citros. **Laranja**, Cordeirópolis, v.20, n.2, p.329 -344, 1999.
10. LOPES, J.R.S.; BERETTA, M.J.G.; HARAKAVA, R.; ALMEIDA, R.P.P.; KRÜGNER, R.; GARCIA-JUNIOR, A. Confirmação da transmissão por cigarrinhas do agente causal da clorose variegada dos citros, *Xylella fastidiosa*. **Fitopatologia Brasileira**, Brasília, v. 21, suplemento, p.343, 1996.
11. LU A.; HOU X., LIN C.; LIU C-L. Insect Species Recognition using Sparse Representation. **Proceedings of the British Machine Vision Conference**, United Kingdom, p.108, 2010.
12. MEJDALANI, G. Morfologia externa dos Cicadellinae (Homoptera, Cicadellidae): comparação entre *Versigonalia ruficauda* (Walker) (Cicadellinae) e *Tretogonia cribrata* Melichar (Proconiini), com notas sobre outras espécies e análise da terminologia. **Revista Brasileira de Zoologia**, Curitiba, v. 15, n.2, p.451-544, 1998.
13. NIELSON, M.W. Leafhopper systematics. In: NAULT, L.R.; RODRIGUES, J.G. (Eds.) **The leafhoppers and planthoppers**. New York: John Wiley & Sons, 1985. p.11-40.
14. PAIÃO, F.G.; MENEGUIM, A.M.; CASAGRANDE, E.C.; LEITE JÚNIOR, R.P. Envolvimento de cigarras (Homoptera, Cicadidae) na transmissão de *Xylella fastidiosa* em cafeeiro. **Fitopatologia Brasileira**, Brasília, DF, v. 27, p. 67, 2002. Suplemento.
15. ROBERTO, S.R. Cigarrinha - como identificar os vetores. **Revista do Fundecitrus**, Araraquara, v.12, n.77, p.10 –11, 1996a.
16. ROBERTO, S.R. Cigarrinhas transmissoras da Clorose Variegada dos Citros. **Correio Agrícola**, São Paulo, n.2, p. 5-6, 1996b.
17. ROBERTO, S.R.; COUTINHO, A.; LIMA, J.E.O.; MIRANDA, V.S.; CARLOS, E.F. Transmissão de *Xylella fastidiosa* pelas cigarrinhas *Dilobopterus costalimai*, *Agrogonia terminalis* e *Oncometopia facialis* em Citros. **Fitopatologia Brasileira**, Brasília, v.21, n.4, p.517-518, 1996.
18. TRIPLEHORN, C.A.; JOHNSON N.F. **Estudo dos insetos**. São Paulo: Cengage Learning, 2011. 809p.
19. TURK, M.A.; PENTLAND, A.P. Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., **IEEE Computer Society Conference on. IEEE**, 1991. p. 586-591.

20. WRIGHT, J., YANG, A.Y., GANESH, A., SASTRY, S.S.; MA, Y. Robust face recognition via sparse representation. **IEEE transactions on pattern analysis and machine intelligence**, v. 31, n. 2, p. 210-227, 2009.
21. YAMAMOTO, P.T.; ROBERTO, S.R. Controle químico das cigarrinhas transmissoras da CVC dos citros. **Correio Agrícola**, São Paulo, 2. ed., p. 3-5, 1997.
22. YAMAMOTO, P.T.; ROBERTO, S.R.; PRIA-JÚNIOR, W.D.; FELIPPE, M.R.; MIRANDA, V.S.; TEIXEIRA, D.C.; LOPES, J.R.S. Transmissão de *Xylella fastidiosa* pelas cigarrinhas *Homalodisca ignorata*, *Acrogonia virescens* e *Molomea cincta* (Hemiptera: Cicadellidae) em plantas cítricas. **Summa Phytopathologica**, Jaboticabal, v.26, n.1, p.128, 2000.
23. YANG, H.; LIU, W.; XING, K.; QIAO, J.; WANG, X.; GAO, L.; SHEN, Z. Research on insect identification based on pattern recognition technology. In: **Natural Computation (ICNC), 2010 Sixth International Conference on**, IEEE, v. 2, p. 545-548, 2010.
24. YOUNG, D.A. Taxonomic study of the Cicadellinae, part 1, Proconiini. **U. S. National Museum Technical Bulletin**, Washington, v. 261, p. 1-287, 1968.
25. ZANOL, K.M.R.; MENEZES, M. Lista preliminar dos cicadélídeos (Homoptera, Cicadellidae) do Brasil. **Iheringia**, Porto Alegre, v. 61, p.9-65, 1982.
26. ZHU, L. Q.; ZHANG, Z. Auto-classification of insect images based on color histogram and GLCM. In: **Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on**, IEEE, v. 6, p. 2589-2593, 2010.