Grading of Cervical Intraepithelial Neoplasia Using Spatial Frequency for Optical Histology

Yang Pu[1], Jaidip Jagtap[2], Asima Pradhan[2], and Robert R. Alfano[*1]

[1]Institute for Ultrafast Spectroscopy and Lasers, and Department of physics, the City College of the City University of New York, New York, NY 10031

[2]The Department of Physics, Indian Institute of Technology (IIT), Kanpur, India 208016

ABSTRACT

It is important to detect cervical dysplasia, Cervical Intraepithelial Neoplasia (CIN). CIN is the potentially premalignant and abnormal squamous cells on surface of cervix. In this study, the spatial frequency spectra of pre-cancer cervical tissues are used to detect differences among different grades of human cervical tissues. Seven sets of thick tissue sections of human cervix of normal, CIN 1, CIN 2, and CIN 3 tissues are studied. The confocal microscope images of the stromal region of normal and CIN human tissues were analyzed using Fast Fourier Transform (FFT) to generate the spatial spectra. It is observed that higher frequency components exist in CIN tissues than those in normal tissue, as well as those in higher grade CIN tissue than those in lower grade CIN tissue. The width of the spatial frequency of different types of tissues is used to create a criterion for CIN grading by training a support vector machine (SVM) classifier. The results show that the randomness of tissue structures from normal to different stages of precancer in cervical tissue can be recognized by fingerprints of the spatial frequency. The efficacy of spatial frequency analysis for CIN grading is evaluated as excellent since high AUC (area under the ROC curve), sensitivity and specificity are obtained by the statistics study. This works lays the foundation of using spatial frequency spectra for a histology evaluation.

Keywords: Cervical Intraepithelial Neoplasia (CIN), spatial frequency spectra, Fast Fourier Transform (FFT), support vector machine classifier (SVM), Gaussian Fitting,

1. INTRODUCTION

Cervical cancer is the second most precedent cancer worldwide in women after breast cancer [1]. Although not a cancer, over 12% of cervical intraepithelial neoplasia (CIN) progress into cancer [1]. Current techniques for CIN include the Papanicolaou or "Pap" smear and colposcopy [2]. The clinical diagnosis of potential CIN depends on histological evaluation of a biopsy of a suspected abnormal area, which is invasive and subjective. It is well known that diagnosis of a premalignant lesion is critical for the success of cancer therapy. Since CIN is a precursor of cervical cancer [1], it is important to find an efficient and objective way to detect CIN.

An image plane with a light intensity distribution is composed of "spatial frequencies", which is similar to the various frequencies making up a signal in the time domain [3]. During the development from normal to high CIN stages, the connective tissue framework of cervix can be impaired during Neoplasia evolution [1]. Such alterations of tissue biochemistry and morphology may be revealed in the tissue scattering and fluorescence [4]. Spatial frequency spectral analysis is able to provide information of the periodic and/or random structures of a sample from the reflected light intensity distributions as would be seen by a pathologist. Moreover, the periodic and/or random structures in the stromal region of CIN tissue is related to the stage of precancer progression [1, 2]; therefore, spatial frequency spectral analysis may offer an alternative and objective way to analyze CIN grades and cancer tissues.

This paper presents the statistical criteria utilization of spatial frequency of cervical tissue microscope images to grade and distinguish tissue change among normal and different stages of dysplasia tissues. Confocal microscope images of stromal region of normal and different grades of CIN human tissues were analyzed using Fast Fourier Transform (FFT) to generate the spatial spectra. The results show that the randomness of tissue structures from normal to different stages of CIN tissues can be recognized by fingerprints of the spatial frequency, which lays the foundation of using spatial frequency spectra from FFT analysis for a histology evaluation. The statistical analyses of the spatial frequency spectra of cervix precancer tissue demonstrate the effective criteria to grade cervix tissue

* alfano@sci.ccny.cuny.edu; phone: (212) 650-5531; fax: (212) 650-5530

Optical Biopsy XII, edited by Robert R. Alfano, Stavros G. Demos, Proc. of SPIE Vol. 8940, 894015 894016 · © 2014 SPIE · CCC code: 1605-7422/14/$18 · doi: 10.1117/12.2035498

states. The analyses of twenty-eight cervical precancer tissues with spatial frequency spectra provide the preliminary demonstration of the effective criteria to grade CIN states.

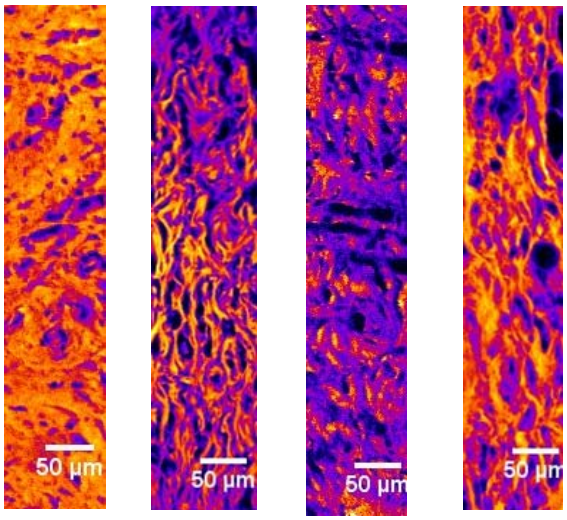
2. SAMPLES AND METHOD

Seven sets of human cervix of normal, CIN 1, CIN 2, and CIN 3 tissues of 5 μm thickness stained by H&E are used in this study. These stained tissue sections were irradiated by 514 nm (Argon ion laser) and fluorescence images captured at emission wavelength 532 nm using a confocal microscope (Leica TCS SP5). Figures 1 (a), (b), (c), and (d) show a typical set of confocal microscope images of the normal, CIN 1, 2, and 3 tissues, respectively. The intensities distribution of these images can be expressed as the two dimensional (2D) functions *f(x,y)* in spatial coordinates (*x*,*y*) [3]:

$$f(x,y) = \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} F[u,v] \exp\left(i2\pi \left(\frac{ux}{L_x} + \frac{vy}{L_y} \right) \right), \quad (1)$$

where u and v are the numbers of cycles of *f*(*x*,*y*) with a periodic L_x and L_y in the *x* and *y* directions, respectively. Discrete Fourier Transform (DFT) is usually-used mathematical tools to convert 2D spatial intensity distribution function *f*(*x*,*y*) into the 2D spatial spectrum *F*(*u*,*v*) by sampling a finite extent $N \times N$ [3]:

$$F(u,v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \left(-i2\pi \left(\frac{ux}{N} + \frac{vy}{N} \right) \right). \quad (2)$$



(a) (b) (c) (d)

Fig.1. A typical set of cropped portions of confocal microscope images of stromal regions of (a) normal, (b) CIN 1, (c) CIN 2, and (d) CIN 3 cervical tissues

The magnitude spectra, |*F*(*u*,*v*)|, are calculated by [3]:

$$|F(u,v)| = \sqrt{R^2(u,v) + I^2(u,v)}. \quad (3)$$

where *R*(*u*,*v*) and *I*(*u*,*v*) are the real and imaginary parts of the spatial frequency spectrum, respectively.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

In order to obtain the spatial spectral information of randomness and aperiodicity of CIN tissues for different grades, the 2D DFT of Figs. 1 was performed by sampling N = 256. Figures. 2 (a), (b), (c), and (d) display the 2D amplitudes spectra of normal, CIN 1, 2, and 3 tissues, respectively. For visual purposes, the amplitude spectra shown in Figs. 2 were truncated by linear mapping of the initial amplitudes and exhibited as logarithms of amplitudes [7] in the color range of [0, 255].

Figures 2 (a) to (d) are typical results of spatial frequencies distribution where the dominant spatial frequency is at the origin ($u = 0$, $v = 0$), and increases in all directions away from the center [8]. The salient differences observed among Figs. 2 (a) to (d) are that higher frequency components are dominant in CIN tissues than in normal tissue, and higher grade CIN over the lower grade CIN tissue. The features displayed in Figs. 2 are thus stated as follows: for the normal tissue and the lower grade CIN tissues, the lower frequency amplitudes mostly dominate over the mid-range and high-frequency ones, but the mid-range and high-frequency amplitude spectrum can be perceived more clearly with the evolution from normal to CIN, and development from low grade to high grade CIN.

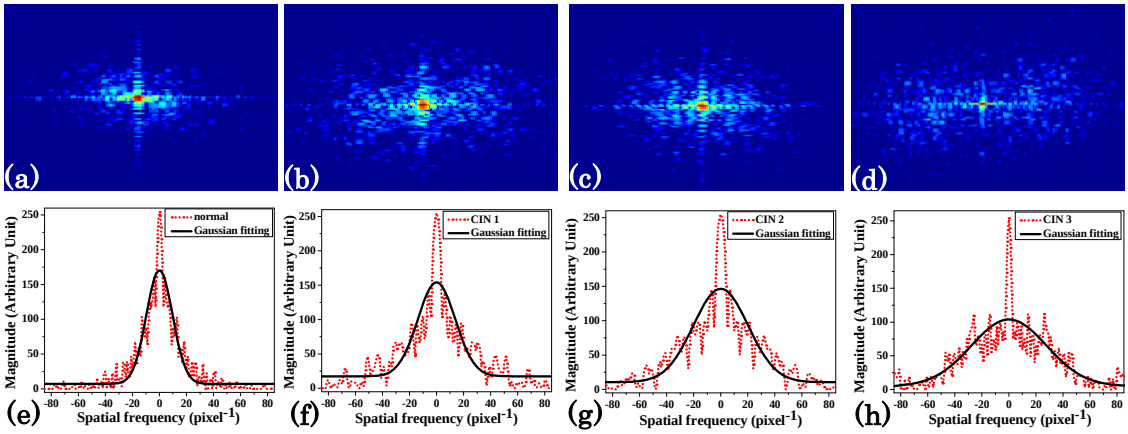


Fig. 2. 2D spatial frequency of (a) normal, (b) CIN 1, (c) CIN 2, and (d) CIN 3 tissues obtained from their corresponding confocal microscope images shown in Figs 1; Spatial frequency (dot line) and their corresponding fitted Gaussian distributions (solid line) at the most dominant cross section along horizontal direction of (e) normal, (f) CIN 1, (g) CIN 2, and (h) CIN 3.

These differences among the different types of tissues can be more distinctly seen from their spatial frequency distributions at the same pixel row along horizontal direction. Figs. 2 (e), (f), (g), and (h) show the digital spatial cross section frequency distributions of the Figs. 2 (a), (b), (c), and (d), respectively using dotted lines. The spatial frequency obtained by Fourier transform of different types of images shows that higher the grade of CIN tissue, more and wider is the spatial frequency range. This observation is in good agreement with the cervical precancer development [2].

CIN can start in any of the three stages, and can either progress, or regress [1, 2]. CIN 1 is the least risky type, confined to the basal 1/3 of the epithelium; CIN 2 is the moderate neoplasia confined to the basal 2/3 of the epithelium; and CIN 3 is the most severe, involving almost the entire thickness [1, 2]. The lesion of CIN 3 is sometimes referred to as cervical carcinoma *in situ* [1]. The patterns of normal and low grade CIN tissues consist of evenly placed uniform epithelial cells supported by a well-structured surrounding extracellular matrix (ECM), which is composed mainly by collagen [1, 2]. As grade advances, the tumor cells proliferate, thus degrading ECM and cause loss and randomness of collagen fibers [1, 2]. The images shown in Fig.1 taken from the stromal region of cervical tissue, show the collagen in the normal tissue to be more ordered in layers and uniform in shape and size while those in CIN precancer tissues are aperiodic random, anti-symmetrical, different sizes, and disordered in structure. The randomness in size and shape gives rise to wider spatial frequency range in the higher grades in comparison with lower grade CIN and normal cervical tissues. Such features of disorder, aperiodic random, anti-symmetrical collagens in different size are also hallmarks of prostate [4, 5], breast [6, 7] and other types of dysplasia and/or tumor.

To create statistical criteria to distinguish the different types of tissue, the width of the spatial frequency spectra needs to be quantified. The data shown as the dotted line in Figs. 2 (e) to 2 (h) were then fitted by the Gaussian distribution function with zero mean:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right), \qquad (4)$$

where μ the mean or expectation of the distribution (which is zero in all of our cases), σ is the standard deviation which is the diversity or "dispersion" of the data away from the mean. A lower σ indicates that the contributing frequencies are close to the mean or zero frequency, whereas higher σ suggests that the data are spread over a larger range of frequencies. The fitting curves obtained by a built-in function of Origin 8.5 are shown as solid lines in Figs. 2 (e) to 2 (d).

The differences in σ among different types of tissues may be used to evaluate different risk levels, which are reflected by their spatial frequencies. Figure 3 (a) shows extracted σ of normal (delta), CIN 1(inversed delta), CIN 2 (circle), and CIN 3 (square) tissues, respectively, using the experimental data to fit Gaussian function of equation (4). An increase of spatial frequency range from normal to CIN tissues, and from low grade CIN to high grade CIN tissues is seen in Fig 3 (a). In spectral analysis, the expanding spectral range refers to more frequency components. This phenomenon is also render as "whitening of signal" [8]. The whiter spatial frequency spectra of CIN tissue compared with normal tissue maybe provide a diagnostic criterion for grading CIN tissues.

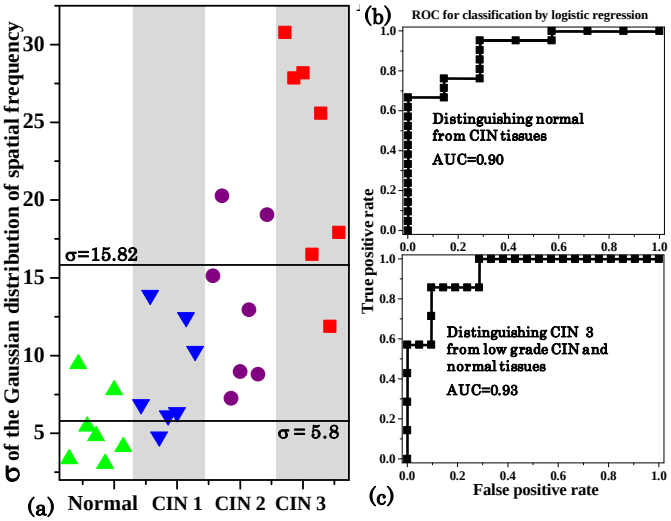


Fig. 3 (a) The extracted σ of the Gaussian distribution for the spatial frequency of total seven sets of normal, CIN 1, 2, and 3 tissues. The separating lines were calculated using SVM. Accuracy evaluated by ROC curve using the extracted σ for distinguishing significant criterion to classify the tissues into two groups for (a) normal vs. three types of CIN tissues; and (c) normal, CIN 1 and CIN 2 vs. CIN 3 tissues.

It is important to provide pathologist a simple but efficient way in form of quantitative criteria to differentiate the normal from the diseased tissues, and discriminate CIN3 from other three types of tissues. The success of the first issue decreases the false positive rate to avoid the overtreatment while the second aim meets the challenges for precancer detection and helps discriminate potential aggressive cancer from indolent disease. In order to evaluate the potential use of extracted σ as a criterion for these goals, the Support Vector Machine (SVM) [9] is used to analyze the alterations of σ reflected in the changes of normal to different grades of CIN tissues.

SVM is well-known one of the most powerful classifiers. In general, the SVM classifier is determined by a number of components for the most effectively discriminating, the support vectors, located at the boundary of the groups of data [9]. Since the data of σ to be categorized is just one dimensional in our case, it may be better to render as Support Component Machine (SCM) and Support Component (SC). In our study for separating normal and CIN tissues, the SCs are chosen from the components of σ between $\sigma_{min}^{CIN} - k$ and $\sigma_{max}^{N} + k$, where k is a self-defined threshold value for optimum, and σ_{min}^{CIN} and σ_{max}^{N} is the minimal σ for CIN tissues, and maximal σ for normal tissues. Same method was applied for chosen SCs for grouping CIN 3 vs. non CIN 3 tissues.

In a diagnosis test, the outcome may be positive (disease) or negative (healthy), which can be either true or false. To evaluate the potential of a diagnosis method using σ, the criteria for categorizing the true or false positive and negative groups in our study were determined using the SCM algorithm on all 28 (7 sets for 4 types of tissues) tissue sections. It was found that σ = 5.80 and σ = 15.82 for separating normal vs. CIN tissues, and non CIN 3 vs. CIN 3 tissues, respectively, which is shown as solid lines in Fig. 3 (a). According to these two separating lines, the sensitivity and specificity for these two cases can be calculated.

Subsequently, the receiver operating characteristic (ROC) curves were generated to evaluate the performance of criterion of the extracted σ of the spatial frequency combined with SCM for grading CIN tissues. Accuracy can be measured by the AUC (area under the ROC curve). ROC curves shown in Figs. 3 (b) and 3 (c) were generated from the cases of normal vs. CIN tissues, and non CIN 3 vs. CIN 3 tissues, respectively, to determine the accuracy of classification using σ of the spatial frequency combined with SCM. The AUC values of the ROC curves shown in Figs. 3(b) and 3(c) were then calculated.

The sensitivity, specificity and the AUC values for using the extracted σ of the spatial frequency combined with SCM for grading CIN tissues are summarized in Table I:

Table I: Evaluation of performance for criterion using σ combined with SCM for separating CIN tissues

evaluated components	sensitivity	specificity	AUC
normal vs. CIN	95.2%	71.4%	0.90
non CIN 3 vs. CIN 3	85.7%	90.5%	0.93

It can seen from Table I that the good sensitivity, specificity, and AUC values demonstrate the excellent efficacy of the spatial frequency combined with SCM as a promising diagnostic tool for CIN detection [7].

Moreover, the increase of σ of the Gaussian distribution of the spatial frequency from normal to CIN tissues, and from lower grade of CIN to higher grade of CIN tissues may provide an alternate diagnostic criterion for grading CIN tissues. In order to evaluate this potential, Fig. 4 shows the extracted σ (square) of the spatial frequency as a function of normal and CIN grade. It is important to note that the σ exhibits a monotonous growth and a good correlation with the CIN grades. This linear dependent property can be schematically shown as the solid line in Fig. 4, which can be characterized by correlation coefficient: $r^2 = 0.91$ using linear regression analysis by taking normal, CIN 1, CIN 2, and CIN 3 as 0, 1, 2, and 3, respectively. The linear increase of the extracted σ as the function of CIN grades actually reflects more spatial frequency components contained in the higher grades of CIN tissues in comparison with lower grade CIN and normal cervical tissues. Identifying the metastatic potential of cancer is critical in cancer detection and staging. This study is highly relevant and serves as a key step to predict the metastatic potential of cervical cancer using spatial frequency analysis [10]. This research will help in understanding the use of native fluorescence to identify potential metastatic cervical cancer using intact tissue specimens. Based on this study, a new optical technique may have potential to be developed as a classifier to determine tumor stages.

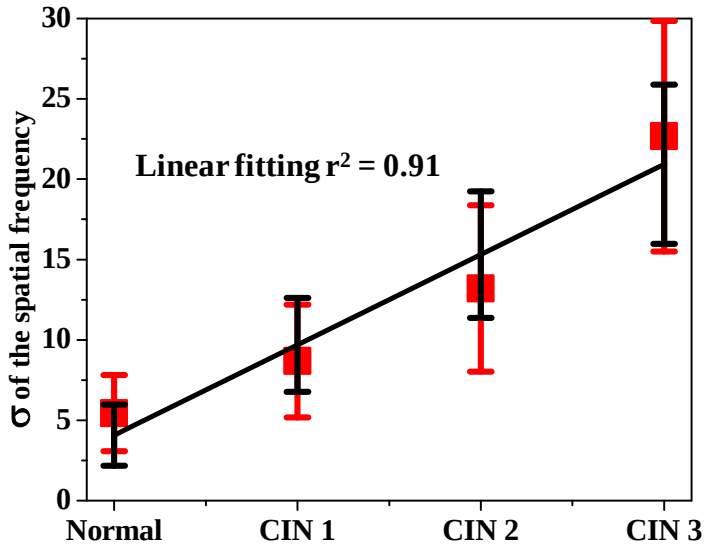


Fig. 4. The increase of σ as a function of normal and CIN grade.

4. CONCLUSION

In conclusion, this preliminary investigation on human normal, CIN 1, CIN 2, and CIN 3 cervical tissues applies Fourier analysis of their confocal microscope images to obtain information of the 2D spatial frequency spectra of these samples. With the evolution from normal to CIN tissues, and the development from low grade to high grade CIN tissues, an increase of σ of the Gaussian distribution of the spatial frequency from normal to CIN tissues, and from lower grade of CIN to higher grade of CIN tissues was observed [5]. This can be understood by more ordered layers and uniform collagen fibers in the normal and low grade CIN tissue, but aperiodic random, anti-symmetrical, different sizes, and disordered structure of collagen in high grade of CIN tissue. Differentiation between normal vs. CIN tissues and non CIN 3 vs. CIN 3 tissues can be highlighted using spatial frequency spectra analysis combined with the SCM. ROC curves were used to evaluate the performance of this criterion. Excellent sensitivity, specificity, and AUC values were achieved. This approach of using spatial frequency could discriminate the normal and three grades CIN tissues *in vitro* and in the future could be useful for *in vivo* detection. Further, based on σ of the Gaussian distribution of the spatial frequency as a function of CIN grade, a spatial frequency grading in parallel with CIN grading could be established with the linear fit in excess of 0.91. This preliminary study presents a potential criterion to diagnose early stage cervical tumor and other types of tumor or dysplasia having tissue structural changes [4-7]; therefore, spatial frequency spectra offers a simple and efficient method in optical biopsy and pathology.

This research is supported in part by U. S. Army Medical Research and Material Command (USAMRMC) under grant of # W81XWH-11-1-0335, and CARE project of Indian Institute of Technology (IIT) at Kanpur.

REFERENCES

[1] Adekunle, O. O., "Cervical Intraepithelial Neoplasia (CIN) (Squamous Dysplasia)", chapter 12 of *Intraepithelial Neoplasia***,** Edited by Dr. Supriya Srivastava, ISBN 978-953-307-987-5, Published in February, (2012).

[2] Hoffman, B., Schorge, J., Schaffer, J., Halvorson, L., Bradshaw, K., Cunningham, F., "Williams Gynecology, Second Edition," McGraw-Hill Professional, ISBN 978-0-07-171672-7, Copyright @ (2012).

[3] Goodman, J., "Introduction to Fourier Optics, Third Edition," *Roberts & Company*, ISBN 0-9747077-2-4, Copyright @ (2005)**.**

[4] Gleason D. F., Mellinger, G. T., "Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging," *J. Urol*, 111, 58-64 (1974).

[5] Pu, Y., Wang, W. B., Yang, Y. L. and Alfano, R. R., "Stokes Shift Spectroscopic analysis of multi-fluorophores for human cancer detection in breast and prostate tissues," J. Biomed. Opt. 18(1), 017005-1-8 (2013).

[6] Bloom, H. J., G., Richardson, W. W. W., "Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years," Br. J. Cancer. 11(3), 359–377 (1957).

[7] Pu, Y., Wang, W. B., Yang, Y. L. and Alfano, R. R., "Native fluorescence spectra of human cancerous and normal breast tissues analyzed with nonnegative constraint methods," Appl. Opt. 52(6), 1293-1301 (2013).

[8] Efford, N., "Digital Image Processing: A Practical Introduction Using Java^{TM}," Pearson Education, Addison-Wesley, ISBN-13: 9780201596236, Copyright © (2000)

[9] Pu, Y., Jagtap, J., Pradhan, A., and Alfano, R. R., "Optical Spatial Frequency Quantitative Pathology of Cervical Intraepithelial Neoplasia in Human Tissues," accepted by J. Biophotonics. (2014).

[10] Laughney, A. M., Krishnaswamy, V., Rice, T. B., Cuccia, D. J., Barth, R. J., Tromberg, B. J., Paulsen, K. D., Pogue, B. W., Wells, W. A., "System analysis of spatial frequency domain imaging for quantitative mapping of surgically resected breast tissues," J. Biomed. Opt. 18(3), 036012-1-11 (2013).