

COMPUTAÇÃO VISUAL NA IDENTIFICAÇÃO DE SEMENTES UTILIZANDO EXTRAÇÃO DE CARACTERÍSTICAS DE FORMA

KIST, Deives M.¹; OLIVEIRA, Lucas F.²; Cavalheiro, Gerson G. H.³;

¹ Universidade Federal de Pelotas, Curso de Bacharelado em Ciência da Computação;

² Universidade Federal do Paraná, Programa de Pós-Graduação em Bioinformática;

³ Universidade Federal de Pelotas, Curso de Bacharelado em Ciência da Computação.

deivesmesquitakist@gmail.com

1 INTRODUÇÃO

O Brasil é um dos principais exportadores de sementes e com papel de destaque no cenário mundial como consumidor e produtor deste produto (OHLSON, 2010). O valor agregado, no entanto, pode ser reduzido devido à detecção de sementes consideradas invasoras as quais crescem de maneira vigorosa em competição por água e nutrientes. Logo, a identificação precoce das sementes invasoras é o primeiro passo para um sistema de controle de qualidade da produção nacional. O atual processo de identificação de sementes feito manualmente por especialistas necessita de laboratórios e sementários com amostragem de espécies e que a descrição de espécies invasoras é rara. No entanto, as características morfológicas das sementes não se alteram muito com as mudanças ambientais e podem ser usadas tão seguramente quanto às características da planta para chegar até a identificação da espécie (MUSIL, 1963). Este fato pode ser explorado em um sistema computacional de apoio à identificação de espécies baseado em características de imagens de sementes.

Nesse trabalho introduzimos uma técnica de identificação de sementes por imagens baseada em estratégias CBIR, do acrônimo em inglês para Busca de Imagens Baseada em Conteúdo (ARAÚJO,2010). Tal técnica foi utilizada para construir uma ferramenta computacional que auxilia na etapa de identificação de espécies de semente. Esta ferramenta opera sobre um banco de dados composto por imagens de sementes de espécies invasoras e, a partir de uma imagem apresentada pelo usuário, identifica na base de dados aquelas espécies que possuem características visuais que mais se aproximam da espécie fornecida como entrada.

2 METODOLOGIA (MATERIAL E MÉTODOS)

O presente trabalho apresenta uma ferramenta para realizar o reconhecimento de sementes por imagens concebidas nos moldes de um sistema de busca de imagens por conteúdo (ARAÚJO,2010). Esse sistema utiliza uma base de dados composta por imagens e de um extrator de características visuais. A identificação da existência de uma imagem na base de dados semelhante a uma determinada imagem fornecida consiste em comparar as características visuais da imagem apresentada com as características das imagens armazenadas na base dados. Diferentes tipos de características visuais podem ser utilizados como cor, forma e textura.

Na ferramenta desenvolvida a base de dados é composta por imagens de espécies de sementes e saída fornecida é um conjunto de espécies similares à imagem de entrada.

A base de dados utilizada contém 68 espécies de sementes classificadas como invasoras. A forma de aquisição das imagens seguiu rigorosamente o mesmo procedimento, uma vez que devem ser garantidas as mesmas condições de

luminosidade para garantir uniformidade na etapa de extração das características das imagens. Este procedimento é descrito na sequência.

As fotografias das sementes foram obtidas por uma câmera fotográfica digital de 6 megapixels, focando grupos de 2 a 8 exemplares da mesma espécie distribuídas sobre um vidro limpo colocado à 15 cm de distância de um fundo azul. Para cada uma das 68 espécies foram adquiridas pelo menos 8 fotografias. Observou-se que em nenhuma situação ocorreram exemplares sobrepostos ou se tocando. A fonte de luz utilizada foi proveniente de uma lâmpada fluorescente redonda, colocada sob o vidro, para garantir luminosidade em todos os lados da semente, sem a formação de sombra e brilho causados pelo reflexo do flash da câmera no vidro.

Em uma etapa anterior ao extrator de características cada exemplar na imagem foi separado em um processo chamado de segmentação (OLIVEIRA, 2010), de forma a representar uma das espécies armazenadas na base de dados. Assim, a base de dados possui armazenada pelo menos 16 sementes de cada espécie. Outro aspecto importante é que se decidiu por utilizar máquina fotográfica digital convencional ao invés de equipamentos mais robustos, como lupa eletrônica ou microscópio.

Esta decisão foi baseada no fato de que o usuário deve obter suas imagens em um ambiente com as mesmas características daquele utilizado para obtenção das imagens da base de dados, de forma a ser possível a leitura das mesmas informações visuais das imagens. A utilização de lupa ou microscópio encareceria o sistema desnecessariamente.

A técnica proposta utiliza características de forma das sementes contidas nas imagens visando uma classificação correta.

A característica de forma é uma das mais importantes, pois ela é bastante discriminante em relação às espécies. Nesse contexto, propõe-se uma técnica de extração de característica por forma. Essa técnica propõe uma adaptação dos atributos de textura de Haralick (HARALICK, 1973) para trabalharem com distância de pontos, determinando um ponto central em cada semente e medindo as distâncias das bordas em relação ao centro. Os atributos mais utilizados são a energia, entropia, contraste, homogeneidade, dissimilaridade e medidas de correlação.

O processo de extração de características de forma é considerado o mais difícil, requerendo a implementação de mecanismos de segmentação e cuidados relacionados à qualidade da obtenção das imagens, considerando tanto a resolução da imagem, quanto a calibragem dos equipamentos envolvidos na captura.

3 RESULTADOS E DISCUSSÃO

A base de dados possui imagens de 68 espécies pertencentes a 12 famílias. Do conjunto de imagens disponíveis, foi retirada uma imagem de cada família para experimentação. Para as imagens restantes, a aplicação da técnica proposta iniciou por extrair as características de todas as imagens da base de dados e registrar a média dos valores obtidos para cada família e os respectivos desvios padrões. Estes valores estão registrados na Tab.1 (duas casas decimais de precisão).

Na Tab.1 observa-se que para algumas famílias os valores para as médias obtidas para determinadas características, computados os respectivos desvios padrões, são semelhantes. No entanto, deve ser considerado que no conjunto de cinco características, a distinção se dará pela combinação das médias obtidas com os valores extraídos para uma determinada espécie a ser identificado.

O procedimento de análise do potencial de identificação do sistema utilizou as imagens retiradas da base de dados antes do cálculo das médias das famílias. Para estas imagens, foram extraídas as características de forma, tendo sido obtidos os valores apresentados na Tab.2. Nesta tabela as espécies são identificadas pelas posições relativas das linhas correspondem às famílias apresentadas na Tab.1. As colunas para característica apresentam como informação adicional quantos desvios padrões o valor obtido se distância da média para a família, ou seja, o valor absoluto da subtração do valor medido pela média da família divide pelo desvio padrão. Desta forma, distâncias menores do que 1,0 indicam que o valor para a característica encontra-se dentro do esperado para a família.

Tabela 1 - Média e Desvios Padrão das características forma

Família	Energia		Entropia		Tendência de Agrupamento		Correlação		Contraste	
	Média	D. Padrão	Média	D. Padrão	Média	D. Padrão	Média	D. Padrão	Média	D. Padrão
1	0,03567	0,00	3,78	0,12	9023,03	3090,32	0,75	0,11	189,30993	146,91782
2	0,01045	0,02	4,92	0,46	18136,51	2892,14	0,77	0,13	1183,89059	471,78101
3	0,00961	0,00	5,11	0,17	17229,72	3875,81	0,81	0,04	740,6556	567,54095
4	0,00755	0,00	5,14	0,09	13449,02	3393,82	0,79	0,10	935,29043	545,08031
5	0,02921	0,02	4,08	0,55	10962,03	8343,28	0,74	0,12	648,29426	539,20218
6	0,01802	0,01	4,51	0,23	22352,59	1746,40	0,85	0,02	1496,41107	766,06162
7	0,00914	0,00	5,05	0,18	11447,23	4299,36	0,79	0,10	636,13963	524,76006
8	0,02196	0,01	4,19	0,34	14453,95	2588,36	0,76	0,11	502,80285	477,10548
9	0,00912	0,00	4,99	0,08	15404,04	3578,30	0,83	0,06	505,86905	291,70067
10	0,0086	0,00	5,07	0,03	10120,78	485,62	0,89	0,00	683,93225	265,58728
11	0,00896	0,00	5,06	0,16	16201,92	4438,45	0,72	0,16	1079,92209	800,4638
12	0,02157	0,00	4,22	0,14	13488,21	945,48	0,82	0,00	220,96398	69,06968

Tabela 2 - Características extraídas do grupo de experimentação

Imagem	Energia		Entropia		Tendência de Agrupamento		Correlação		Contraste	
	Valor	Distância	Valor	Distância	Valor	Distância	Valor	Distância	Valor	Distância
1	0,03	1,34	3,75	0,23	8505,50	0,17	0,96	1,91	56,94	0,90
2	0,01	0,07	4,98	0,13	15613,24	0,87	0,93	1,20	691,90	1,04
3	0,01	0,75	5,17	0,36	20060,76	0,73	0,97	3,96	1766,52	1,81
4	0,01	0,32	5,12	0,24	17996,53	1,34	0,96	1,84	1577,25	1,18
5	0,01	0,80	4,69	1,11	33387,85	2,69	0,99	2,14	1118,30	0,87
6	0,02	0,50	4,31	0,88	24440,49	1,20	0,97	5,41	573,40	1,20
7	0,01	0,34	5,05	0,03	10938,67	0,12	0,95	1,65	345,24	0,55
8	0,03	0,56	3,93	0,77	16245,81	0,69	0,96	1,83	137,72	0,77
9	0,01	0,24	5,09	1,20	17523,00	0,59	0,97	2,36	3371,53	9,82
10	0,01	1,76	5,00	2,30	11917,72	3,70	0,98	92,90	616,29	0,25
11	0,01	1,51	4,98	0,47	19534,18	0,75	0,98	1,62	611,04	0,59
12	0,02	0,41	4,21	0,06	19276,64	6,12	0,99	137,42	168,79	0,76

A Tab.2 exemplifica a comparação de algumas imagens que foram retiradas de cada espécie durante o processo de composição da base de dados. Análise dos dados obtidos para a imagem7, verifica-se que 4 características encontram-se dentro do desvio padrão observado para a família a qual pertence. Logo, a imagem7 tem altíssima similaridade com a família7.

Ao contrário da imagem10, somente uma das características está dentro do desvio padrão, portanto, apresenta uma baixa similaridade com a família10.

Assim, o sistema mostrará ao usuário o grau de similaridade de sementes contidas em imagens com as famílias armazenadas na base de dados. O usuário terá um pequeno número de famílias para identificar.

Verifica-se na Tab.2 que a coluna de Correlação tem todos os valores acima de 1,0. Logo, para este experimento a Correlação não serve para identificar se as sementes pertencem à família testada.

4 CONCLUSÃO

A identificação de imagens pela forma é uma tarefa complexa e muitos algoritmos para essa função já foram desenvolvidos. A técnica proposta de extração de atributos de forma é mais uma metodologia para executar essa tarefa. O trabalho proposto aborda uma metodologia de classificação de imagens pela forma do objeto, utilizando uma adaptação do conceito de matriz de co-ocorrência e o cálculo dos atributos de Haralick. Os testes mostraram que a técnica proposta tem potencial, visto que mesmo para sementes rotacionadas os valores dos atributos calculados tiveram uma variação mínima. Isso mostra uma tendência dos valores ficarem dentro de uma faixa mesmo com o objeto em diversas posições. Considerando o potencial dessa técnica é possível utilizá-la junto a outras metodologias de extração de características existentes na bibliografia, para desenvolver um sistema robusto de identificação de sementes. O sistema de identificação de sementes será uma ferramenta com grande utilidade para os laboratórios de análises de sementes.

5 REFERÊNCIAS

ARAÚJO, F.; Recuperação de Imagens Baseado em Conteúdo Disponível em <<http://www.fabioaraujo.com.br/recuperacao-de-imagens-baseadas-em-conteudo/>> Acesso em : 17 de abril de 2011.

HARALICK, R. M.; SHANMUGAN, K.; and DINSTEN, I.; **Textural Fetures for Image Classification**. IEEE Transactions on Systems, Man and Cybernetics, 1973. SMC-3(6): p. 610-621.

LOPES, J.C.; DIAS, P.C.; PEREIRA, M.D.; Maturação fisiológica de sementes de quaresmeira, **Pesquisa agropecuária brasileira**, Brasília, v.40, n.8, p.811-816, ago. 2005.

MUSIL, A.F. **Identification of crop and weed seeds**. Washington: Department of Agriculture, 1963. 171p. (Agriculture Handbook, 219).

OHLSON, Osvaldo de Castro; SOUZA, Camila Ribeiro de; PANOBIANCO, Maristela. Qualidade Física e Fisiológica de Sementes de Capim-Colonião e Milheto, Comercializadas no estado do Paraná. **Informativo Abrates**, vol.20, no. 1,2 p.030 - 036, 2010.

Oliveira, L. F. de; Kist, D. M.; Cavalheiro, G. G. H.; Meneghello, G. E. ; Tillmann, M. A. A..Segmentação de Imagens com Fundo Azul utilizando a multiplicação dos Canais HSV. In: **IV Workshop de Visão Computacional**, 2010, Presidente Prudente, 2010