

An Adaptive CAD System to Detect Microcalcification in Compressed Mammogram Images

Ayman AbuBaker

Electrical and Computer Engineering Dept.
Applied Science Private University
Amman, Jordan

Abstract—Microcalcifications (MC) in mammogram images are an early sign for breast cancer and their early detection is vital to improve its prognosis. Since MC appears as small dot in the mammogram image with size less than 1 mm and maybe easily overlooked by the radiologist, the Computer Aided Diagnosis (CAD) approach can assist the radiologist to improve their diagnostic accuracy. On the other hand, the mammogram images are a high resolution image with large image size which makes difficult the image transfer through the media. Therefore, in this paper, two image compressions techniques which are Discrete Cosine Transform (DCT) with entropy coding and Singular Value Decomposition (SVD) were investigated to reduce the mammogram image size. Then a novel adaptive CAD system is used to test the quality of the processed image based on true positive (TP) ratio and number of detected false positive (FP) regions in the mammogram image. The proposed adaptive CAD system used the visual appearance of MC in the mammogram to detect a potential MC regions. Then five texture features are implemented to reduce number of detected FP regions in the mammogram images. After implementing the adaptive CAD system on 100 mammogram images from USF and MIAS databases, it was found that the DCT can reduce the image size with a high quality since the ratio of TP is 87.6% with 11 FP/regions while in SVD the TP ratio is 79.1% with 26 FP/regions.

Keywords—Mammogram image; texture features; Discrete Cosine Transform (DCT); Singular Value Decomposition (SVD)

I. INTRODUCTION

Nowadays, breast cancer has become the second leading cause of cancer deaths in women after lung cancer [1]. In 2007, about 40460 women and 450 men died from breast cancer in the United States [1]. Breast cancer's death rates continue to decline primarily in young women less than 50 years old [1], [2]. These declines are believed to be the result of earlier detection, through screening and increased awareness, as well as improved treatment.

Radiologist used three methods to diagnose breast cancer tumors: mammography, fine needle aspirate and surgical biopsy. The Mammography is specialized medical imaging that uses a low-dose x-ray system to see inside the breasts. A mammography exam, called a mammogram, aids in the early detection and diagnosis of breast diseases in women. The malignant sensitivity of mammography has a reported between

68% and 79% [3]. In fine needle aspirate biopsy (FNAB), a flexible needle that is thinner than ones used for blood tests is employed. The needle is guided into the area of the breast change while the doctor is feeling (palpating) the lump to extract some lump fluids. These fluids extracted from the breast lump are inspected under the microscope and diagnosed as benign or malignant. The sensitivity of this method varying from 65% to 98% [3], since cancer lumps are sometimes missed or benign cells are taken from near a cancer. Surgical biopsy is more evasive and costly but it is the only test that can confirm malignancy.

Most of the researchers try to focus in detection one or more of abnormal structures in mammogram images which are microcalcifications, circumscribed masses and speculated lesions [4]. Therefore, many researchers investigate many approaches to automatically and accurately detect these tumors in the mammogram images.

Approximately 65% of women who undergo biopsy for histopathologic diagnosis as a second stage in detecting the breast cancer are found to be healthy, which was a false positive result from diagnosis mammogram images [5]. These false positives may be due to different circumstances such as poor image quality, eye fatigue, or oversight by the radiologist. Therefore, the Computer Aided Diagnosis (CAD) which is the application of computational techniques that is used to solve the problem of interpreting mammogram images [6], [7]. The CAD system is usually used as a second mammogram reader. However, the final decision regarding the likelihood of the presence of a cancer and patient management is left to the radiologist.

Microcalcification is one of the most difficult detection tumors by the radiologists since its size less than 1 mm. Therefore, mammogram images have a high resolution in order to show the microcalcification clearly for the radiologist. As a result, the digital mammogram image is one of the medical images that varies in the size between 8 MB and 55 MB based on the breast size. In this case, image compression techniques is crucial to easily process and transfer these images. At the same time, it is important to maintain the mammogram image details which are microcalcifications as in the original image.

The main objective of this paper is to evaluate the performance of the proposed adaptive CAD system that will be

implemented on two different images' compression techniques which are Discrete Cosine Transform (DCT) with entropy coding and Singular Value Decomposition (SVD) techniques. Moreover, the mammogram image quality resulted from both compression techniques will be evaluated using the adaptive CAD system based on True Positive (TP) ratio and number of detected False Positive (FP) regions.

In this paper, mammogram database that are used in this work is described in Section II. A brief literature survey is presented in Section III. An adaptive CAD system implementation on two different compression techniques is introduced in Section IV. While Section V presents the algorithm evaluations. Finally, concluding remarks are given in Section VI.

II. DATABASE

One hundred mammogram images are collected from both University of South Florida (USF) and MIAS databases. USF mammogram images are collected from different medical schools. Then they were scanned using high resolution (3000 pixel \times 4500 pixel and 16-bit pixel depth). The MIAS database also been used in this work, where 40 % of the MC mammogram images from this database. This database were collected from the United Kingdom National Breast Screening Program. The mammogram images have been expertly diagnosed and the positions of the MCs in each image are recorded.

In this paper, 60 USF and 40 MIAS MC mammogram images are used in order to be processed on the proposed algorithm.

III. LITERATURE REVIEW

Reduction image size such as mammography, multimedia and electronic publishing is a critical stage in processing such these images [7], [8]. Many techniques are available to magnify or reduce images ranging from linear interpolation to cubic spline interpolation [8]-[11].

Image interpolation is one of image reduction techniques that has a central role in many applications [12], [13]. These techniques change digital image size according to the nature of the display device. Chuah and Leou [14] presented three categories of image interpolation techniques which are static image interpolation [15], [16]; multi-frame image interpolation [11], [17]; and image sequence (video) interpolation [17].

Many authors investigate different types of interpolation techniques to reduce the image size as in Herasa, *et al.* [18]. They investigate three interpolation methods to reduce the image size which are linear, bi-cubic and the parametric spline method. Then they try to compare between them using the visual appearance which was a compromise between the previous two methods. Then a robust algorithm OPED for the reconstruction of images from Radon data using these interpolation methods were build. The results show that the performance of bi-cubic interpolation were better than other interpolation techniques showed significantly lower

Normalized Mean Square Error (NMSE) comparing with other methods.

On the other hand, other authors modified the algorithm of interpolation processing such in Kim, *et al.* [19]. They were proposed a new image scaling algorithm called the Winscale algorithm. In this algorithm a scaling (up/down) is used based on an area pixel model rather than a point pixel model. As a result, the Winscale algorithm produced effective results for image processing systems that required high visual quality and low computational complexity. However, its performance was similar to the bi-linear interpolation technique [20]. An adaptive algorithm to interpolate low resolution image frames was proposed by Chuah and Leou [14]. In this algorithm, two nonlinear filters were used to generate high-frequency components iteratively that were lost during the implementation of the resolution reduction procedure, then a blocking artefacts-reduction scheme was adopted to improve the image quality. Abe and Iiguni [21] investigated the discrete cosine transform (DCT) in down-sampled images and proposed high-resolution (HR) image restoration from a down-sampled low-resolution (LR) image using the discrete cosine transform (DCT). Their algorithm showed a superior performance compared to cubic spline interpolation in the HR image restoration as long as the amount of the additive noise was small.

Another authors used image interpolation techniques that was followed by a re-sampling process in order to reduce the image size. These techniques used in image reduction that utilizes the mean of each non-overlapping 8×8 pixel neighborhood [22]. This blurred the breast boundary since the boundary was averaged with the background. As a result, an important breast regions such as microcalcification are lost which means that there are a significant loss of the information in the original mammogram images.

IV. PROPOSED ALGORITHM

CAD system is used to identify regions where abnormalities may be found. Therefore, some researchers have suggested that a combination of computer plus radiologist can produce better results than a radiologist alone, and have concluded that CAD systems can be used to prompt radiologists [23]. Recently, some studies have suggested that a CAD system and a radiologist could be used as an alternative to using two radiologist film readers to look at each mammogram case [23], [24].

CAD systems can improve the performance of a radiologist by detection the MC in the mammogram image with minimum number of false positive regions. So, in order to have a robust CAD system, a preprocessing image size reduction is needed to speed up and easily transfer the mammogram images. Therefore, in this work, the mammogram image size is reduced using two image compression techniques which are SVD and DCT. Then an adaptive CAD system is investigated to evaluate the performance of these image compression techniques based on TP ratio and FP regions. The proposed work stages are summarized in Fig. 1.

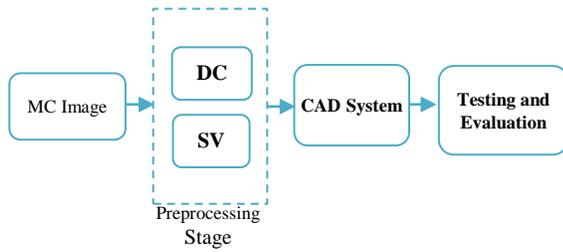


Fig. 1. The paper work flowchart.

A. Preprocessing Stage

The growth in social media and multimedia increases the need to transfer and save multimedia files. The mammogram images are part of these files that have large image size with high resolution. Compression the mammogram images is crucial to be transferred or processed but without degradation the image quality. In this paper, two types of image compression techniques were used as in [25] which are SVD and DCT.

1) Singular Value Decomposition (SVD)

SVD is one of the image compression techniques which take high data set dimensions and reduces it to fewer dimensions with retaining the original substructure of the data. In this work, the SVD technique is applied on 100 mammogram images from both USF and MIAS database as in [25]. Fig. 2, show the result of the processed mammogram images using SVD compression technique.

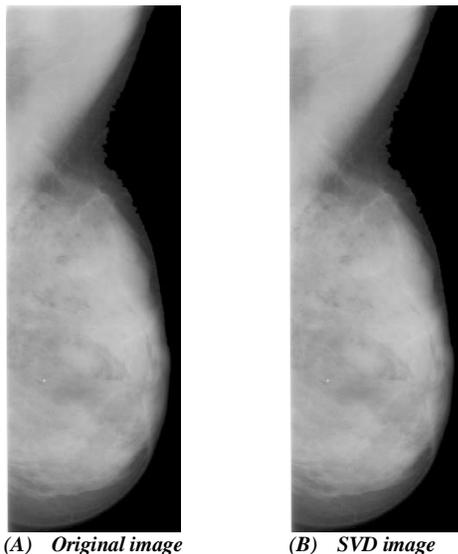


Fig. 2. Mammogram image compression using SVD.

The mammogram images were processed using SVD at different threshold value and found that the best singular value is in the range of 150-160 singular values. The maximum image compression ratio was 42% with high quality image appearance.

B. Discrete Cousin Transform (DCT)

As a DCT works in a frequency domain, the image is initially transformed from the spatial domain to frequency

domain. Then the image is separated into parts of differing frequencies. These image frequencies presented as a sum of sinusoids of varying magnitude and frequencies. DCT is high energy compact so the most significant image information is concentrated in few coefficient of DCT.

One hundred mammogram images are also processed using the DCT technique in order to compare between these compression techniques. The DCT image compression is applied as in [25]. Fig. 3 show the result of implementing the DCT compression technique on the mammogram images.

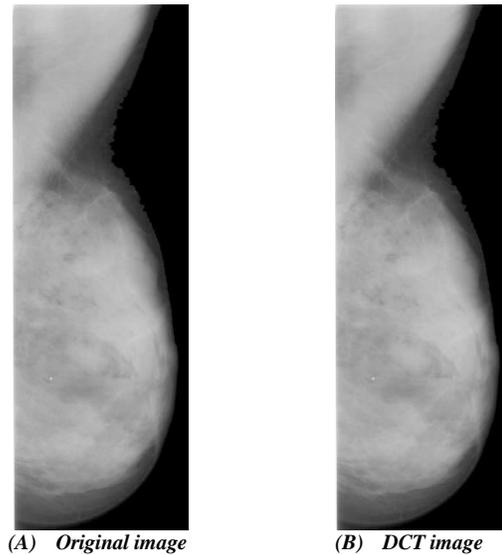


Fig. 3. Mammogram image compression using DCT.

After many trials, it was found that the average of the best threshold value is 10 where the image appearance is good. The DCT can successfully reduce the mammogram image size by 62% of the original image.

C. Proposed Adaptive CAD System

This section presents the novel adaptive CAD technique that is implemented on the two types of compressed mammogram images. The CAD system performance in detecting the MCs in mammogram images for each one of the image compression techniques will be recorded in this paper. The proposed CAD system is mainly divided into two stages. In the first stage a potential MC regions (PMR) will be detected based on the MC appearance on the mammogram images. Then, in stage 2, five texture feature will be generated to reduce number of the detected FP regions in the mammogram image consequently increase the CAD system sensitivity in detection the MC in the mammogram images. The following subsection will present the steps in detecting the MC in compressed mammogram images.

1) Detection of Potential MC Region

MCs are one of most difficult cases in diagnosis by the radiologist since they are appear as small regions, with intensity values higher than their surrounding background in mammogram images. Usually the MCs size is less than 1 mm [26]. So, the proposed CAD algorithm is designed to be sensitive in detecting MC in compressed mammogram images.

Therefore, two dynamic concentric masks are used to detect potential MC regions as shown in Fig. 4. So, the potential peak (MCs) will be when an average value of the inner mask is greater than the average of the outer mask excluding the inner mask pixels values.



Fig. 4. Two concentric masks.

In [27] the mask size for both inner and outer mask are set to be fixed based on image resolution. Where the image resolution of the USF and MIAS databases are $45 \mu\text{m} \times 45 \mu\text{m}$ and $50 \mu\text{m} \times 50 \mu\text{m}$ respectively. Based on the image resolution, an inner mask was design to be in a radius of 9 pixels. Also, the suitable outer mask is design to be in a radius of 13 pixels. These concentric masks are processed on different USF and MIAS mammogram images and the CAD system can effectively detect a potential MC regions.

In this work, the MCs size in the image will be changed corresponding to the image compression technique since the image compression change the image resolution. Therefore, the proposed CAD system is designed using two dynamic concentric masks. The masks size will be varied based on the image resolution.

In preprocessing stage, the mammogram images are processed using SVD and DCT compression techniques. So we have to datasets which are 100 SVD mammogram images and 100 DCT mammogram images. These images are individually processed using the first stage of the proposed adaptive CAD system and the MC regions are tested for each image. As a result, the algorithm can successfully detect the PMC regions in both SVD and DCT mammogram images as shown in Fig. 5.

As shown in Fig. 5, the PMC regions are detected but with large number of FP regions which will reduce the proposed CAD system sensitivity. Therefore, five texture features are generated for each detected PMC region in order to reduce number of detected FP regions as will be presented in the next section.

2) FP Reduction Algorithm

In order to have a high sensitive CAD system, number of detected FP regions should be very low. This can be achieved by using five texture features which are mean, entropy, standard deviation, moment and kurtosis. These five texture features are implemented on *actual* TP MC regions that copped manually from 100 mammogram images. As a result, a dataset of the *actual* TP features is generated. This dataset is used to

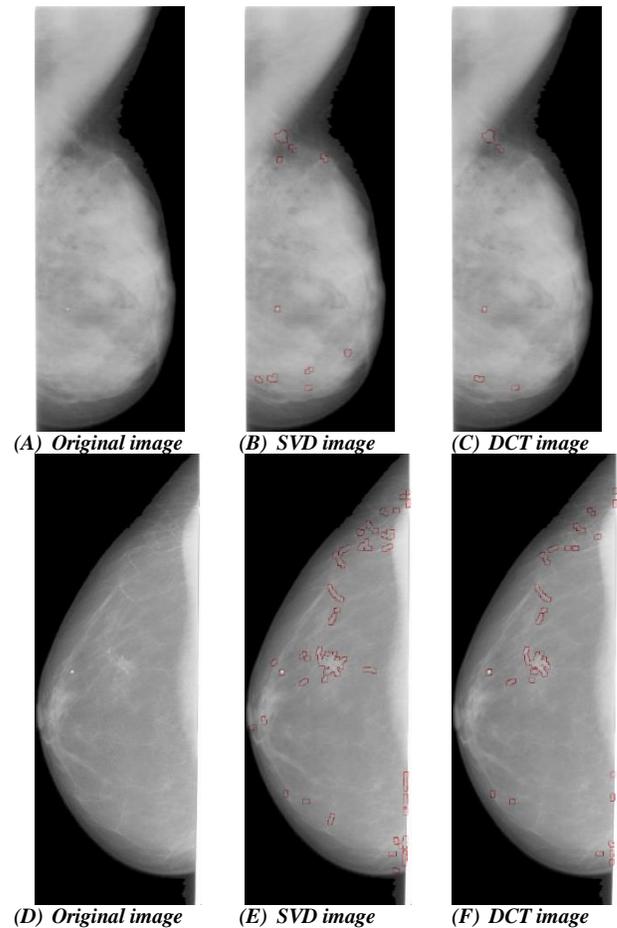


Fig. 5. Detected PMC regions in SVD and DCT compressed images.

classify the detected PMC to TP and eliminate FP regions from the mammogram images.

In order to investigate feature extraction algorithm, initially, the histogram probability $P(i)$ that is calculated from the intensity histograms renormalized was implement on each *actual* TP and FP cluster as in (1).

$$P(i) = \sum_{i=40}^{240} h(i) / MN \quad (1)$$

Where, $h(i)$ is the intensity histogram and M, N are the image region's height and width respectively.

Then five features are calculated and modified considering the upper and lower grey levels of the MC (range from 40 to 240 grey levels) in the mammogram images as shown in (2) to (6).

- 1) The Modified Mean Feature (μ)

$$\mu = \sum_{i=40}^{240} iP(i) \quad (2)$$

- 2) The Modified Entropy Feature(E)

$$E = -\sum_{i=40}^{240} P(i) \log_2 [P(i)] \quad (3)$$

- 3) The Modified Standard Deviation Feature (σ)

$$\sigma = \sqrt{\sum_{i=40}^{240} (i - \mu)^2 P(i)} \quad (4)$$

4) The Modified third order of moment Feature (M3)

$$M_3 = \sum_{i=40}^{240} (i - \mu)^3 P(i) \quad (5)$$

5) The Modified Kurtosis Feature (K)

$$K = \sigma^{-4} \sum_{i=40}^{240} (i - \mu)^4 P(i) - 3 \quad (6)$$

The texture feature algorithm is again implemented on 100 mammogram images. The results show that texture feature algorithm can successfully reduce the detected FP regions in the mammogram images as shown in Fig. 6.

V. ALGORITHM EVALUATION

The adaptive CAD system is implemented on 100 mammogram images resulted from both SVD and DCT image compression techniques. Initially, the CAD system is implemented to detect the PMC regions in the mammogram images and the TP and FP regions are recorded for both SVD and DCT techniques as shown in Table 1. Then texture feature CAD system is implemented on SVD and DCT mammogram images. Also the TP and FP regions are recorded as shown in Table 1.

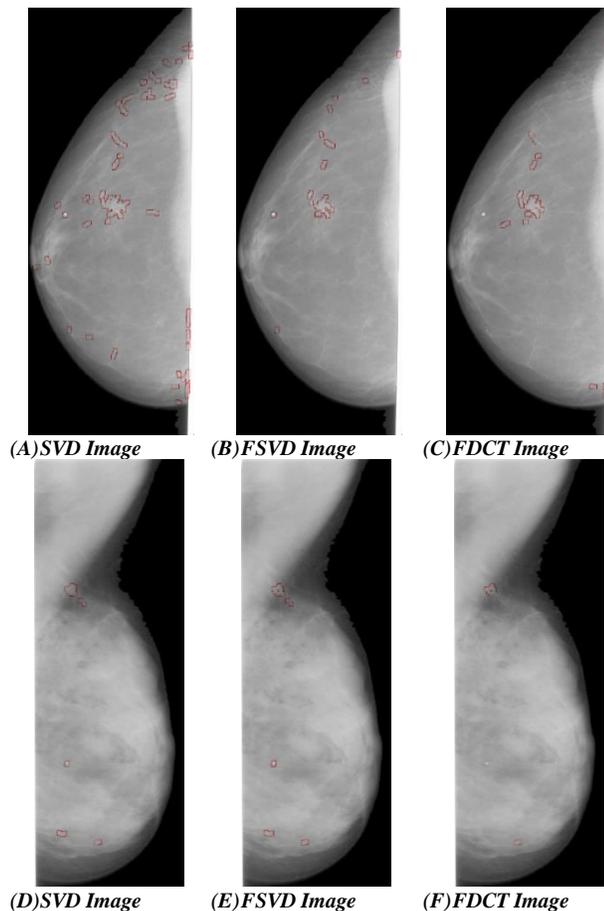


Fig. 6. Processed SVD and DCT compressed images after implementing texture features.

TABLE. I. COMPARISON BETWEEN SVD AND DCT IMAGE COMPRESSION TECHNIQUES BASED ON TP AND FP REGIONS

Image Name	SVD		SVD Feature		DCT		DCT Feature	
	TP (%)	FP	TP (%)	FP	TP (%)	FP	TP	FP
A_1116_1.LEFT_CC	79	30	76	10	83	18	83	2
A_1874_1.RIGHT_MLO	71	16	70	7	80	12	79	3
A_1894_1.RIGHT_MLO	69	65	68	31	82	53	82	9
A_1178_1.LEFT_CC	85	120	81	24	93	87	93	18
Average of 100 mammogram images	79.6	81	79.1	26	87.9	49	87.6	11

As shown in Table 1, the proposed CAD system can effectively detect the MC in the mammogram image in both DCT and SVD image compression techniques. In SVD the average values for TP for 100 mammogram images is 79.1% with minimum FP regions 26 clusters, whereas in DCT the average values for 100 mammogram images is 87.6% with 11 FP clusters. As a result, the DCT image compression techniques is better the SVD image compression technique in maintaining a small details in the image. So the image quality of the DCT is better than SVD since the performance of the CAD system show good result when using the DCT image compression technique.

VI. CONCLUSIONS

A novel adaptive CAD system is proposed in this paper. The proposed CAD system is used to test the image quality resulted from SVD and DCT image compression techniques. Initially, SVD and DCT image compression techniques are used to reduce 100 selected MC mammogram images from both USF and MIAS databases. The resulted images which are SVD and DCT mammogram images are processed using adaptive CAD system. The proposed CAD system can effectively detect most of the MC regions as a PMC but number of FP regions were extremely high. So, five texture features are used to reduce the detected FP regions in the mammogram images. As a result, the adaptive CAD system with texture features can successfully detect the MC in the mammogram images with minimum number of FP regions.

After implementing the proposed adaptive CAD system on 100 mammogram images from USF and MIAS database, it was found that the performance of the DCT image compression technique is better than SVD image compression techniques. The results show that the proposed CAD system can effectively detect the MC in DCT mammogram image with ratio 87.6% TP and 11 FP/regions while in SVD mammogram images the TP ratio was 79.1% with 26 FP/regions.

ACKNOWLEDGMENT

The author is grateful to the Applied Science Private University, Amman, Jordan, for the full financial.

REFERENCES

- [1] American Cancer Society, "Cancer Facts and Figures 2007," Atlanta, Ga: American cancer Society, 2007.
- [2] V. Vogel, "Breast cancer prevention: a review of current evidence," Ca. Cancer J. Clinic, 2000, pp.156-170.
- [3] O.L. Mangasarian, "Breast cancer diagnosis and prognosis via linear programming," Oper. Res., Springer, Vol. 43, pp. 570-577, 1995.

- [4] S. Liu, and E.J. Delp, "Multiresolution detection of stellate lesions in mammograms," Proc. IEEE Intl. Conf. Image Processing, Santa Barbara, 1997, pp. 109-112.
- [5] D. B. Kopans, "The positive predictive value of mammography," AJR Am J Roentgenol, Vol. 158, pp. 521-526, 1992.
- [6] K. Doi, H. MacMahon, S. Katsuragawa, RM Nishikawa, Y. Jiang, "Computer-aided diagnosis in radiology: potential and pitfalls," Radiology, Elsevier, Vol. 31, pp. 97-109, 1999.
- [7] ML. Giger, "Computer-aided diagnosis of breast lesions in medical images," IEEE Computer Science Eng, Vol. 2, pp. 39-45, 2000.
- [8] A. Aldroubi, M. Unser and M. Eden, "Cardinal spline filters: Stability and convergence to the ideal sinc interpolation," IEEE Signal Process, Vol. 28, 1992, pp. 127-138.
- [9] M. Unser, A. Aldroubi and E. Eden, "Fast B-spline transforms for continuous image representation and interpolation," IEEE Trans. Pattern Anal. Machine Intell., Vol. 13, 1991, pp. 277-285.
- [10] M. Unser, A. Aldroubi and M. Eden, "Enlargement or reduction of digital images with minimum loss of information," IEEE Trans. Image Processing, Vol. 4, 1995, pp. 247-258.
- [11] R. kumar Mulemajalu , Shivaprakash Koliwad, "Lossless compression of digital mammography using base switching method", J. Biomedical Science and Engineering, 2009, 2, 336-344
- [12] Z. Liang, Xiangying Du, Jiabin Liu, Yanhui Yang, Dongdong Rong, Xinyu Yao & Kuncheng Li, "Effects Of Different Compression Techniques On Diagnostic Accuracies Of Breast Masses On Digitized Mammograms", Acta Radiologica, Vol. 49, No.(7), 2009, 747-751.
- [13] W.K. Pratt, "Digital Image Processing," 3rd edition, John Willey & Sons Inc, 1991.
- [14] C.S. Chuah, and J.J. Leou, "An adaptive image interpolation algorithm for image/video processing," Pattern Recognition, Elsevier, Vol. 34, 2001, pp. 2383-2393.
- [15] H.S. Hou, and H.C. Andrews, "Cubic splines for image interpolation and digital altering," IEEE Trans. Acoust. Speech Signal Process, ASSP, Vol. 26, 1978, pp. 508-517.
- [16] R.R. Schultz, and R.L. Stevenson, "A Bayesian approach to image expansion for improved definition," IEEE Trans. Image Process., Vol. 3, 1994, pp. 233-242.,
- [17] A.J. Patti, M.I. Sezan and A.M. Tekalp, "High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying blur," Proc. IEEE Intl. Conf. Image Processing, Austin, TX, 1994, pp. 343-347.
- [18] Hugo de las Herasa, Oleg Tischenkoa, Yuan Xub, Christoph Hoeschena, "Comparison of interpolation functions to improve a rebinning-free CT-reconstruction algorithm," IEEE of Medical Physics, Vol. 4, 2007, pp. 430-445.,
- [19] C.H. Kim, S.M. Seong, J.A. Lee and L.S. Kim, "Winscale: An image-scaling algorithm using an area pixel model," IEEE Trans. Circuits and Systems for Video Technology, Vol. 13, No. 6, 2003, pp. 1-4.
- [20] E. Aho, J. Vanne, K. Kuusilinnä and T.D. Hämmäläinen, "Comments on Winscale: An image-scaling algorithm using an area pixel model," IEEE Trans. Circuits and Systems for Video Technology, Vol. 15, No.3, 2005, pp. 345-355.
- [21] Y. Abe, Youji Iiguni, "Image restoration from a down-sampled image by using the DCT," IEEE of Signal Processing Vol. 87, 2007, pp. 2370-2380.
- [22] M. Masek, "Hierarchical segmentation of mammograms based on pixel intensity," Ph.D. Thesis, The University of Western Australia, Australia, 2004
- [23] A. A. AbuBaker, R. S. Qahwaji, M.J. Aqel, M. H. Saleh, "Efficient Pre-processing of USF and MIAS Mammogram Images," Journal of Computer Science, Science Publications, Vol. 3, No. 2, pp. 67-75, 2006.
- [24] A. A. AbuBaker, R.S. Qahwaji, M.J. Aqel, M.H. Saleh, "Mammogram Image Size Reduction Using 16-8 bit Conversion Technique," International Journal of Biomedical Sciences, Enformatica, Vol. 1, No. 2, pp. 103-110, 2006.
- [25] A. Abubaker, "Adaptive Enhancement Technique for Cancerous Lung Nodule in Computed Tomography Images", International Journal of Engineering and Technology (IJET), Vol 8, No 3, 2016, pp. 1444-1450.
- [26] H.S. Sheshadri and A. Kandaswamy, "Computer Aided diagnosis of digital mammograms," ACM Information Technology Journal, Vol. 5, No. 2, 2005, pp. 342-346.
- [27] A. AbuBaker. "Automatic Detection of Breast Cancer Microcalcifications in Digitized X-ray Mammograms". Ph.D., thesis, School of Informatics, University of Bradford-UK. (2008).