Seeded Region Growing Features Extraction Algorithm; Its Potential Use in Improving Screening for Cervical Cancer

N. A. Mat-Isa[1]**, M. Y. Mashor**[2]**,** and **N. H. Othman**[3]

[1,2]Control and Electronic Intelligent System (CELIS) Research Group,
School of Electrical & Electronic Engineering,
Universiti Sains Malaysia, Engineering Campus Seri Ampangan,
14300 Nibong Tebal, Seberang Perai Selatan, Malaysia.
E-mail [1]: ashidi75@yahoo.co.uk, E-mail [2]: yusof@eng.usm.my

[3]Pathology Department, School of Medical Science
Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia.
E-mail[3]: hayati@kb.usm.my

Abstract

Region growing algorithm has successfully been used as a segmentation technique of digital images. The current study went one step further by utilizing the potential use of thresholding the region growing algorithm as features extraction technique. The proposed features extraction algorithm is called seeded region growing features extraction (SRGFE). This algorithm was used to extract four features of cervical cells; size of nucleus, size of cytoplasm, grey level of nucleus and grey level of cytoplasm. Correlation test was applied between data extracted using the proposed SRGFE algorithm with the data extracted manually by cytotechnologists. The high correlation value obtained in the correlation test show that the SRGFE algorithm is suitable and has high capability to be used as an image extraction technique to extract important features of cervical cells. This would assist cytopathologists and cytotechnologists in the cervical cancer screening by providing accurate value of size and grey level of nuclear and cytoplasmic features.

Keywords

Seeded region growing features extraction (SRGFE), cervical cells, Pap smear, medical imaging, features extraction.

1. Introduction

Cervical cancer is the second common form of cancer among women (Othman, 2003, WebMD, 2003). Cervical cancer is a silent cancer. Unlike other cancers that cause pain, noticeable lumps or other early symptoms, cervical cancer has no telltale symptoms until it is so advanced that is usually unresponsive to treatment (WebMD, 2003). Only in its late stage, cervical cancer causes pain in the lower abdominal or back regions. However, most cervical cancer takes many years to develop from normal to dangerous stage. The mortality related to cervical cancer can be substantially reduced through early detection and treatment.

Currently, Pap test is the most popular method to detect the presence of abnormal

cells arising from the cervix. Several previous studies by Cohn & Herzog (2001), Breen *et al.* (2001), Framer (2001), Kuie (1996) and Adami *et al.* (1994) showed that the chances for a woman acquiring cervical cancer is reduced if she has Pap test regularly. A computer model analysis has found that the proportion of cervical cancer reduction is approximately 30% for a smear taken once every 10 years, 80% for one done every five years, 90% for every three years, 91% for every two years and 92% for an annual smear (Kuie, 1996).

Cervical cancer has been classified in a variety of ways. The new and commonly used is the Bethesda system. Abnormal cervical cells are classified into two types; *low-grade squamous intraepithelial lesions* (LSIL) and *high-grade squamous intraepithelial lesions* (HSIL) (WebMD, 2003, Stoler, 2002, Kuie, 1996). Cytopathologists differentiate both types of abnormal cervical cells and normal cells based on several morphologies. The abnormal cervical cells show changes in nucleocytoplasmic ratio. The cytoplasm size decreases but the nucleus size increases from normal cells to HSIL cells through LSIL cells (Crum, 1994, Govan *et al.*, 1986). This phenomenon increases the nucleocytoplasmic ratio. Besides that, the abnormal cervical cells also show changes in colour (grey level) of nucleus and cytoplasm (WebMD, 2003, Crum, 1994). The grey levels for the cells' structures become darker from normal cells to HSIL cells through LSIL cells.

Some studies (Nanda *et al.*, 2000, Othman *et al.*, 1997, Kuie, 1996, Hislop *et al.*, 1994) proved that sometimes the Pap test is not effective. The determination of abnormal cervical cells can sometimes be missed in certain situation. Generally, the accuracy of Pap test depends on the quality of the Pap smear samples. The heavily stained Pap smear may be masked by menstrual blood, vaginal discharge, air artefacts etc thus obscuring the abnormal cervical cells. Sometimes, overexposing or underexposing to the microscope light may also blur the Pap smear images. Thus, the cytopathologists may have difficulty in extracting the important morphologies of cervical cells due to these problems. In this study, feature extraction algorithm based on region growing is proposed to extract the size and grey level of the region of interest. The algorithm is called seeded region growing features extraction (SRGFE). Then, the SRGFE algorithm will be used to extract the size of nucleus and cytoplasm of cervical cells as well as their grey level.

2. Region Growing Based Features Extraction

Various medical imaging modalities such as X-ray, computer tomography (CT) and magnetic resonance imaging (MRI) are widely available and used in routine clinical practice. It has been generally accepted that visual interpretation of such images is highly subjective. As a result, there is a great need for image processing techniques that must be applied on medical images to assist the pathologists and radiologists for easier and better diseases diagnosis. One of the common image processing techniques is features extraction. Applications of various features extraction techniques in medical images have grown in recent years (Christoyianni *et al.*, 1999, Chen *et al.*, 1998, Pham *et al.*, 1997, El-Faramawy *et al.*, 1996, Chitre *et al.*, 1993).

In medical field, region growing algorithm has successfully been used as segmentation technique of medical images (Ngah *et al.*, 2002, Venkatachalam *et al.*, 2002, Tuduki *et al.*, 2000, Ooi *et al.*, 2000, Justice & Stokely, 1997). The current study

will utilize the potential use of thresholding the region growing algorithm as features extraction technique. The proposed algorithm is called seeded region growing features extraction (SRGFE). As shown in Figure 1, the SRGFE is used to extract the size and grey level of certain region of interest on a digital image.

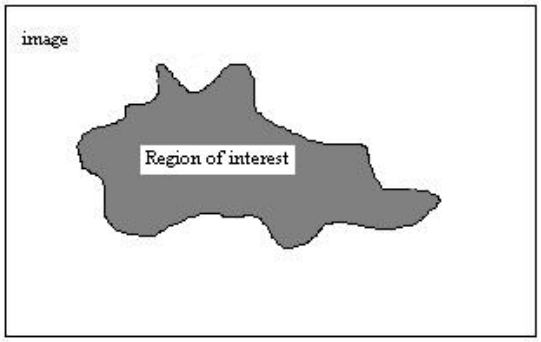


Figure 1: The region of interest for features extraction process.

Referring to Figure 1, the size of the region is calculated as a total number of pixels in the region and given by the following equation.

$$Size = Total\ \ of\ \ \ pixels\ \ in\ \ the\ \ region \quad (1)$$

The grey level of the region is calculated as mean value of all pixels in the region and given by the following equation.

$$Grey\ level = \frac{Total\ of\ \ grey\ level\ \ for\ \ all\ \ pixels\ in\ \ the\ \ region}{Total\ of\ \ \ pixels\ in\ \ the\ \ region} \quad (2)$$

In the SRGFE algorithm, the user needs to determine the region of interest by clicking mouse on any pixels in the region. The user also needs to determine the threshold value. The algorithm of the proposed SRGFE algorithm can be implemented as:

1. Apply three preprocessing techniques to the image, which are median filter, histogram normalization and histogram equalization.
2. Determine the threshold value.
3. Click mouse in the region of interest.

(Note: The pixel, which mouse is clicked on it will be used as initial seed pixel)

4. Choose $N \times N$ neighbourhood as shown in Figure 2 (for $N = 5$)

(Note: Initial seed point must be located at the centre of all its neighbours and N must be an odd number)

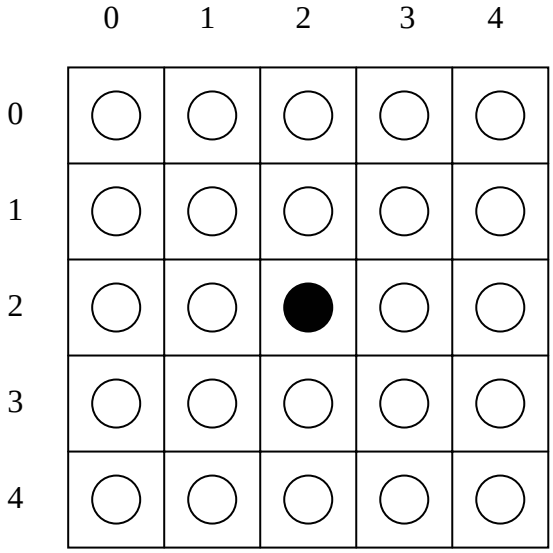


Figure 2: Location of initial seed pixel (point) and its 5×5 neighbourhood

5. Set the initial value for *Total of pixels in the region* with 1 and *Total of grey level for all pixels in the region* with original grey level value of the initial seed pixel.
6. Calculate the mean value, $\bar{x}$ (which is known as region mean) and standard deviation, σ of the $N \times N$ neighbourhood using Equation 3 and 4 respectively.

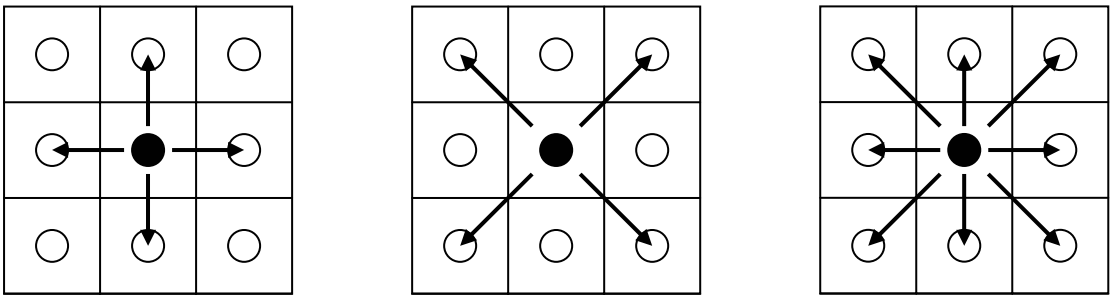
$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (3)$$

$$\sigma = +\sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} \quad (4)$$

where x_i is grey level of i-th pixel in the $N \times N$ neighbourhood and n is total of pixels in $N \times N$ neighbourhood.

7. Grow the seed pixels to its neighbour's pixels. Compare the grey level of the seed pixel with its neighbour's pixel. Include the neighbour pixel into the region if it satisfy one of the conditions listed below (Romberg *et al.*, 1997):
   a. If the gradient of the pixel is less than 95% of the equalized histogram AND the grey level of the pixel is less or equal to the preselected threshold.
   b. If the gradient of the pixel is more than or equal to 95% of the equalized histogram AND the grey level of the pixel is not more than or equal to one standard deviation away from the region mean.
8. If the neighbour pixel is included into the region:
   a. Add one (1) to *Total of pixels in the region* value.
   b. Add original grey level of the neighbour pixel to *Total of grey level for all pixels in the region* value.
9. Set the neighbour pixel, which is added to the region in Step (7) as a new seed pixel.
10. Repeat Step (6) to (9) until all pixels have been considered to be grown or the pixel cannot be grown anymore.
11. Calculate the value of the size and grey level of the region according to Equation (1) and (2) respectively.

For Step (7) in the proposed algorithm, there are three possible ways for seed pixel to grow as shown in Figure 3(a), (b) and (c) respectively.



(a) (b) (c)

Figure 3: The seed pixel growing towards (a) its 4 adjacent neighbours (b) its 4 diagonal neighbours and (c) its 8 surrounding neighbours.

3. Methodology and Materials

In this study, the SRGFE algorithm is used to extract the size and grey level of two cervical cells' structures; the nucleus and the cytoplasm. Therefore, the original SRGFE algorithm needs to be modified. Two thresholds must be determined by the user, which represent threshold value for the nucleus and the cytoplasm respectively. The modification is made in the Step (2) of the SRGFE algorithm. For initial seed pixel determination as in Step (3), the mouse must be clicked in the nucleus regions of the cell of interest as shown in Figure 4. Thus, the growing process will start in the nucleus regions. Step (4) to (11) of the SRGFE algorithm needs to be repeated, which the first and second implementation will extract the nucleus and cytoplasm features respectively. For extraction process for both structures, the initial seed pixel is set to location of the clicking mouse by the user.

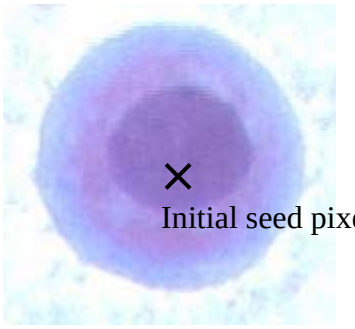
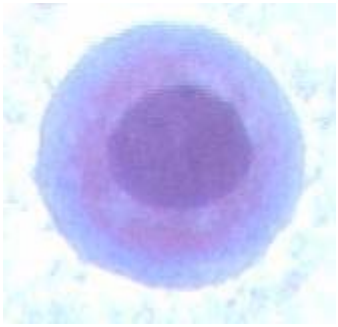


Figure 4: The location of initial seed pixel for extraction of cervical cells' features.

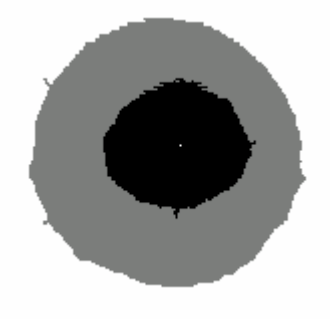
350 images of cervical cells (161 normal cells, 93 LSIL cells and 96 HSIL cells) were used to determine the capability and the suitability of the SRGFE algorithm to extract the size and grey level of the nucleus and cytoplasm. The images have been captured from Pap smear samples. The samples were provided by Hospital Universiti Sains Malaysia (HUSM). The cells features were also extracted manually by experienced and certified cytotechnologist from HUSM using an image analyser. In the current study, the size of nucleus and cytoplasm that are calculated by SRGFE algorithm are represented by total number of pixels, while cytotechnologist calculated the size of nucleus and cytoplasm in micrometer. However, the colour of nucleus and cytoplasm that are determined by SRGFE algorithm and cytotechnologist are same, which are in grey level value. The current study determined the accuracy of the features extraction using the proposed SRGFE algorithm by comparing with those done by the cytotechnologist. The comparison is done using correlation test. The relationship is described as strong if the correlation value is higher than 0.8, moderate if equal or more than 0.5 but less than 0.8 and weak if less than 0.5 (Devore, 2000).

4. Results and Discussion

The proposed SRGFE algorithm was implemented on 350 images of cervical cells, which were taken from Pap smear samples. Figure 5, 6 and 7 show the feature extraction result using SRGFE algorithm for three samples called CervixCell1, CervixCell2 and CervixCell3 respectively. Image (a) shows the original image, while image (b) shows the image after applying SRGFE algorithm. For image (b), the black region (grey level of 0) represents the nucleus region and the grey region (grey level of 127) represents the cytoplasm region that both were extracted by SRGFE algorithm. The white region (grey level of 255) represents the unwanted region (background region). Table 1 shows the results of features extraction of those images using SRGFE algorithm.



(a) Original CervixCell1 image



(b) CervixCell1 image after applying SRGFE

Figure 5: Results of features extraction using SRGFE algorithm for CervixCell1 image.

(a) Original CervixCell2 image

(b) CervixCell2 image after applying SRGFE

Figure 6: Results of features extraction using SRGFE algorithm for CervixCell2 image.

(a) Original CervixCell3 image

(b) CervixCell3 image after applying SRGFE

Figure 7: Results of features extraction using SRGFE algorithm for CervixCell3 image.

Table 1: Results of features extraction using SRGFE algorithm

Data Type	CervixCell 1	CervixCell 2	CervixCell 3
Size of nucleus	3531 pixels	3201 pixels	4261 pixels
Size of cytoplasm	14027 pixels	19562 pixels	15790 pixels
Grey level of nucleus	146.88	139.38	149.22
Grey level of cytoplasm	180.76	165.73	169.27

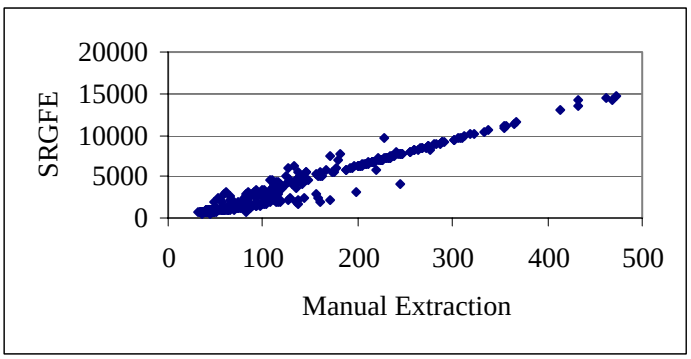
With reference to Figures 5, 6 and 7, it can be observed that the SRGFE algorithm has a high ability to extract the nucleus and cytoplasm size of the cervical cells that closely represents the original images. The shape and size of both cervical cells structures are maintained.

As mentioned in the previous section, the accuracy of the features extraction using the proposed SRGFE algorithm was compared with those manually done by the cytotechnologist using correlation test. The correlation test results between data extracted using SRGFE algorithm and data extracted manually by cytotechnologist are shown in Table 2. Figure 8(a), (b), (c) and (c) show plotted graph of data extracted using SRGFE algorithm versus data extracted manually by cytotechnologist, where each graph represents data for size of nucleus, size of cytoplasm, grey level of nucleus and grey level of cytoplasm respectively. For data size, the value of size extracted using SRGFE algorithm is calculated in total of pixels, while for size extracted manually, the value is calculated in micrometer.

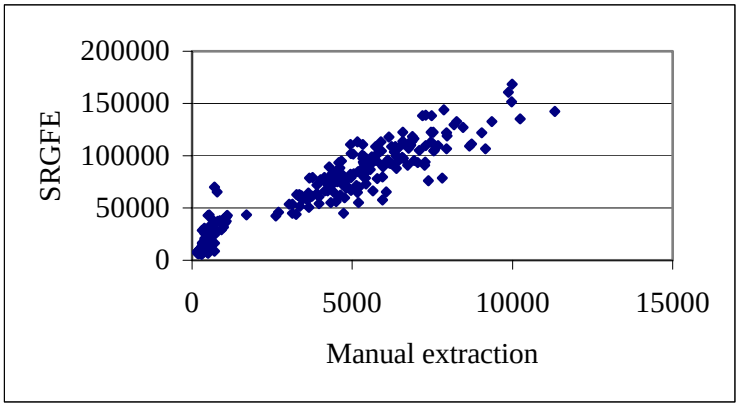
The results in Table 2 and Figure 8 show that SRGFE algorithm extracts the cervical cells' features with accuracy as good as manual technique. The nucleus size, cytoplasm size and cytoplasm grey level show correlation higher than 0.95 and depicts very strong linear relationship between both methods. Correlation test for data of nucleus grey level is slightly lower, 0.884525, which is still within the strong linear relationship range. Superior correlation test suggests that the proposed SRGFE algorithm is a good feature extraction technique. The application in Pap smear would be able to assist cytopathologists for better cervical cancer screening by providing accurate value of size and grey level of nucleus and cytoplasm of cervical cells.

Table 2: Results for correlation test.

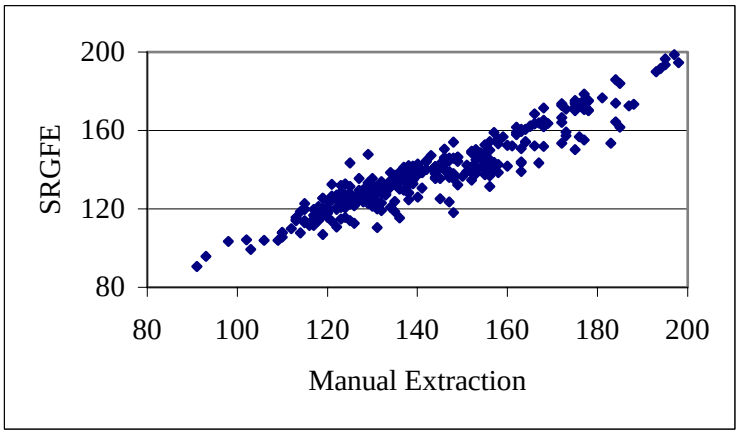
Data Type	Correlation Value
Size of nucleus	0.973237
Size of cytoplasm	0.963314
Grey level of nucleus	0.884525
Grey level of cytoplasm	0.953263



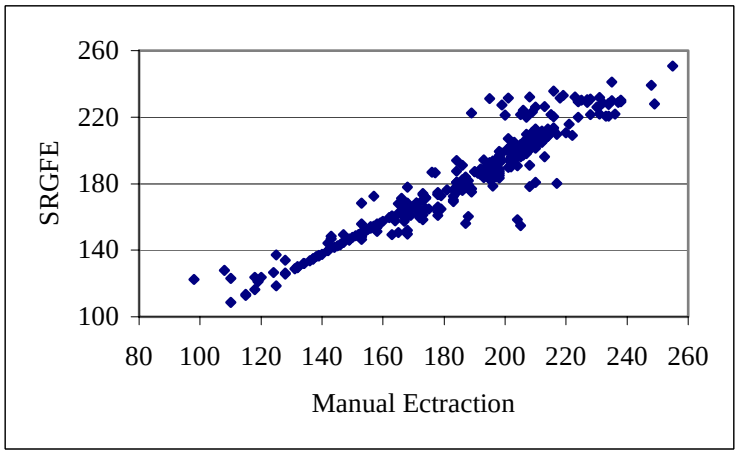
(a) Size of nucleus



(b) Size of cytoplasm



(c) Grey level of nucleus



(d) Grey level of cytoplasm

Figure 8: Graph of data extracted using SRGFE versus data extracted manually by cytotechnologist.

5. Conclusion

The SRGFE algorithm has been proposed as features extraction technique for digital images. The current study used the proposed algorithm to extract 4 cervical cells features; size of nucleus and cytoplasm, and grey level of nucleus and cytoplasm. The data extracted using SRGFE algorithm gave high correlation value when compared with data extracted manually by cytotechnologist using image analyser. The strong linear relationship between both types of data clearly shows that the SRGFE algorithm is suitable to be used to extract cell features. The SRGFE algorithm, when applied to Pap smear could assist the cytopathologists in cervical cancer screening by providing accurate value of size and grey level of nucleus and cytoplasm.

References

[1] Adami, H. O., Ponten, J., Sparen, P., Bergstrom, R., Gustafsson, L. & Friberg, L. G. (1994). "Survival Trend After Invasive Cervical Cancer Diagnosis in Sweden Before and After Cytologic Screening". *Cancer*. 73(1). 140-147.

[2] Breen, N., Wagener, D. K., Brown, M. L., Davis, W. W. & Barbash, R. B. (2001). "Progress in Cancer Screening Over a Decade: Results of Cancer Screening From the 1987, 1992 and 1998 National Health Interview Surveys". *Journal of the National Cancer Institute*. 93(22). 1704-1713.

[3] Chen, C. W., Luo, J. & Parker, K. J. (1998). "Image Segmentation via Adaptive K-Mean Clustering and Knowledge-Based Morphological Operations with Biomedical Applications". *IEEE Transaction on Image Processing*. 7(12).1673-1683.

[4] Chitre, Y, Dhawan, A. P. & Moskowitz. (1993). "Artificial Neural Network Based Classification of Mammogramic Microclassification Using Image Structure Feature". *International Journal of Pattern Recognition and Artificial Intelligence*. 7(6). 1377-1401.

[5] Christoyianni, I., Dermatas, E. & Kokkinakis, G. (1999). "Neural Classification of Abnormal Tissue in Digital Mammography Using Statistical Features of The Texture". *Proceedings of The 6th IEEE International. Conference on Electronics, Circuits and Systems*. 1. 117-120.

[6] Cohn, D. E. & Herzog T. J. (2001). "New Innovations In Cervical Cancer Screening". *Clinical Obstetric and Gynaecology*. 44(3). 538-549

[7] Crum, C. P. (1994). Female Genital Tract. In *Robbins: Pathologic Basis of Disease*.(Cotran, R. S., Kumar, V. & Robbins, S. L.). 1033-1088. 5th Edition. Philadelphia: W. B. Saunders.

[8] Devore, J. L. (2000). *Probability and Statistics for Engineering and The Sciences*. 5th Edition. New York: Duxbury Press.

[9] El-Faramawy, N. M., Rangayyan, R. M., Desautels, J. E. L. & Alim, O. A. (1996). "Shape Factor for Analysis of Breast Cancer Tumors in Mammograms". *Proceedings of Canadian Conference on Electrical and Computer Engineering*. 1. 355-358.

[10] Framer, P. S. (2001). "Screening for Cancer: Progress but More Can Be Done". *Journal of the National Cancer Institute*. 93(22). 1676-1677.

[11] Govan, A. D. T., MacFarlane, P. S. & Callander, R. (1986). *Pathology Illustrated*. 2nd Edition. Edinburgh: Churchill Livingstone.

[12] Hislop, T. G., Band, P. R., Deschamps, M., Clarke, H. F., Smith, J. M. & Ng, V. T. Y. (1994). "Cervical Cancer Screening in Canadian Native Women: Adequacy of The Papanicolaou Smear". *The Journal of*

Clinical Cytology and Cytopathology. 38(1). 29-32.

[13] Kuie, T. S. (1996). *Cervical Cancer: Its Causes and Prevention*. Singapore: Times Book International.

[14] Nanda. K, McCrory, D. C., Myers, E. R., Bastian, L. A., Hasselblad, V., Hickey, J. D., Matchar, D. B. (2000). "The Accuracy of The Papanicolaou Test in Screening for and Follow-Up of Cervical Cytologic Abnormalities: A Systematic Review". *Annual International Medical*; 132. 810-819.

[15] Ngah, U. K., Ooi, T. H., Sulaiman, S. N. & Venkatachalam, P. A. (2002). "Embedded Enhancement Image Processing Techniques on A Demarcated Seed Based Grown Region". *Proceedings of Kuala Lumpur International Conference on Biomedical Engineering.* 170-172.

[16] Ooi, T. H., Ngah, U. K., Khalid, N. E. A. & Venkatachalam, P. A. (2000). "Mammagraphic Calcification Clusters Using The Region Growing Technique". *Proceedings of The New Millennium International Conference on Pattern Recognition, Image Processing and Robot Vision*. 157-163.

[17] Othman, N. H. (2003). *Cancer of The Cervix: From Bleak Past to Bright Future*. Kelantan, Malaysia: Pustaka Reka.

[18] Othman, N. H., Ayub, M. C., Aziz, W. A. A., Muda, M., Wahid, R. & Selvarajan, S. (1997). "Pap Smears – Is It An Effective Screening Methods for Cervical Cancer Neoplasia? – An Experience with 2289 Cases". *The Malaysian Journal of Medical Sciences*. 4(1). 45-50.

[19] Pham, D., Prince, J. L., Chen, Y. X. & Dagher, A. P. (1997). "An Automated Technique for Statistical Characterization of Brain Tissues in Magnetic Resonance Imaging". *International Journal of Pattern Recognition and Artificial Intelligence*. 11(8). 1189-1211.

[20] Romberg, J., Akram, W. & Gamiz, J. (1997). "Image Segmentation Using Region Growing. Citing from internet source" URL http://www.owlnet.rice.edu/elec539/projects97/WDEKnow/ index. (2002).

[21] Stoler, M. H. (2002). "New Bethesda Terminology and Evidence-Based Management Guidelines for Cervical Cytology Findings". *JAMA*. 287(16). 2140-2141.

[22] Tuduki, Y., Murase, K., Izumida, M., Miki, H., Kikuchi, K., Murakami, K. & Ikezoe, J. (2000). "Automated Seeded Region Growing Algorithm for Extraction of Cerebral Blood Vessels from Magnetic Resonance Angiographic Data". *Proceedings of The 22nd Annual International. Conference of The IEEE Engineering in Medicine and Biology Society*. 3. 1756-1759.

[23] Venkatachalam, P. A., Ngah, U. K., Hani, A. F. M. & Shakaff, A. Y. M. (2002). "Seed Based Region Growing Technique in Breast Cancer Detection and Embedded Expert System". *Proceedings of International Conference on Artificial Intelligence in Engineering and Technology*. 464-469.

[24] WebMD. (2003). "How Can Cervical Cancer Be Prevented?". Citing from internet source. URL *http://www.webmd.com/content/dmk/dmk_article_3961643*