

APLICANDO FUZZY C-MEANS PARA O RECONHECIMENTO DE SPOTS EM IMAGENS ORIUNDAS DE GÉIS DE ELETROFORESE BIDIMENSIONAL

GEANCARLO MAYDANA¹; MARLON DIAS²; MARILTON SANCHOTENE DE AGUIAR³

¹Universidade Federal de Pelotas – gsmaydana@inf.ufpel.edu.br

²Universidade Federal de Pelotas – mdsdias@inf.ufpel.edu.br

³Universidade Federal de Pelotas – marilton@inf.ufpel.edu.br

1. INTRODUÇÃO

O proteoma indica as proteínas expressas em um organismo ou tecido. O proteoma altera com o estado de desenvolvimento do tecido ou mesmo sob as condições nas quais o indivíduo se encontra. Portanto, investigar diretamente os produtos dos genes é uma forma de estudar doenças e qualquer problema biológico complexo. O termo proteoma foi introduzido em 1996 para descrever as proteínas expressas em um genoma. (ANDERSON; ANDERSON, 1996; WILKINS et al., 1996).

A proteômica pode ser vista como uma metodologia de seleção da biologia molecular, a qual tem como objetivo documentar a distribuição geral de proteínas da célula, identificar e caracterizar proteínas individuais de interesse e principalmente elucidar as suas associações e funções. (CIERO; BELLATO, 2002).

Atualmente, uma das principais técnicas usadas na proteômica é a eletroforese bidimensional. Em um experimento de eletroforese bidimensional em gel, as proteínas (ou outras moléculas) são separadas em duas dimensões, de modo que todas as proteínas/moléculas se espalhem por todo o gel.

Estas proteínas podem ser detectadas por uma variedade de reagentes de revelação. Após a coloração do gel, observa-se um perfil bidimensional de pontos (*spots*), sendo que em cada ponto há múltiplas cópias de uma proteína. Por fim, este gel é escaneado e a imagem resultante pode ser processada. (ROCHA; SILVA; SÁ, 2005).

As imagens de eletroforese bidimensional podem conter ruídos, bem como partículas de poeira e, até mesmo, rachaduras no gel, e isso pode interferir no resultado final da análise de reconhecimento dos *spots* (SAVELONAS; MYLONA; MAROULIS, 2012).

2. METODOLOGIA

Em um primeiro momento foi feito um estudo sobre técnicas de processamento de imagens *fuzzy* e *crisp* baseado principalmente em (CHI; YAN; PHAM, 1996; VARGAS; BEDREGAL; PALMEIRA, 2011; PAL, 1999). Ainda, um estudo das técnicas usuais para análise de géis de eletroforese bidimensional, baseando-se em trabalhos como os de (RABILLOUD et al., 2010; DASZYKOWSKI et al., 2009). A partir deste estudo, foi proposto um modelo a fim de fazer a clusterização das imagens com o intuito de encontrar *spots*, utilizando o algoritmo de clusterização *Fuzzy C-means*. A análise é feita com base na imagem provida ao final do exame, onde as moléculas já sofreram os efeitos dos campos elétricos e pararam de se mover. Partindo da imagem, os únicos dados presente são os tons de cinza que compõem a imagem. Baseando-se nos

princípios da clusterização, acredita-se que os *spots* apresentam tons de cinza similares e, assim, pertençam ao mesmo cluster.

A Figura 1 (a) é a imagem original. Ela representa uma parte da imagem maior. Cada cor nas imagens representam um cluster. Em (HOOGLAND et al., 2004) é apresentado uma base de imagens oriundas de eletroforese bidimensional. As imagens presentes nessa base de dados já foram testadas e apresentam *spots* confirmados, os quais foram usados para medir a precisão do modelo. Logo, há uma lista informando os pontos na imagem que representam um *spot*. Os pontos apresentados nessa lista são *spots* confirmados, entretanto, ainda há outros pontos na imagem que não foram confirmados *spots*.

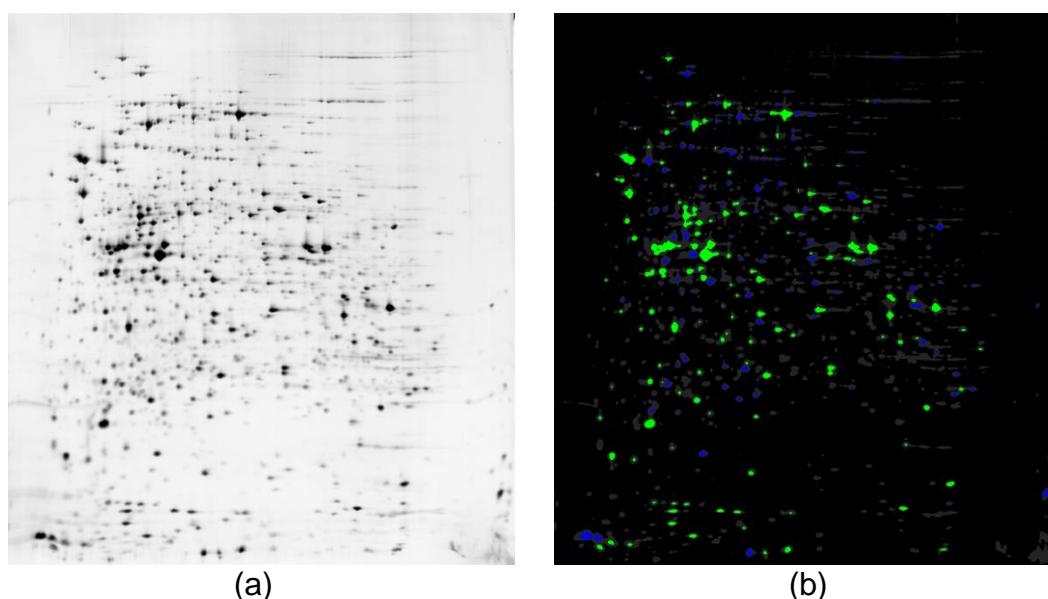


Figura 1. Imagens de eletroforese utilizadas para clusterização. (a) Imagem original – ECOLI. (b) Imagem após a clusterização.

3. RESULTADOS E DISCUSSÃO

Diversos testes foram realizados com diferentes configurações do modelo. O resultado mais promissor foi utilizando três clusters. Com base nessa configuração, observa-se os clusters representando o fundo da imagem, redondezas ou rastros de um *spot* e, por fim, os *spots*. A Figura 1 (b) mostra um teste realizado com uma das imagens da base utilizando essa configuração. A Figura 1 (a) representa a imagem original, chamada de ECOLI. Os três clusters, mostrados na Figura 1 (b), são representados pelas cores preto (cluster 1); cinza (cluster 2); e azul e verde (cluster 3). O cluster 3 representa as áreas na imagem consideradas *spots*.

O modelo informa o grau de pertinência de cada ponto a um dos clusters. A função máximo é aplicada como método de defuzzificação, ou seja, o cluster do ponto é designado com base no maior grau de pertinência entre clusters. Assim, tem-se a informação do cluster de cada ponto. Para poder contar um *spot* é necessário saber a área que representa uma *spot*, o conjunto de pontos próximo pertencentes ao cluster que representa os *spots*.

Com o objetivo de formar e contar os *spots*, um algoritmo de preenchimento é usado. Parte-se de um ponto que é *spot* e dele verifica-se a

vizinhança. Os vizinhos que também são *spots* serão os próximos a serem verificados. Esse processo se repete até não haver vizinhos classificados como *spot*. Com isso, todos esses pontos avaliados são considerados um *spot*. Esse processo é aplicado a toda imagem.

O cluster 3 possui duas cores, ambas representam *spots* encontrados pelo modelo. A diferença é que as áreas verdes são *spots* confirmados por um especialista em (HOOGLAND et al., 2004). Os azuis não foram confirmados, mas isso não diz diretamente que eles não possam ser *spots*. Na verdade, são áreas que poderiam ser consideradas *spots*, indicando que elas precisam de uma análise mais detalhada. De acordo com o protocolo deste exame, pode ser necessário recortar uma porção do gel para uma análise mais detalhada e, assim, esta região em que o modelo indica ter *spots* não comprovados poderia ser uma sugestão de recorte. Os pontos vermelhos, são *spots* confirmados em (HOOGLAND et al., 2004) que não foram encontrados pelo modelo. No total, a lista de *spots* confirmados contém 206 *spots*. O modelo encontrou 279 *spots*, onde 151 estão na lista de confirmados, e 55 estão faltando, resultando em uma precisão de 73%.

4. CONCLUSÕES

Imagens de eletroforese bidimensional são difíceis de analisar. Há o problema de possíveis ruídos na imagem, o que pode dificultar na análise. Porém, além disso, os *spots* podem apresentar características bem diferentes. Por exemplo, dos *spots* confirmados na lista, mas não encontrados pelo modelo, apresentam tons de cinza muito distintos. A maioria dos confirmados possui um tom mais forte, entretanto, alguns dos confirmados possuem tons de cinza bem baixos. Isso é um grande desafio a ser superado pela clusterização.

Como trabalhos futuros, é previsto a utilização de algumas técnicas de pré-processamento. É possível utilizar algumas técnicas como remoção do fundo, normalização, equalização e segmentação da imagem. As imagens utilizadas são relativamente grandes, contendo muita informação. A primeira técnica tem o intuito de reduzir os dados a serem processados, o que pode acelerar o processo. A segunda e terceira, visam mudar os dados apresentados, podendo influenciar diretamente na maneira com que a função *fuzzy* os clusteriza. E, por fim, a última técnica, a segmentação da imagem, tem como objetivo dividir a imagem, resultando na clusterização de imagens menores. A imagem original apresenta regiões com intensidades diferentes, com isso pretende-se achar aqueles *spots* com tons de cinza mais discrepantes.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, N. G.; ANDERSON, N. L. "Twenty years of two-dimensional electrophoresis: past, present and future," **Electrophoresis**, vol. 17, pp. 443–453, 1996.

BEZDEK, J. C.; PAL, S. K., Eds., **Fuzzy Models for Pattern Recognition**, ser. A selected reprint volume. IEEE Press, 1992, methods That Search for Structures in Data.

BEZDEK, J. C. **Pattern Recognition with Fuzzy Objective Function Algorithms**. Norwell, MA, USA: Kluwer Academic Publishers, 1981.

BLOCH, I.; MAÎTRE, H. "Fuzzy mathematical morphology," Ann. **Math. Artif. Intell**, vol. 10, pp. 55–84, 1993.

CHI, Z.; YAN, H.; PHAM, T. **Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition**, ser. Advances in fuzzy systems applications and theory. World Scientific, 1996.

CIERO, L.; BELLATO, C. "Proteoma: avanços recentes em técnicas de eletroforese bidimensional e espectrometria de massa," **Biotecnologia Ciência e Desenvolvimento**, vol. 29, pp. 158–164, 2002.

HOOGLAND, C.; MOSTAGUIR, K.; SANCHEZ, J.-C.; D. HOCHSTRASSER, F.; APPEL, R. D. "**Swiss-2dpage, ten years later**," Proteomics, vol. 4, no. 8, pp. 2352–2356, 2004.

ROCHA, T. L.; SILVA, M. C.; SÁ, M. F. "Eletroforese bidimensional e análise de proteomas," **Comunicado Técnico da Embrapa**, vol. 136, 2005.

SAVELONAS, M. A.; MYLONA, E. A.; MAROULIS, D. "Unsupervised 2d gel electrophoresis image segmentation based on active contours," **Pattern Recogn.**, vol. 45, no. 2, pp. 720–731, 2012.

WILKINS, M. R.; PASQUALI, C.; APPEL, R. D.; OU, K.; GOLAZ, O.; SANCHEZ, J. C.; JAN, J. X.; GOOLEY, A. A.; HUGHES, E.; HUMOHERY-SMITH, I.; WILLIAMS, K. L.; HOCHSTRASSER, D. F. "From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis," **Nature Biotechnology**, vol. 14, pp. 61–65, 1996.