

Computer Vision for Larval Structures Identification applied to Forensic Science

CARINA ELISEI OLIVEIRA, LUCAS RODRIGUES DE LIMA, GLAUCIA RAQUEL ASSIS DE OLIVEIRA, ARIADNE BARBOSA GONÇALVES, HEMERSON PISTORI

UNIVERSIDADE CATÓLICA DOM BOSCO, CAMPO GRANDE - MS, BRASIL

carinaelisei@yahoo.com.br, lukas6923@gmail.com,
glauciaraque.assis@gmail.com, ariadne.gon@gmail.com,
pistori@ucdb.br.

WILSON WERNER KOLLER
EMBRAPA GADO DE CORTE
CAMPO GRANDE - MS, BRASIL
wilson.koller@embrapa.br

Abstract—The diptera maggots are used in forensic entomology to estimate the post-mortem interval (PMI). Maggots have a wide range of morphological and structural features that aid in the identification. In order to assist in the necrophagous larvae identification, this research aims to develop a software using computer vision and machine learning to automate the classification process. Diptera maggots were collected in a dead pig at the capital of Mato Grosso do Sul state, Campo Grande. The maggots were identified and photographed at a light microscope (5x objective). Next, the images were processed, the features extraction was performed using an extractor in Python language. The classification of the images were tested with AdaBoost, Random Forest, Random Tree and SMO classifiers. The SMO the best performance of 64.41%, using the area under a ROC curve metric.

Keywords—Entomology, fly larvae, computer vision

I. INTRODUÇÃO

Entomologia forense é o estudo de insetos, dentre outros artrópodes, associados a questões criminais, sendo um instrumento útil na investigação de crimes [1]. Entre as diversas aplicações da Entomologia na prática forense, a mais utilizada é a identificação de insetos envolvidos na decomposição de cadáveres para a estimativa do intervalo post-mortem (IPM), particularmente nos casos em que o ciclo de vida de insetos necrófagos é a única ferramenta disponível para estimar o IPM [5]. Entretanto, a identificação de algumas espécies de insetos pode ser complicada por diversos fatores, mesmo para taxonomistas experientes, sobretudo dos espécimes mais comumente encontrados em cadáveres, estágios imaturos e pupários vazios [1].

As moscas da família Calliphoridae apresentam distribuição cosmopolita. Existem mais de 1000 espécies e

cerca de 150 gêneros conhecidos. Mais de 40 espécies em 30 gêneros são restritas ao oriente. Os dípteros califorídeos apresentam os seguintes estágios de desenvolvimento: ovo, larva, o estágio larval em três instares: L1, L2 e L3[9].

Este trabalho tem como principal contribuição a validação de técnicas de visão computacional para identificação de larvas de insetos necrófagos através de um inédito banco de imagens construído para esta finalidade. Na próxima seção serão apresentados alguns trabalhos correlatos. As Seções III, IV e V tratam da metodologia, do banco de imagens e das técnicas exploradas. Por fim, apresenta-se os experimentos, resultados, discussão e conclusões.

II. TRABALHOS CORRELATOS

Para [5] o sistema automático ideal para detecção de pragas teria de ser levado a cabo juntamente com a gravação de informações diárias sobre cada espécie de inseto. Além disso, seria melhor que o sistema fosse capaz de identificar a fase de desenvolvimento com base no tamanho, forma e cor dos insetos, além de discriminar entre qualquer material estranho, de tal maneira que possibilitasse integrar estes dados em um sistema de diagnóstico automático [8]. Neste contexto, a visão de máquina que analisa imagens computadorizadas demonstrou ter um grande potencial para detecção e identificação dos insetos [8].

Segundo [3] há muitas aplicações de ciência ambiental que poderiam se beneficiar de métodos de visão computacional de baixo custo. Poucos projetos podem justificar a despesa de ter entomologistas peritos para classificar manualmente espécimes coletados em campo. Reconhecimento de espécies de inseto precisa de um conhecimento mais especializado. Isso significa que sem alguma experiência profissional é quase impossível para um leigo determinar uma categoria de inseto no nível de espécie. Como resultado, o reconhecimento de espécies de insetos usando métodos de visão por computador é cada vez mais exigida em aplicação [4].

III. COLETA DE LARVAS

Para coleta de larvas necrofágicas foram adotadas duas metodologias, uma principal e outra complementar. A principal foi feita com carcaça de suíno e a complementar com armadilha contendo isca. O método principal foi realizado na fazenda escola da Universidade Católica Dom Bosco – UCDB. O suíno, com cerca de 10 Kg foi adquirido morto de estabelecimento comercial de Campo Grande, ele foi colocado sobre o solo de uma região de transição entre mata ciliar e pasto sob as coordenadas UTM x: 749580 e y: 7743829 (21K) a 650m de altitude. Neste local a carcaça do suíno foi exposta a condições ambientais naturais, e atraiu indivíduos dípteros que depositaram seus ovos na carcaça. Uma jaula de metal revestida por tela foi colocada para a proteção da carcaça contra mamíferos, mas possibilitando a entrada de insetos (Fig.1).

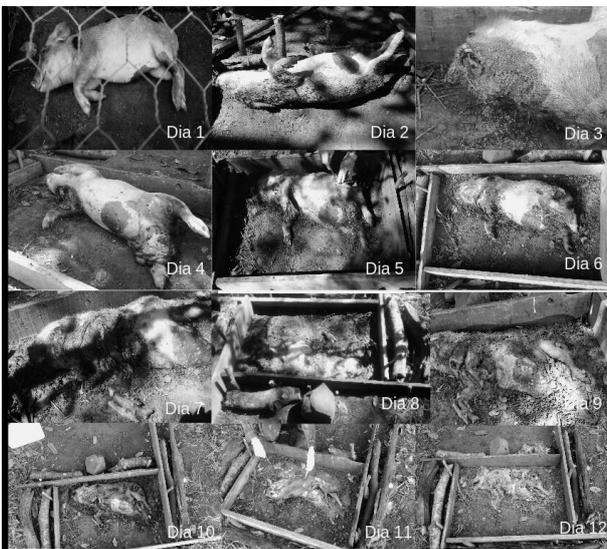


Fig.1: Aparato de metal (principal) onde o suíno foi alocado para decomposição.

O método complementar consistiu na coleta de larvas de moscas utilizando-se armadilhas (Fig.2). Foram confeccionadas duas armadilhas utilizando seis garrafas PET (três para cada armadilha), tinta preta do tipo bisnaga (não tóxica), fita adesiva, barbante e carne crua. A armadilha foi montada da seguinte maneira: foram cortadas três metades superiores das garrafas, sendo que duas dessas metades foram unidas por fita adesiva e em seguida feito um furo do meio de junção entre as duas metades, a extremidade superior teve sua ponta inserida no furo. Em seguida uma extremidade inferior foi unida por fita adesiva, a extremidade superior que teve sua ponta inserida do furo essas duas extremidades foram pintadas de preto com a tinta do tipo bisnaga e em seguida feito um furo do meio das duas extremidades.



Fig.2 : Armadilha (complementar) utilizada na captura dos dípteros.

IV. BANCO DE IMAGEM

Os indivíduos coletados foram primeiramente separados em morfo espécies e posteriormente identificados quanto ao gênero com o auxílio de um microscópio óptico e uso da chave de identificação da referência [10], sendo que os adultos destas larvas foram levados ao especialista para confirmação da identificação. Após a identificação de cada larva, ela foi colocada em uma lâmina e levada ao microscópio óptico CARL ZEISS Axio Scope A1 na lente objetiva de 5x, variando a luminosidade e o foco para uma melhor visualização das características. Para a captura de imagem das larvas, elas foram fotografadas em três planos de visão sem o processo de clarificação, pois suas estruturas externas eram muito visíveis, sendo eles, dorsal, lateral e ventral, para ser possível capturar um maior número de características por região de larva (Fig.3). Devido à variação numérica entre as espécies identificadas foi definido o número de 30 indivíduos a serem fotografados. A maioria das fotos foram capturada de larvas que estavam no segundo ou terceiro instar devido à dificuldade da identificação correta no primeiro instar.

V. APLICAÇÃO COMPUTACIONAL

Para extração dos atributos nas imagens das larvas, foi utilizado o software python-extrai-atributos, desenvolvido pelo grupo INOVISAO da UCDB. Este programa é composto por vários extratores de atributos dentre os mais importantes estão a matriz co-ocorrência que extrai atributos de textura baseados em matrizes de coocorrência, LBP (Local Binary Patterns) que extrai atributos de textura e o HOG (Histogramas de Gradientes Orientados) que extrai atributos de gradiente. O software python-extrai-atributos gera arquivos de mineração de dados (arff) para cada diretório do último subnível de diretório que compõe o banco de dados, ou seja, para os diretórios cranial e caudal de cada espécie respectivas a cada instar. Desta forma, foi possível testar os arquivos arff específicos de cada característica individualmente. Também

forma feitos testes concatenando todas as imagens de um mesmo instar concatenadas de cada instar, tornando possível observar se há diferença de classificação entre a quantidade e diversidade de informações extraídas.

Para testes com aprendizagem automática foi utilizado o Weka [2] com os algoritmos Random Tree, Random Forest, Adaboost (versão M1) e SMO. As métricas utilizadas para mensurar o desempenho dos classificadores foi a porcentagem de acerto, área sob a curva ROC e medida F. Após o término das análises foram realizados o teste estatístico denominado ANOVA com o pós teste de Tukey para confirmação dos resultados utilizando o programa R Studio.

VI. RESULTADOS E DISCUSSÕES

As larvas utilizadas para o experimento foram: *Atherigona orientalis*, *Chrysomya albiceps*, *Chrysomya megacephala*, Família Sarcophagidae (Fig.4). Com as larvas necrofágicas coletadas tanto no método principal quanto no complementar foi possível construir um banco de imagens contendo 480 imagens com quatro espécies de larvas necrofágicas que provavelmente são as mais abundantes nas duas região de Campo Grande. Na Fig. 5 é demonstrado o resultado obtido com o classificador SMO utilizando a métrica porcentagem de acerto. É possível visualizar que a espécies melhores identificadas foram *Chrysomya albiceps* no segundo instar e família Sarcophagidae no terceiro instar.

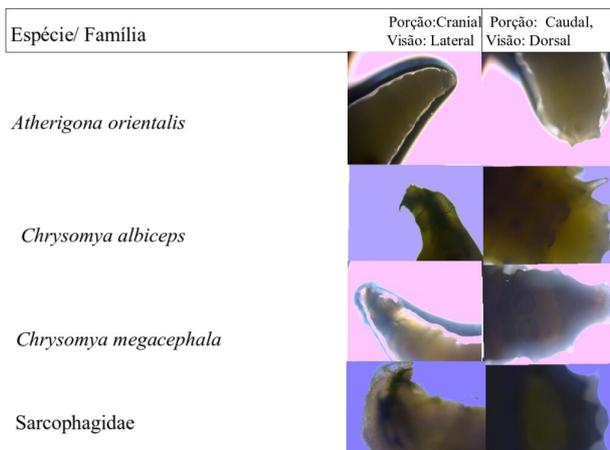


Fig.4: Visão dorsal da porção caudal (a direita) e porção cranial (a esquerda) das larvas que foram utilizadas.

Na métrica Área sob a curva ROC os algoritmos SMO e Random Forest obtiveram resultados acima de 90% com poucas variações. Os algoritmos Adaboost e Random Tree obtiveram resultados entre 30 e 80% com a mesma métrica. Na métrica de medida F os valores variaram consideravelmente, o algoritmo SMO variou entre 20 e 80 %, o classificador Random Forest obteve desempenho entre 65 a 80%, o Adaboost de 40 a 80% e o Random Tree de 27 a 98%. Demonstrando resultados semelhantes ao trabalho realizado por [3] onde obtiveram uma variação de acerto entre 60 e 95% utilizando uma metodologia diferente, o que indica a necessidade

de um aprofundamento no estudo e aplicação de técnicas de visão computacional.

	SMO																	
	Instar 1				Instar 2				Instar 3									
	D. Cra.	D. Cau.	V. Cra.	V. Cau.	L. Cra.	L. Cau.	D. Cra.	D. Cau.	V. Cra.	V. Cau.	L. Cra.	L. Cau.						
A	0	0	0	0	0	0	50	100	25	40	0	33	0	33	0	0	33	0
CM	0	0	0	0	0	0	20	40	50	75	50	40	100	25	25	33	0	0
CA	0	0	0	0	0	0	25	33	100	20	100	16	67	0	0	29	0	0
S	0	0	0	0	0	0	40	50	50	100	0	33	50	0	100	75	40	50

Fig.5: Desempenho do classificador SMO na classificação das imagens larvais. Dorsal Cranial (D. Cra.), Dorsal Caudal (D. Cau), Ventral Cranial (V. Cra.), Ventral Caudal (V. Cau.), Lateral Cranial (L. Cra), Lateral Caudal (L. Cau), *Atherigona orientalis* (A), *Chrysomya megacephala* (C.M), *Chrysomya albiceps* (C.A), Sarcophagidae (S).

Na métrica Área sob a curva ROC os algoritmos SMO e Random Forest obtiveram o resultado acima de 90% com poucas variações. E os algoritmos Adaboost e Random Tree obtiveram resultados entre 30 e 80% com a mesma métrica. Na métrica de medida F os valores variaram consideravelmente, o algoritmo SMO variou entre 20 e 80 %, o classificador Random Forest obteve desempenho entre 65 a 80%, o Adaboost de 40 a 80% e o Random Tree de 27 a 98%. Demonstrando resultados semelhantes ao trabalho realizado por [3] onde obtiveram uma variação de acerto entre 60 e 95% utilizando uma metodologia diferente, o que indica a necessidade de um aprofundamento no estudo e aplicação de técnicas de visão computacional.

Na Fig. 6 é apresentado o resultado do desempenho dos quatro algoritmos testados, Adaboost, Random Forest, Random Tree e SMO em relação a classificação dos instares de cada espécie de larva.

O teste estatístico ANOVA apresentou o valor-p igual a 0.279 para arffs concatenados e 0.100 para arffs não concatenados, demonstrando que não houve diferença estatística entre os classificadores (Fig.7 e Fig.8).

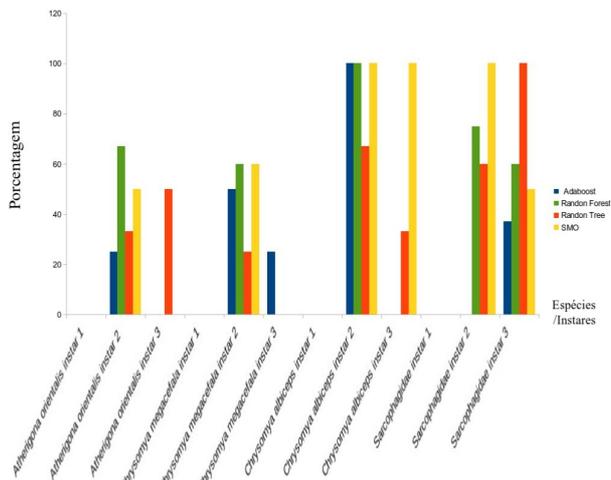


Fig.6 Porcentagem de acerto dos algoritmos classificadores em relação ao instar e espécie das larvas necrofágicas. No eixo X consta a espécie das larvas (exceto Sarcophagidae que é família) seguido do instar da larvas. No eixo Y consta a porcentagem de acerto dos algoritmos testados. A direita consta o nome dos algoritmos testados juntamente com a cor que cada um é representado no gráfico.

Foi possível perceber alguns fatores que influenciaram a expressão dos resultados, o primeiro é a quantidade de imagens, notou-se que nos testes, todos os exemplos referentes ao instar 1 apresentaram resultados iguais ou próximos a zero, como os diretórios referentes aos instares 2 e 3 tem exatamente a mesma quantidade de imagens e somente os diretórios do instar 1 tem um número extremamente menor de imagens, podemos dizer que um dos motivos que explica o mau desempenho em classificar imagens referentes ao instar 1 é a melhor quantidade de imagens.

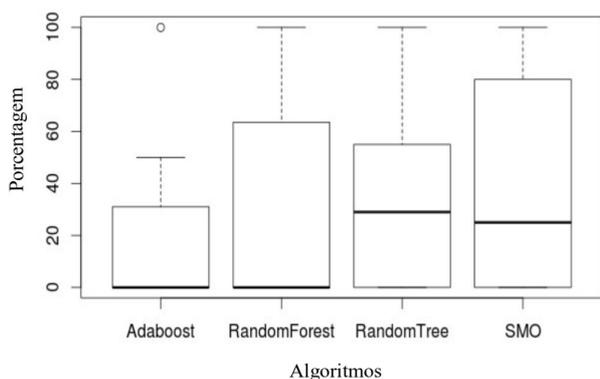


Fig.7 Gráfico tipo Boxplot demonstrando o desempenho estatístico dos classificadores utilizando arff's sem concatenação.

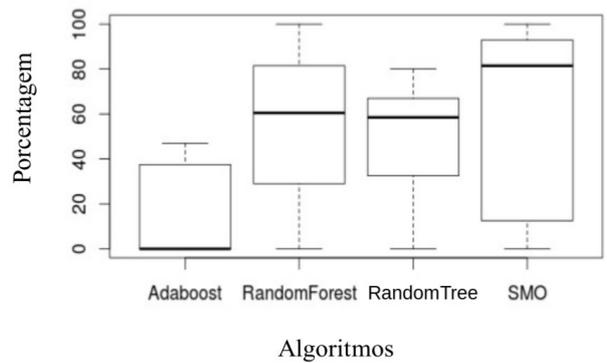


Fig.8 Gráfico do tipo Boxplot demonstrando a diferença estatística de desempenho dos classificadores que foram concatenados.

O segundo fator que influenciou os resultados de modo geral foi a quantidade e variedade de características. As larvas em instar 1 independentemente da espécie apresentam pouquíssimas características morfológicas que as diferenciam o que torna ainda mais difícil sua classificação, o que justifica a menor quantidade de imagens no banco. Em contra partida as larvas da espécie *Chrysomya albiceps* e da família Sarcophagidae a partir do instar 2 apresentam várias características morfológicas que as diferenciam, a ponto de todos os classificador conseguirem captar a maioria das características.

O terceiro e talvez principal fator é o fato do algoritmo SMO ter conseguido apresentar resultados mais satisfatórios em classificar a larvas considerando as métricas utilizadas, tanto na porcentagem de acerto quanto ter a menor variação entre o pior resultado e o melhor resultado que os demais algoritmos testados. A explicação para isso pode ser a seguinte: Todos os algoritmos com exceção do SMO trabalham utilizando o método de árvore de decisão ou uma combinação de vários algoritmos que utilizam esse método [6].

VII. Conclusão

Foi construído um banco de imagens de acesso livre que poderá ser utilizado para testes com outras técnicas computacionais. Com a técnica utilizada foi possível automatizar a identificação das espécies de larvas necrofágicas em seus respectivos instares de desenvolvimento. Os classificadores Random Forest e SMO tiveram um desempenho parecido em todas as métricas utilizadas. Para trabalhos futuros espera-se um aprofundamento nos testes utilizando os classificadores com o melhor desempenho, coleta para a aquisição e classificação de larvas em L1 para

enriquecer o banco de imagens e a construção da interface do usuário.

Agradecimentos

Ao Centro Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) pelas bolsas concedidas. Ao Dr. Wedney Rodolpho de Oliveira pelos ensinamentos da coleta das larvas necrofágicas. À equipe SINOVA que auxiliou nos procedimentos laboratoriais .

REFERÊNCIAS

- [1] CATTS & Goff, *Forensic Entomology in Criminal Investigations*, ed. Annu. Rev. Entomol., Palo Alto CA, 1992.
- [2] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explorations*. 2009.
- [3] LARIOS, E, et al, *Automated Insect Identification through Concatenated Histograms of Local Appearance Features*, 2014.
- [4] LU, A., HOU, X., CHEN, X., LIU, C., *Insect Species Recognition Using Sparse Representation*, BMVC Conference, Aberystwyth, 2010.
- [5] NEETHIRAJAN S, KARUNAKARAN C, JAYAS DS, WHITE ND, *Detection techniques for stored-product insects in grain*. *Food Cont.* 18, 157–162, 2007.
- [6] PLATT, J. C. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. Microsoft Research, Technical Report MSR-TR-98-14, USA, 1998.
- [7] PUJOL-LUZ, J.R.; ARANTES, L.C.; CONSTANTINO, R. *Cem anos da entomologia Forense no Brasil (1908 – 2008)*. ed. *Revista Brasileira de Entomologia*. Volume 52 (4): 485-492., Curitiba PR, 2008.
- [8] RAMOS A. L. & SANTOS C.N., *Combinando Algoritmos de Classificação para Detecção de Intrusão em Redes de Computadores*, Fortaleza CE, 2015.
- [9] SOLIS-SÁNCHEZ, L. O., *Machine vision algorithm for whiteflies*, ed. *Journal of applied entomology* Vol.133(7), pp.546-552, Querétaro MEX, 2009.
- [10] SHEWELL, G. E., In: MC ALPINE, J. F.; PETERSON, B. V.; SHEWELL, G. E.; TESKEY, H. J.; VOCKEROTH, J. R.; WOOD, D. M. ed. *Manual of Nearctic Diptera*, 1987.
- [11] SMITH G.V., *A Manual of Forensic Entomology*, London 1986.