

# DIAGNOSTICO AUTOMATICO DE CÉLULAS DO COLO DE ÚTERO

Franco, Ramon<sup>1</sup>; Carvalho, Marco<sup>2</sup>

<sup>1</sup> Doutorado em Tecnologia, Unicamp, Limeira Sao Paulo ramon@ft.unicamp.br

<sup>2</sup> Professor Doutor, Unicamp, Limeira São Paulo, magic@ft.unicamp.br

**RESUMO** - Las tecnologías computacionales han permitido la disminución de los índices de muertes por Cáncer de Cuello de Utero en los países desarrollados, pero aun así, sigue siendo el segundo cáncer con más muertes en el mundo. Este artículo tiene como objetivo presentar algunas de las tecnologías investigadas para minimizar este tipo de cáncer, junto con la discusión de por qué Brasil sigue siendo uno de los países con mayor índice de muertes por esta enfermedad a pesar de la disposición de estas tecnologías. Se propone realizar un trabajo interdisciplinar que involucre el procesamiento digital de imágenes, en imágenes coleccionadas por el método de citología esfoliativa convencional. .

**Palavras-chave:** Tecnologías; Cáncer colo de Útero; Procesamiento imágenes.

## INTRODUÇÃO

O câncer de colo do útero (CCU) atinge mulheres de todas as idades, mesmo sendo raro em mulheres jovens, esse tipo de câncer pode ter início ainda na adolescência. No Brasil, para o ano de 2016 são esperados 16.340 casos novos de câncer de colo do útero, isto representa um risco médio de 15,85 casos a cada 100 mil mulheres (MINISTÉRIO DA SAÚDE, 2016)?. Em 2012, esse tipo de neoplasia foi o responsável pelo óbito de 265 mil mulheres no mundo. Vale ressaltar que 87% desses óbitos ocorreram em países em desenvolvimento. Tanto a incidência quanto a mortalidade por câncer de colo do útero podem ser reduzidas através de programas organizados de rastreamento. Uma expressiva redução na mortalidade causada por essa doença foi alcançada nos países desenvolvidos após a implantação de programas de rastreamento de base populacional, a partir de 1950 e 1960 (GLOBOCAN, 2012)?.

O exame Papanicolau usando uma lâmina de vidro e leitura convencional em laboratórios clínicos é o procedimento mais praticado pelos patologistas no rastreamento e detecção do câncer de colo de útero no Brasil. Isto ocorre devido ao baixo custo de implementação do referido exame. Mas, esta não é a maneira mais eficiente de se fazer o Papanicolau. Para mulheres que moram em populações afastadas e que não têm laboratórios clínicos próximos, o Papanicolau convencional é uma tarefa muito complicada pois é necessário o deslocamento das mulheres para a coleta e para a entrega dos resultados que demoram em muitos casos até 40 dias no sistema público de saúde brasileira SUS (LORENTE, 2013)?. Uma das primeiras tecnologias estudadas para auxiliar o diagnóstico do câncer do colo de útero de maneira automática foram pesquisadas pelo grupo de pesquisa do centro de Análise de Imagens da Universidade de Uppsala, na Suécia nos anos 80. Este foi um dos primeiros centros de pesquisa voltados para o estudo da segmentação de imagens de células do colo do útero no mundo. As pesquisas desenvolvidas pelo referido grupo inspiraram vários pesquisadores a fazer estudos de automatização de imagens. Durante os últimos 50 anos os estudos de segmentação e classificação de imagens do colo do útero, tiveram várias evoluções. Nos anos 1980, a maior parte das pesquisas da área se concentravam em encontrar maneiras de segmentar imagens usando o histograma na escala de cinzas. Assim, apareceram tecnologias como Cybest, Cytoview e Papne. E foram geradas as primeiras pautas na automatização do diagnóstico do CCU. Na década de 1990 surgiram as tecnologias de coleta em meio Líquido LBC1 e as tecnologias na apresentação de mostra ThinPrep para a realização do exame do Papanicolau. Estas tecnologias abriram novos precedentes para pesquisa e prevenção do CCU. Já no final do século XX, as técnicas de redes neurais e Fuzzy-based foram amplamente pesquisadas (CENCI; NAGAR; VECCHIONE, 2000)?, aproveitando a melhoria na qualidade das imagens fornecidas pela tecnologia ThinPrep. Esta última, é uma tecnologia desenvolvida para a separação mecânica das células por centrifugação, este processo permite as células fiquem mais visíveis, tornando mais fácil sua identificação. O ThinPrep começou a ser investigado na década de 1990, nos Estados Unidos. Posteriormente, essa tecnologia se disseminou em diversos países (LOZANO; MIGUEL; COUSILLAS, 2011)? ?

Técnicas como otsu, k-mens (LIU; YU, 2009)?, watershed (BEUCHER, 1992)?, Gray level cooccurrence matrix (KINOSHITA et al., 2007)?, Local Binary Pattern histogram LBP (OJALA; PIETIKAINEN; HARWOOD, 1994)?, Edge Orientation Histogram (JAYASINGH; STEPHEN, 2014)???, Maximal Gray-level-gradient-difference (MGLGD) method (YUNG -FU, 2014)? são técnicas amplamente utilizadas na segmentação de células do colo de útero. No entanto, muitos

destes métodos são usados com imagens captadas por dispositivos tecnológicos modernos, incluindo câmaras digitais de alta qualidade e o método ThinPrep (ANSCHAU; GINECOLOGIA; LUCAS, 2006)?. A tecnologia ThinPrep permite uma análise mais precisa e eficaz das células. O uso da tecnologia mencionada acarretou numa mudança na forma como o Papanicolau era coletado. E então, o método de esfoliação comum que usa uma única lâmina de vidro foi substituído pelo o método LBC. As referidas tecnologias possibilitaram avanços na segmentação e classificação de imagens de células do colo do útero, nos países em que foram implementadas (BUDGE et al., 2005)?. Em países em desenvolvimento como o Brasil, os órgãos de saúde pública costumam usar a tecnologia de coleta em lâmina de vidro. O principal motivo do uso deste método antigo são seus baixos custos e sua fácil implementação. Mesmo com avanços, como a vacina ofertada a mulheres jovens (com idade inferior a 15 anos) contra alguns tipos de câncer do colo do útero, as pesquisas na área de processamento de imagens do colo de útero não diminuiram e os estudos neste campo do conhecimento tem até aumentado nos últimos anos. (FRANCO, 2015)? Um motivo para o crescimento deste campo de estudos é a evolução da capacidade de processamento computacional; fator que facilita o processamento de imagens. A maior parte dos estudos recentes usam imagens de alta qualidade fornecidas pelo o método ThrinPet. As imagens de baixa qualidade coletadas pelo o método convencional (com a lâmina de vidro) não são mais usadas para pesquisa (BERGMEIR; SILVENTE; BENÍTEZ, 2011), (USHIZIMA et al., 2013)? (YOSHIO FUKUDA; MARIA Q. FENG; MASANOBU SHINOZUKA, 2011)???. Nesse contexto, vale endossar que devido aos seus custos elevados, a implementação de técnicas de última geração como o ThrinPet no sistema público de saúde não são viáveis para os países em desenvolvimento. Dessa maneira, faz-se necessário que pesquisas sobre o aproveitamento do método convencional ainda sejam realizadas. Em razão disso, o grupo IMAGELab2 desenvolve uma pesquisa que visa segmentar e classificar imagens de baixa qualidade obtidas de lâminas de vidro por meio de técnicas de processamento de imagens.

## IMAGELab

O grupo de pesquisa IMAGELab realizou um contato com o centro de pesquisa de imagens da Universidade de Uppsala na Suécia, para compartilhar conhecimentos e avanços da área nestes últimos anos. O centre for image Analysis da Universidade de Uppsala (Sweden) é liderado pelo o professor dr. Ewert Bengtson, pela professora dra. Carolina Wahlby e pela professora dra. Ingela Nystrom. Eles tem publicado artigos importantes na área de análises de células do colo de útero e, assim, conseguiram notoriedade mundial. As pesquisas do referido centro de análise de imagens são realizadas desde o ano de 1978, eles são os pioneiros em pesquisas sobre Sistemas de inspeção de células de baixa resolução de amostras ginecológicas usando o analises do histograma nas tonalidades de cinza (Y. NOGUCHI, B. STENKVIST, J. HOLMQUIST, 1975)?. Mais recentemente, o Centre Image Analysis se destacou no desenvolvimento de pesquisas sobre o uso da técnica PAPSINTH, técnica que simula o campo brilhante em imgenes de Esfregaço de células de colo do útero (MALM; BRUN; BENGTTSSON, 2010)?. Com uma experiência de 40 anos, o grupo de pesquisa de Image Analysis, é o lugar mais adequados para aprender novas técnicas de segmentação e classificação de célula uterinas. Além disso, uma aproximação com o grupo Image Analysis ajudaria no fortalecimento da atividade de pesquisa realizada pelo grupo IMAGELab da FT Unicamp. Atualmente, o grupo IMAGELab da FT Unicamp está desenvolvendo uma pesquisa para classificar e segmentar células obtidas através esfrego convencional de células do colo do útero em lâmina de vidro. A pesquisa já conseguiu alguns resultados como o desenvolvimento de uma metodologia para a separação do fondo do porta objetos de vidro e a detenção do citoplasma sobrepostos usando, regressão não linear e os algoritmos Watersheet e k-means . Os resultado preliminares foram apresentado no Campus Party de Inovação e Tecnologia no 2016 em São Paulo.

## FRONTEIRA TECNOLOGIACA

A fronteira tecnológica no campo da automatização do diagnóstico do câncer de colo do útero está na segmentação e classificação de amostras com múltiplas e variadas células. A maior parte das pesquisas desenvolvidas nesta área trabalham com a classificação de células isoladas . Contudo, para fazer um diagnóstico adequado os patologistas precisam ver a amostra coletada como um todo (ver figura 1). Ushizima et al (2013) realizou uma segmentação de núcleos e citoplasma de várias células do colo do útero aplicando RGB Mean-shift e Watershet, esse experimento apresentou como resultado inicial imagens de alta qualidade, obtidas pelo sistema ThrinPet.

No geral, as pesquisas recentes que conseguem segmentar e classificar um alto volume de imagens de núcleos e de citoplasmas de células do colo do útero costumam usar os métodos de clusterização, métodos de tresholding automáticos e análises de textura (GUAN et al., 2011)?, (LORCA; ARZOLA; PEREIRA, 2010)?. Contudo, as referidas pesquisadas só trabalham com imagens de alta qualidade. Isso dificulta a aplicação das referidas técnicas e métodos nos países em desenvolvimento. Em países como o Brasil, a Índia e a Colômbia são adotadas técnicas baratas e convencionais para a realização do Papanicolau, a técnica usada na coleta do exame é o esfregaço de células em lâminas de vidro (MINISTÉRIO

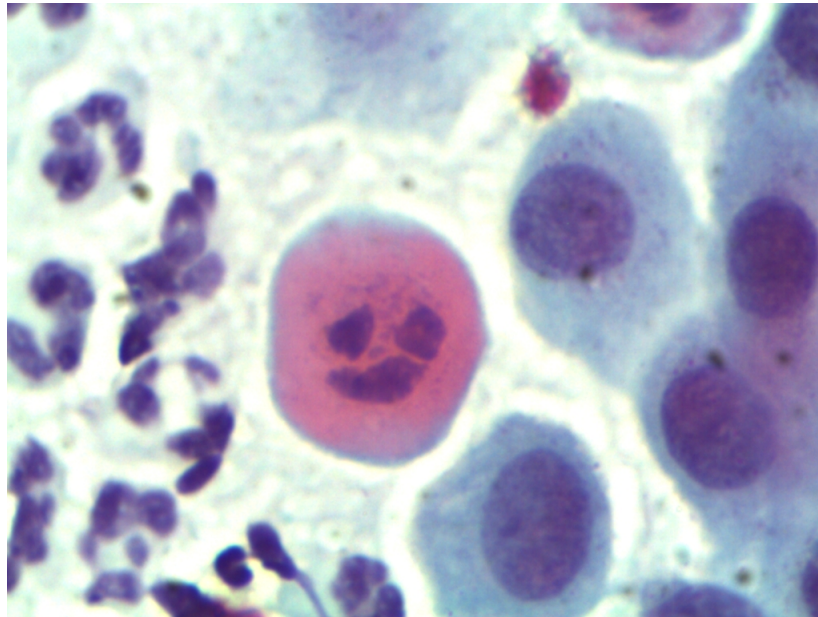


Figura 1: Conjuntos de multiples células do colo de útero

DA SAÚDE, 2016)?. Nesse cenário, a imagem do material coletado é frequentemente de baixa qualidade.

Liu (2009)?, Guan (2011)?, Chankong (2014)?, Braga (2015)?, Jurjo (2015)? e Mader (2016) empregam o processamento de imagens e a clusterização em suas pesquisas, um dos algoritmos mais utilizados é o k-means.

A clusterização usando k-means é baseada na minimização da distância normalizada (HARTIGAN; WONG, 1979)?. A distância é normalizada com respeito ao centro do cluster  $M$  e definido para ser  $D$

$$D_j = (X_k - M_j)^T K_j^{-1} (X_k - M_j), j = 1, 2, \dots, M \quad (1)$$

Onde

$$K_j = \frac{1}{N_j} \sum (X_k - M_j)(X_k - M_j)^T, j = 1, 2, \dots, M \quad (2)$$

e  $X_k$ ,  $K = 1, 2, \dots, N$  são os vetores característicos avaliados,  $K$  é a estimativa da matriz de covariância,  $N$  é o número total de amostras, e  $M$  é o número de grupos.

A clusterização pelo algoritmo k-means precisa de várias somatórias e, portanto, necessita de vários recursos computacionais. Mas, devido ao aumento de processamento computacional nos últimos anos é possível realizar a clusterização de imagens de maneira simples e eficiente. E o algoritmo k-means pode ser aproveitado para delimitar imagens de baixa qualidade. Um dos motivos é a capacidade de identificar grupos de dados relacionados de maneira eficiente.

O maior desafio na segmentação de imagem de baixa qualidade é o pré-processamento da amostra, é necessário ter cautela para não perder informações importantes. Normalmente, a perda de informação acontece quando são usados filtros para diminuir os ruídos da imagem. Os filtros mais adotados na área de segmentação de células do colo de útero são os gaussianos e os de média (USHIZIMA et al., 2013)?, (YUNG -FU, 2014)?.

## REFERÊNCIAS

ANSCHAU, F.; GINECOLOGIA, D. DE; LUCAS, S. Citologia Cervical em Meio Líquido Versus Citologia Convencional Liquid-based Cervical Cytology Versus Conventional Cytology Comparação entre a Citologia Convencional e a Citologia em Meio Líquido. *Femina*, v. 34, n. 05, p. 7, 2006.

BERGMEIR, C.; SILVENTE, M. G.; BENÍTEZ, J. M. Segmentation of cervical cell nuclei in high-resolution microscopic images: A new algorithm and a web-based software framework. *Computer Methods and Programs in Biomedicine*, v. 107, n. 3, p. 497-512, 2011.

BEUCHER, S. The Watershed Transformation Applied to Image Segmentation Proceedings of the 10th Pfefferkorn Conference on Signal and Image Processing in Microscopy and Microanalysis. *Anais...*1992

BRAGA, A. M.; MARQUES, R. C. P.; USHIZIMA, D. M. Segmentação de Células Cervicais Utilizando Level Set Binário-Hierárquico Conference on Graphics, Patterns and Images, 28 (SIBGRAPI). *Anais...*Porto Alegre: Sociedade

Brasileira de Computação, 2015

BUDGE, M. et al. Comparison of a self-administered tampon ThinPrep test with conventional pap smears for cervical cytology. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, v. 45, n. 3, p. 215-219, 2005.

CENCI, M.; NAGAR, C.; VECCHIONE, A. PAPNET-assisted primary screening of conventional cervical smears. *Anticancer Research*, v. 20, n. 5 C, p. 3887-3889, 2000.

CHANKONG, T.; THEERA-UMPON, N.; AUEPHANWIRIYAKUL, S. Automatic cervical cell segmentation and classification in Pap smears. *Computer Methods and Programs in Biomedicine*, v. 113, n. 2, p. 539-556, 2014.

FRANCO, R. A. S. M. G. M. A. G. Fronteras en el desenvolvimiento tecnológico: Una visión de las distancias entre países en la prevención del Cáncer de Cuello de Útero. In: *Debates Interdisciplinarios VII*. Santa Catalina: Editora Unisul, 2015. p. 17.

FRANCO, R. A. S.; MARTINS, P. S.; CARVALHO, M. A. G. DE. Cytological Low-Quality Image Segmentation Using Nonlinear Regression, K-means and Watershed. *WSCG 2016 - 24th WSCG*, p. 8, 2016.

GLOBOCAN. Estimate Cancer Incidence, Mortality and prevalence Worldwide in 2012. Disponível em: <http://globocan.iarc.fr/Default>

GUAN, R. et al. Text clustering with Seeds Affinity Propagation. *IEEE Transactions on Knowledge and Data Engineering*, v. 23, n. 4, p. 627-637, 2011.

HARTIGAN, J. A.; WONG, M. A. A K-Means Clustering Algorithm. *Applied Statistics*, v. 28, n. 1, p. 100-108, 1979.

JAYASINGH, E.; STEPHEN, A. Nominated Texture Based Cervical Cancer Classification. *Computational and Mathematical Methods in Medicine*, v. 2015, p. 10, 2014.

JURJO, D. L. B. R. et al. Analysis of the structural behavior of a membrane using digital image processing. *Mechanical Systems and Signal Processing*, v. 54, p. 394-404, 2015.

KINOSHITA, S. K. et al. Content-based retrieval of mammograms using visual features related to breast density patterns. *Journal of digital imaging*, v. 20, n. 2, p. 172-190, jun. 2007. LIU, D.; YU, J. Otsu method and K-means. *Proceedings - 2009 9th International Conference on Hybrid Intelligent Systems, HIS 2009*. Anais...2009

LORCA, G.; ARZOLA, J.; PEREIRA, O. Segmentación de Imágenes Médicas Digitales mediante Técnicas de Clustering Digital Medical Image Segmentation with Clustering Techniques. *revista aporte santiagino*, n. 20708361x, 2010.

LOZANO, M.; MIGUEL, C. DE; COUSILLAS. Nuevas tecnologías en Citopatología. *Sociedad Española de Anatomía Patológica*. Anais...Pamplona: Ainzúa Navarra, 2011

MADER, K.; STAMPANONI, M. Identifying layers in random multiphase structures. *AIP Conference Proceedings*, v. 020046, p. 1-13, 2016.

MALM, P.; BRUN, A.; BENGTSSON, E. Papsynth: Simulated bright-field images of cervical smears. *2010 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2010 - Proceedings*. Anais...2010

MINISTÉRIO DA SAÚDE. Estimativa 2016 Incidência de Câncer no Brasil. Coordenação ed. Rio de Janeiro: Instituto Nacional de Câncer José Alencar, 2016.

OJALA, T.; PIETIKAINEN, M.; HARWOOD, D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision and Image Processing., Proceedings of the 12th IAPR International Conference on*. Anais...1994

USHIZIMA, D. M. et al. Automated Pap Smear Cell Analysis: Optimizing the Cervix Cytological Examination. *2013 12th International Conference on Machine Learning and Applications*. Anais...2013

Y. NOGUCHI, B. STENKVIST, J. HOLMQUIST, E. B. AND B. O. B. Low Resolution Inspection System of Gynecological Cell Specimens. *Bull. Electrotech*, n. 39, p. 293 - 323, 1975.

YOSHIO FUKUDA; MARIA Q. FENG; MASANOBU SHINOZUKA. Cost-effective vision-based system for monitoring dynamic response of civil engineering structures. *Structural Control and Health Monitoring*, n. May 2011, p. 918-935, 2011.

YUNG -FU, C. Semi-Automatic Segmentation and Classification of Pap Smear Cells. *IEEE Journal of Biomedical and Health Informatics*, v. 18, p. 94 - 108, 2014.