International Journal of Advancements in Electronics and Electrical Engineering– IJAEEE
Volume 4 : Issue 2 [ISSN : 2319-7498]
Publication Date : 30 October, 2015

DNA Sequence Matching Technique based on Optical Correlation

A. K. Alqallaf and A. K. Cherri

***Abstract*— We present a new approach, which is capable of providing real-time processing of DNA sequences using an optical joint Fourier transform correlator (JTC). The DNA sequences are treated as 2-D images instead of numeric data. Simulation experiments are employed to demonstrate the effectiveness of the proposed approach to search for and to obtain locally and/or globally similarity/dissimilarity between the two DNA sequences.**

***Keywords*— DNA sequencing, data visualization, joint fourier transform optical correlation**

I. Introduction

Biologists are looking for fast real-time processing techniques to analyze the huge amounts of data in the form of DNA sequences. In this regard, various computer-based algorithms have been used to map and detect the similarities and/or dissimilarities across the tested DNA sequences such as local and global alignments, and multiple-sequence alignments where symbolic structures of the DNA sequences as string, tree, and graph are used [1-3].

DNA sequences are usually treated as the symbolic data A, G, C, and T corresponding to the four types of nucleic acids: Adenine, Guanine, Cytosine, Thymine, respectively. Alternatively, the structures of genomic sequences can also be represented by numeric form where well-known digital signal processing techniques (such as discrete Fourier transform or wavelet transform [4-6]) are developed to analyze the numeric signals for many applications [7-9]. For instance, hypercomplex signals which are considered as the general form of quaternion signals [10-12] are used for pairwise alignment and determining the cross-correlation of DNA sequences to find both the global and local matching between two DNA sequences based on cross-correlation algorithm.

On the other hand, the optical processing techniques, which are well established in many fields of research and technology, present solutions over the conventional digital techniques for processing biometric features [13-17]. For instance, optical correlator approaches are used for various information processing applications. The 2-D processing capability, the fast processing speed, and the non-interfering communication of optical correlators, provide fast real-time pattern recognition. In this regard, the joint Fourier transform correlator (JTC) has been found to be inherently suitable for real-time applications and it has been widely investigated [18-22]. In this paper, we propose a framework based on optical joint Fourier transform correlation to efficiently analyze and explore the genetic data. First, the examined DNA sequences are prepared as 2-D images for optical processing. Then, these images are processed to detect local and global matching among the examined sequences. The effectiveness of the proposed method is verified and demonstrated by simulation experiments.

A. K. Alqallaf and A. K. Cherri
College of Engineering and Petroleum; Electrical Engineering Department;
Kuwait University; P. O. Box 5969; Safat 13060,
Kuwait;

II. Optical Correlation Processing For DNA Sequence Matching

In optical pattern recognition, Fourier transform based information processing is widely used [23, 24]. In particular, the joint Fourier transform correlator (JTC) is an attractive technique for real-time optical processing [22-25] due to the fact the JTC does not require a complex filter in the Fourier plane as well as meticulous alignment requirement when compared to the classical VanderLugt correlator. Improvement JTC-based techniques to enhance some features of the correlation in the output plane are found in literatures [18, 20, 22] such as the binary JTC (BJTC) and the fringe-adjusted JTC (FJTC). Unlike BJTC, the FJTC does not produce higher-order harmonic correlation peaks in the output plane. However, for multiple target detection, the FJTC cannot eliminate the false alarms for input scenes containing multiple identical targets as well as multiple non-target objects. However, a multi-step JTC using Fourier plane image subtraction technique was proposed to suppress the zero-order term and false alarms generated by identical objects in the input scene [20]. This technique is used in this paper to illustrate the usefulness of the proposed method for DNA sequences.

A classical joint transform correlator is shown in **Fig. 1**, where the input joint image is displayed at the input plane of a spatial light modulator (SLM) and may be expressed as

$$f(y) = r(y + y_0) + s(y - y_0) \quad \text{(1a)}$$

$$s(y - y_0) = \sum_{i=1}^{n} t_i(y - y_i) + \sum_{i=1}^{n} n_i(y - y_i) \quad \text{(1b)}$$

where $r(y)$ represents the reference pattern and $s(y)$ represents the unknown input scene that may contain multiple target patterns denoted by $\sum_{i=1}^{n} t_i(y - y_i)$ and multiple identical

non-target patterns denoted by $\sum_{i=1}^{n} n_i(y-y_i)$. Note that one dimensional notation is used for simplicity. Applying Fourier transform to the input joint image of Eq. (1), we get

$$F(v)=|R(v)|e^{j[\phi_r(v)+vy_0]}+|S(v)|e^{j[\phi_s(v)-vy_0]} \quad (2a)$$

$$S(v)e^{-jvy_0}=|S(v)|e^{j[\phi_s(v)-vy_0]} =\sum_{i=1}^{n}|T_i(v)|e^{j[\phi_{ti}(v)-vy_i]}+\sum_{i=1}^{n}|N_i(v)|e^{j[\phi_{ni}(v)-vy_i]} \quad (2b)$$

where $|R(v)|$, $|S(v)|$, $|T_i(v)|$, and $|N_i(v)|$, are the amplitudes, and $\phi_r(v)$, $\phi_s(v)$, $\phi_{ti}(v)$, and $\phi_{ni}(v)$ are the phases of the Fourier transforms of $r(y)$, $s(y)$, $t_i(y)$, and $n_i(y)$, respectively; v is a frequency domain variable scaled by a factor $2\pi/\lambda f$, λ is the wavelength of collimating light, f is the focal length of lenses L1 and L2 [19].

A square-law detector, placed in the back focal plane of L1, produces the intensity or the joint power spectrum (JPS) corresponding to Eq. (2):

$$|F(v)|^2=F(v)F^*(v)=|R(v)|^2+|S(v)|^2+2|R(v)||S(v)|\cos[\phi_r(v)-\phi_s(v)+2vy_0] \quad (3a)$$

$$|S(v)|^2=S(v)S^*(v)=\sum_{i=1}^{n}|T_i(v)|^2+\sum_{i=1}^{n}|N_i(v)|^2 +2\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}|T_i(v)||T_k(v)|\cos[\phi_{ti}(v)-\phi_{tk}(v)-vy_i+vy_k] +2\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}|N_i(v)||N_k(v)|\cos[\phi_{ni}(v)-\phi_{nk}(v)-vy_i+vy_k] +2\sum_{i=1}^{n}\sum_{k=1}^{n}|T_i(v)||N_k(v)|\cos[\phi_{ti}(v)-\phi_{nk}(v)-vy_i+vy_k] \quad (3b)$$

Eq. (3) contains the desired cross-correlation between the reference and the targets, and other correlation produced due to the autocorrelation between the identical input scene targets and due to the autocorrelation between the identical non-targets themselves. The JPS is displayed on the second SLM and then Lens L2 of **Fig. 1** performs an inverse Fourier transform. The correlation results appear on the output plane and they are detected by a square-law device. False correlation peaks are expected to appear in the output plane as it will be shown later.

From (3), it is evident that the correlation output will consist of a pair of autocorrelation peaks corresponding to each target in the input scene, a zero-order terms, cross-correlation peaks between the reference and the non-target objects that may be present in the input scene, and false alarms produced due to the autocorrelation between identical targets and identical non-targets present in the input scene.

A Fourier plane image subtraction technique can be used to eliminate the false alarms that are generated automatically in a JTC when multiple targets or identical non-target objects are present in an input scene [20]. In this technique, both the input-scene-only power spectrum ($|S(v)|^2$) and the reference-image-only power spectrum ($|R(v)|^2$) are subtracted from the JPS of Eq. (3). Thus, the modified JPS is:

$$P(v)=|F(v)|^2-|R(v)|^2-|S(v)|^2 =2|R(v)||S(v)|\cos[\phi_r(v)-\phi_s(v)+2vy_0] \quad (4)$$

An inverse Fourier transform of this modified JPS yields the correlation output:

$$\gamma_{ff}(x',y')=\gamma_{rs}(x',y')*\delta(x',y'-2y_o) +\gamma_{sr}(x',y')*\delta(x',y'+2y_o) \quad (5)$$

Eq. (5) shows that the subtraction operation eliminates the zero-order terms, the false alarms generated by similar input scene targets and identical non-targets, and the cross-correlation terms between other objects that may be present in the input scene.

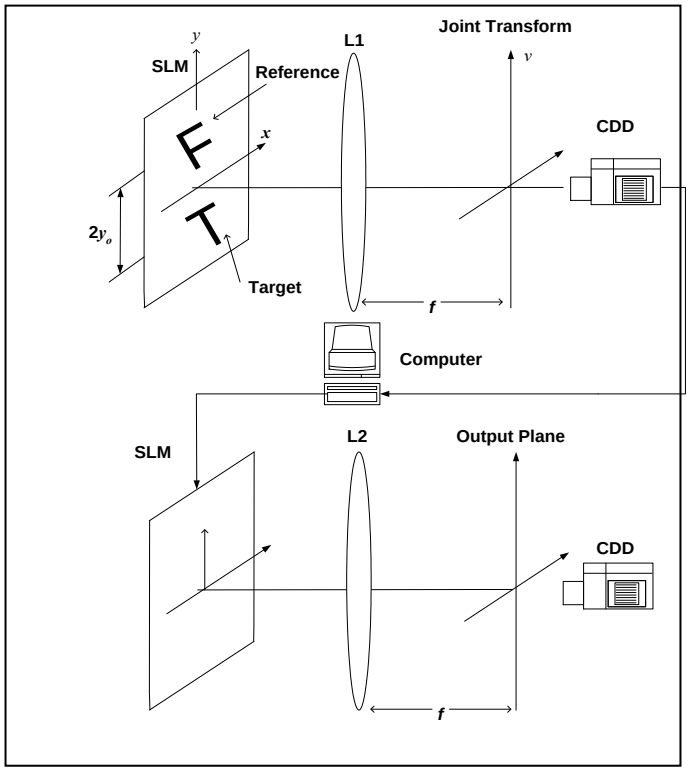
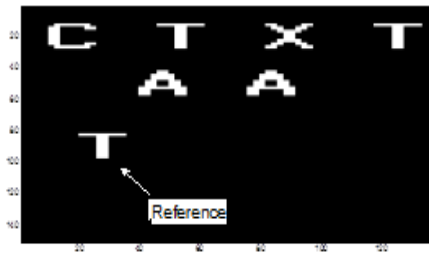


Figure 1. A classical JTC architecture.

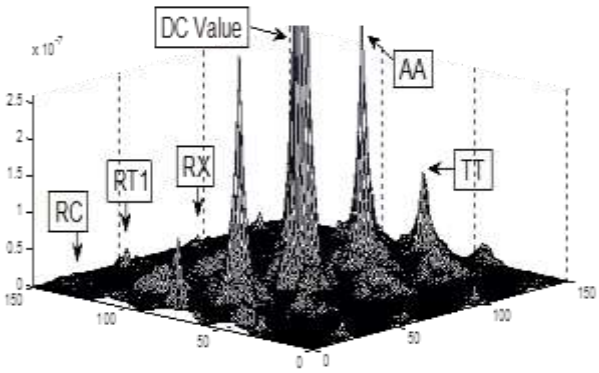
Fig. 2 illustrates the JTC technique. **Fig. 2(a)** is a joint image that contains a reference letter **T** in the lower part of the joint image; whereas the upper part has an input scene that contains multiple target letters (**T**), multiple similar non-target letters (**A**), and other non-target letters (**C** and **X**). **Fig. 2(b)** illustrates the correlation output produced by a classical JTC. From this figure, the correlation peaks values of the targets are much lower than the other correlation peaks (for instance for

AA or **TT** in the input seen). In addition the DC value dominates all the correlation peaks values in the output plane. **Fig. 2(c)** show how the Fourier plane image subtraction technique eliminates the correlation peaks for the targets letter (letter **T**), the non-targets letter (letter **A**) as well as the DC values peaks within the input seen image.

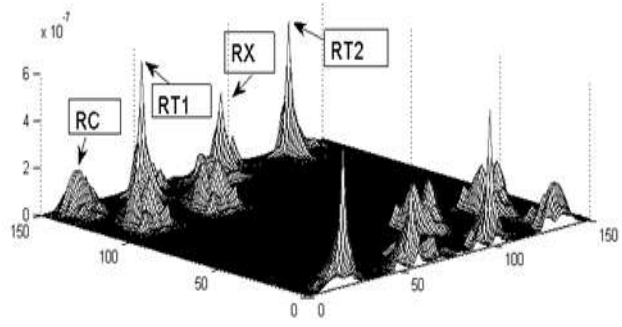
Next, the symbolic structure representation of DNA sequences are prepared for optical processing in the JTC set-up. In this regard, the symbolic structure of the DNA strings are text-to-image converted into optical images.



(a)



(b)



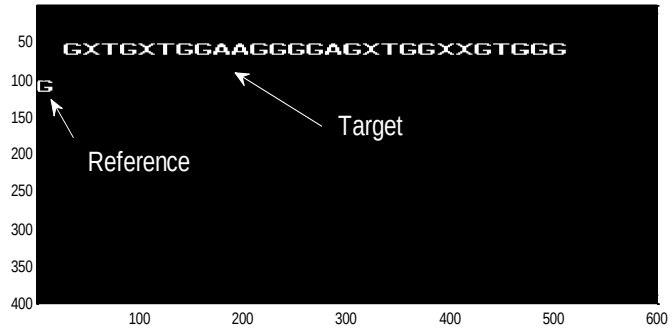
(c)

Figure 2. (a) Joint input image; (b) classical correlation output; and (c) correlation output after using subtraction technique.

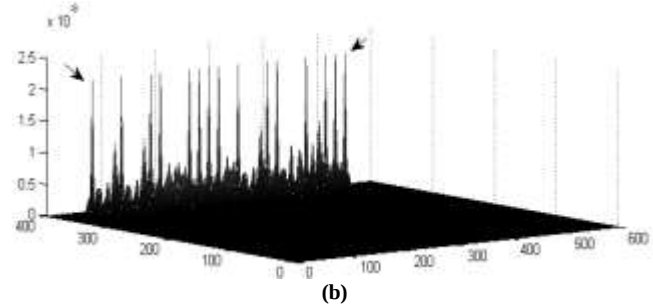
III. Experimental Results

A typical DNA sequence has multiple appearances of the symbols letters 'A', 'C', 'G', and 'T'. It is clear that the JTC is well-suited to process this type of images. Note that the letters 'C' and 'G' resemble each other and the cross-correlation between them produces very close results and may cause misses in the recognition process. For this reason, in any DNA sequence, the letter 'C' is replaced with the letter 'X' to make the discrimination process more successful.

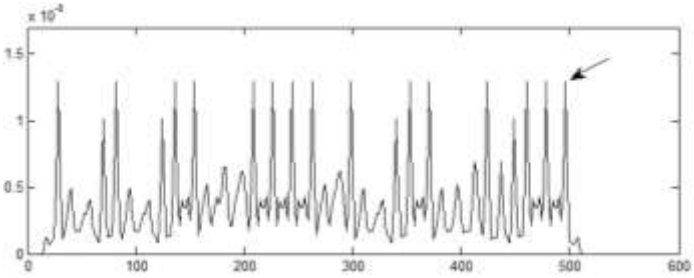
To examine the performance of the proposed method, DNA sequences are employed in the following three experiments. The first experiment illustrates the detection of single DNA element in a given sequence. Regardless of how many elements in the sequence, the JTC detects their presences and their locations as well their counts. **Fig. 3** shows the results of this experiment. **Fig. 3a** is the input joint image and the target is the letter 'G'. **Fig. 3b** is the correlation output showing the correlation peaks where the letter 'G' is detected. Note that in **Fig. 3b** only half of the correlation spots are displayed since the other half is a duplicate of the first half. **Fig. 3c** illustrates a trace line across the correlation output to easily visualize the correlation peaks. Note that the count of the letter 'G' is fifteen as expected.



(a)



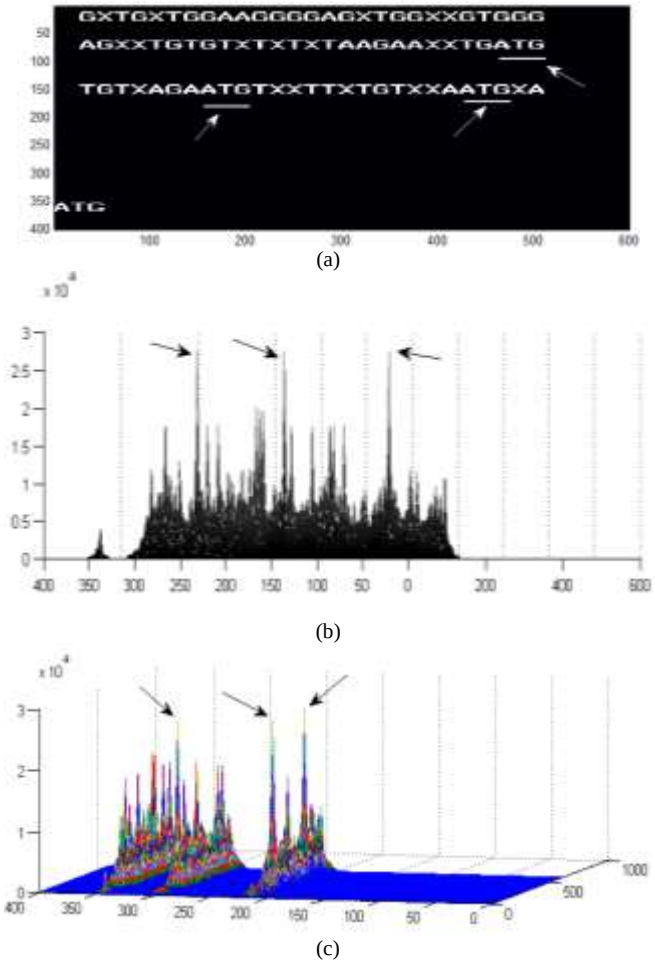
(b)



(c)

Figure 3. Single-detection: detection of DNA element (nucleotide) "G" in a given sequence. (a) DNA sequence and the target letter; (b) the correlation output; and (c) 2-D trace curve of the correlation output.

The results of the second experiment are shown in **Fig. 4**. **Fig. 4b** presents the detection of the start codon 'ATG' within a DNA sequence as indicated in **Fig. 4a**. The start codon 'ATG' was successfully detected in three different locations as pointed out by the arrows in **Figs. 4b** and **4c**. The third experiment demonstrates how a whole sequence can be easily detected by the optical correlation as shown in **Fig. 5.**



(a)

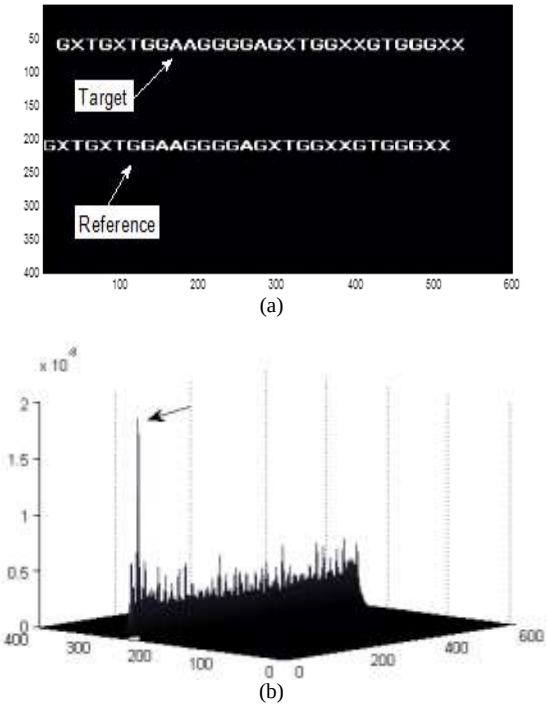
(b)

(c)

Figure 4. The detection of the start codon (**ATG**); (a) DNA sequence; (b) the correlation output; (c)Top view of the location of the start codon.

I. Conclusion

In this paper, we have proposed a real-time DNA sequencing analysis by using the well-known joint Fourier transform correlation (JTC). The DNA sequences are first converted to 2-D images, which are now considered as target and reference images for the optical JTC processor. Simulation experiments demonstrated that the proposed optical approach is suitable to locate similarities in addition to other important information between the studied DNA sequences.



(a)

(b)

Figure 5. Self-detection: (a) target and reference DNA sequences are the same; and (b) the correlation output.

References

[1] C. Zhang and A.K. Wong, "A genetic algorithm for multiple molecular sequence alignment," Comput. Appl. Biosci., vol. 13, pp. 565–581, 1997.

[2] W. Choe, O.K. Ersoy, and M. Bina, "Neural network schemes for detecting rare events in human genomic DNA," Bioinformatics, vol. 16, pp. 1062–1072, 2000.

[3] L. Hirschman, J.C. Park, J. Tsujii, L. Wong, and C.H. Wu, "Accomplishments and challenges in literature data mining for biology," Bioinformatics, vol. 18, pp. 1553–1561, 2002.

[4] W. Wang and D.H. Johnson, "Computing linear transforms of symbolic signals," IEEE Trans. on Signal Processing, vol. 50, pp. 628–634, 2002.

[5] S. Tiwari, S. Ramachandran, A. Bhattacharya, S. Bhattacharya, and R. Ramaswamy, "Prediction of probable genes by Fourier analysis of genomic sequences," Computer Applications in the Biosciences, vol. 13, pp 263–270, 1997.

[6] D. Anastassiou, "Frequency-domain analysis of biomolecular sequences," Bioinformatics, vol. 16, pp. 1073–1081, 2000.

[7] D. Anastassiou, "Genomic signal processing," Signal Processing Magazine, vol. 18, pp. 8–20, 2001.

[8] J. Astola, E. Dougherty, I. Shmulevich, and I. Tabus, "Genomic signal processing," Signal Processing, vol. 83, pp. 691–694, 2003.

[9] A.V. Oppenheim, R.W. Schafer, and J.R. Buck, Discrete-Time Signal Processing, 2nd ed., Chapter 1, Prentice Hall International, Inc., 1999.

[10] I. L. Kantor and A. S. Solodovnikov, Hypercomplex Number: An Elementary Introduction to Algebras, New York: Springer-Verlag, 1989.

[11] H. T. Chang, "DNA sequence visualization," Chapter 4 in Advanced Data Mining Technologies in Bioinformatics, pp. 63–84, Edited by Dr. H.-H. Hsu, Idea Group, Inc., 2006.

[12] H. T. Chang, Chung J. Kuo, Neng-Wen Lo, and Wei-Z. Lv, "DNA sequence representation and comparison based on quaternion number system," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 3, 2012.

[13] F. Tsalakanidou, S. Malassiotis, M. G. Strintzis, "A 3D face and hand biometric system for robust user-friendly authentication," Pattern Recognition Letters, vol. 28, pp. 2238-2249, 2007.

[14] B. Da and N. Sang, "Local binary pattern based face recognition by estimation of facial distinctive information distribution," Opt. Eng. 2009; 48(11):117203. doi: 10.1117/1.3258349

[15] S. Zhang and X. Gu, "Palm recognition method based on score level fusion", Optik, vol. 124, pp. 3340-3344, 2013.

[16] K. L. Rice, T. M. Taha, A. M. Chowdhury, A. A. S. Awwal, and D. L. Woodard, "Design and acceleration of phase-only filter-based optical pattern recognition for fingerprint identification", Opt. Eng., vol. 48(11), 2009; 117206. doi: 10.1117/1.3265549

[17] M. Jedynski, K. Chalasinska-Macukow, "Wavelet transform for preprocessing in an optical correlator with multilevel composite filter," Opt. Eng., vol. 43, pp. 1759-1766, 2004.

[18] M. S. Alam and M. A. Karim, "Multiple target detection using a modified fringe-adjusted joint transform correlator," Opt. Eng., vol. 33**,** pp. 1610-1617, 1994.

[19] H. A. Kamal and A. K. Cherri, "Complementary-reference and complementary-scene for real-time fingerprint verification using joint transform correlator," Optics & Laser Technology, vol. 41, pp. 643-650, 2009.

[20] M. Schonleber, G. Cedilnik, and H. J. Tisiani, "Joint transform correlator subtracting a modified Fourier spectrum," Appl. Opt., vol. 32, pp. 7532-7537, 1995.

[21] F. T. S. Yu, F. Cheng, T. Nagata and D. A. Gregory, "Effects of fringe binarization of multi-object joint transform correlation," Appl. Opt., vol. 28, pp. 2988-2990, 1989.

[22] [22] B. Javidi, J. Li, A. H. Fazlollahi, and J. Horner, "Binary nonlinear joint transform correlator performance with different thresholding methods under unknown illumination," Appl. Opt., vol. 24, pp. 886-896, 1995.

[23] J. W. Goodman. Introduction to Fourier Optics. New York, NY: McGraw-Hill Book Co.: 1968, pp 181, 203-208.

[24] A. VanderLugt, "Signal detection by complex spatial filtering," IEEE Trans. Inf. Th., **IT-10**, p. 139, 1964.

[25] M. S. Alam, M. Akhteruzzamana, A. K. Cherri, "Real-time fingerprint identification," Optics & Laser Technology, vol. 36, pp. 191-196, 2004.

Abdullah Alqallaf

Dr. Abdullah Alqallaf is an assistant professor with the Department of Electrical Engineering at the Kuwait University. He received his BS and MS degrees in electrical engineering from Kuwait University in 1996 and 1999 respectively, and his Ph.D. degree in Electrical Engineering from University of Minnesota - Twin Cities in 2009. Dr. Alqallafs research interests are Microwave Imaging Techniques and Analysis for Tumor Detection- Genomics Signal Processing and Bioinformatics- Statistical and Wavelet Signal Processing Machine Learning for Signal Processing- Design and Implementation of Signal Processing- Image and Multidimensional Signal Processing- Medical image analysis/Feature extraction, Feature selection, Segmentation, Detection/Estimation, and Classification.

Abdallah Cherri

Dr. Abdallah Cherri is a professor with the Department of Electrical Engineering at the Kuwait University. He received his BS and MS degrees in electrical engineering from the Wichita State University (Wichita, Kansas) in 1984 and 1985 respectively, and his Ph.D. degree in electrical engineering (with specialization in optical computing) from the University of Dayton (Dayton, Ohio) in 1989. Dr. Cherri's research interests are optical computing, display devices, optical pattern recognition, digital/optical image and signal processing, and electrooptics devices. He has published numerous journal and conference proceedings papers.