

## Hybrid Artificial Intelligent System For Human Gender Classification

Laheeb M. Al-Zoubaidy\*

Susan H. Mohammad\*\*

### Abstract

This paper introduces an automatic human gender classification (male or female ) depending on ultrasound images using artificial neural network to classify the gender . This system consists of three units : preprocessing the ultrasound images of human gender (noise removed , normalization and segmentation ), Feature extraction and gender classification. After preprocessing the ultrasound images then extracting the features by Kernel principal component analysis (kernel PCA) after that, linear neural network used to a classify training and testing these gender images . The system produces promising results for gender classification .

الملخص:

( )

.

)

:

.

, (

. ( kernel PCA)

\*Assistant .Prof/ Software Engineering dept./ College of Computer Sciences and Mathematics

\*\*Asst. Lecture/ Computer Science dept./ College of Computer Sciences and Mathematics

## 1. Introduction

It is often useful to have a machine performance pattern classification to classify ultrasound human gender (male or female) images. Neural network have been a natural choice as trainable pattern classifiers because of their capability to approximate functions and to generalize [1]. Artificial Neural Networks (ANN) are powerful computational systems consisting of many simple processing elements connected together to perform tasks analogously to biological brains. They are massively parallel, which makes them efficient, robust, fault tolerant and noise independent. They can learn from training data and generalize them to new situations. The learning process of ANN is similar to the learning function of the brain. During training, samples are presented to the input layer that yields changes of the activation state of output processing element [1]. The calculated output value is compared to the required value which is also given in the training set.

In this paper, we address the problem of creating an automated automatic human gender classification (male or female) depending on ultrasound images based. We apply artificial Kernel principal component analysis (kernel PCA) as feature extraction and Linear artificial neural network technique to classify the gender.

## 2. Automatic Ultrasound Human Gender Classification System

The automatic ultrasound human gender classification system consists of mainly three modules, preprocessing the ultrasound images of human gender (noise removed, normalization and segmentation), feature extraction using Kernel principal component analysis (kernel PCA) and gender classification using Linear artificial neural network as illustrated in Figure (1).

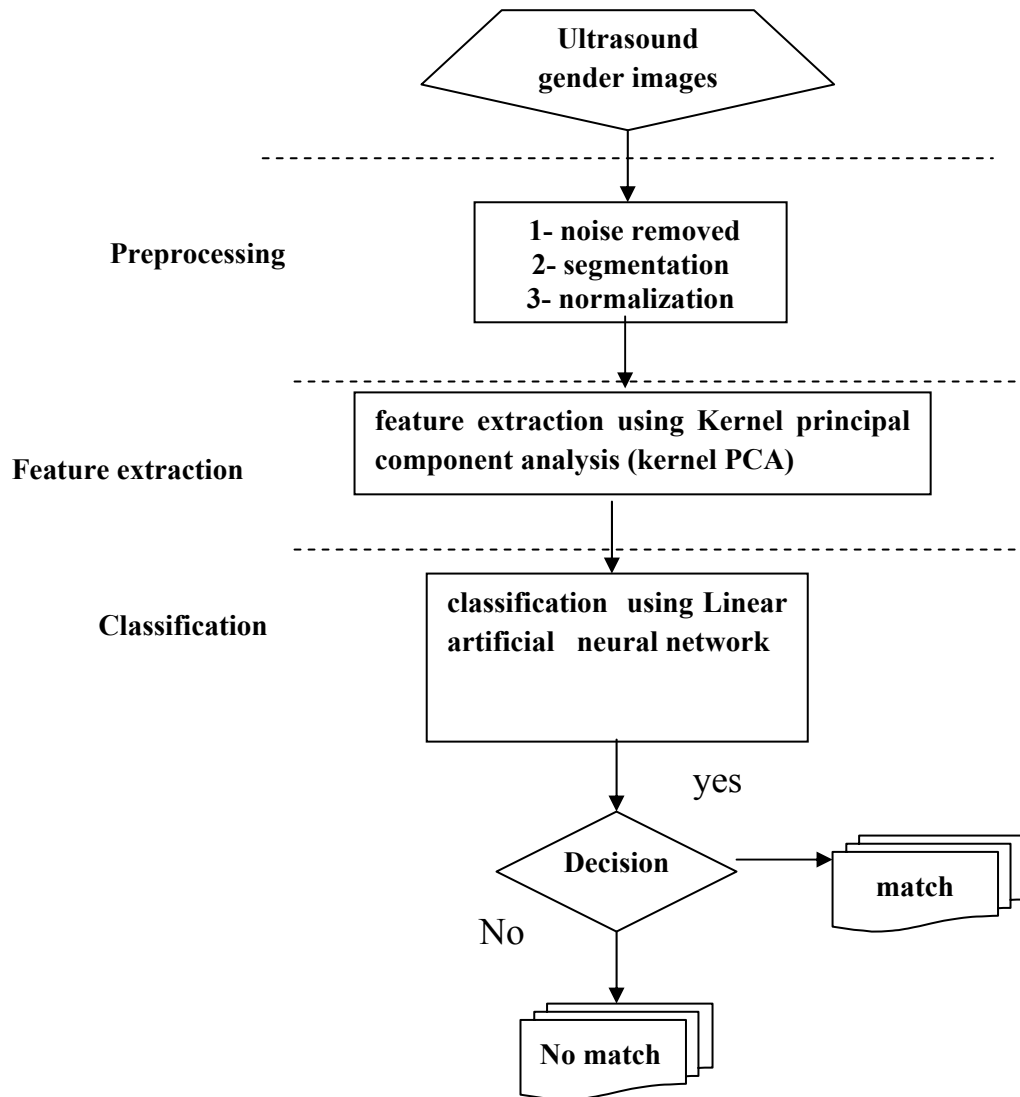


Figure (1) Automatic gender classification system

## 2.1 Preprocessing Module

The Ultrasound images, see figure (2), are rarely of perfect quality. They may be degraded and corrupted due to image noise. Impression conditions and variations in the images. Thus, image enhancing techniques must be used prior to feature extraction, The preprocess module involves a series of image enhancement and steps that can be classified in the following phases:

- Noise removal
- Segmentation
- normalization

For each of the steps, research was conducted on the state of the art of biometric algorithms, looking for the best quality/complexity ratios for each of them. The main goal being a robust implementation of Kernel principal component analysis (kernel PCA) extraction which could be found .

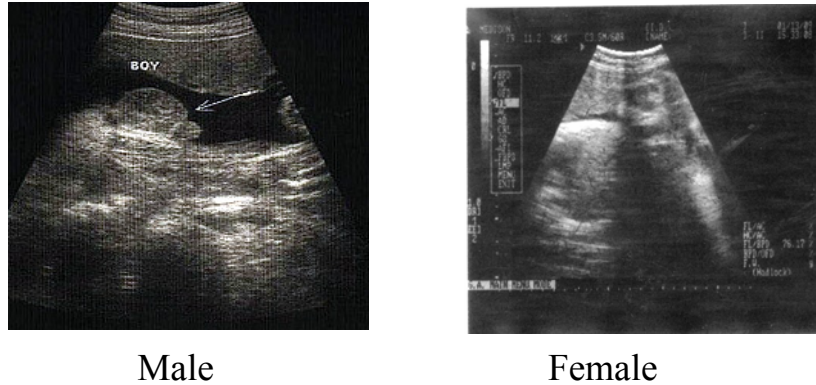


Figure (2) Ultrasound images

**2.1.1 Noise Removal :** The principal source of noise in digital images arise during image acquisition(digitization) and /or transmission, the performance of imaging sensors is affected by a variety of factors , such as environmental conditions during image acquisition, and by quality of the sensing elements themselves [2]. The modeling of image noise is not new and has led to investigations in various fields. Several authors have characterized the noise on aerial images [3,4] because texture analysis (based on image variance) is a fundamental tool of remote sensing for terrain classification. Similar techniques were also used in medical imaging for tissue classification and segmentation [5,6].

The Canny edge detection operator was used In this paper to remove noise and detected edges . uses a multi-stage algorithm to detect a wide range of edges in images. Canny's aim was to discover the optimal edge detection algorithm. In this situation, an "optimal" edge detector means good detection, good localization and minimal response Stages of the Canny algorithm [7,8]

To reduce noise in images, Canny algorithm uses four filters to detect horizontal, vertical and diagonal edges in the blurred image. The edge detection operator (Roberts, Prewitt, Sobel) returns a value for the first derivative in the horizontal direction ( $G_y$ ) and the vertical direction ( $G_x$ ). From this the edge gradient and direction can be determined :

$$G = \sqrt{G_x^2 + G_y^2} \quad \dots\dots\dots (1)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad \dots\dots\dots (2)$$

The edge direction angle is rounded to one of four angles representing vertical, horizontal and the two diagonals (0, 45, 90 and 135 degrees for example). The same binary map shown on the left after non-maxima suppression. The edges are still colored to indicate direction. Given estimates of the image gradients, a search is then carried out to determine if the gradient magnitude assumes a local maximum in the gradient direction. From this stage referred to as non-maximum suppression, a set of edge points, in the form of a binary image, is obtained. These are sometimes referred to as "thin edges". Intensity gradients which are large are more likely to correspond to edges than if they are small. It is in most cases impossible to specify a threshold at which a given intensity gradient switches from corresponding to an edge into not doing so. Therefore Canny uses thresholding with hysteresis. Once this process is complete we have a binary image where each pixel is marked as either an edge pixel or a non-edge pixel. From complementary output from the edge tracing step, the binary edge map obtained in this way can also be treated as a set of edge curves, which after further processing can be represented as polygons in the image domain. A more refined approach to obtain edges with sub-pixel accuracy is by using the approach of differential edge detection, where the requirement of non-maximum suppression is formulated in terms of second- and third-order derivatives computed from a scale-space representation [7,8,9]. See figure (3)

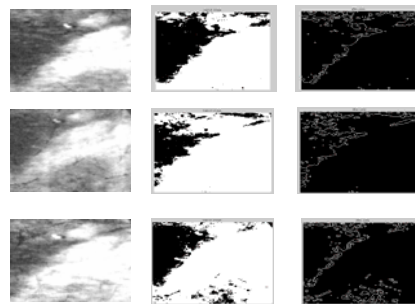


Figure (3) Canny filter on Ultrasound images

**2.1.2 Segmentation:** In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels) (Also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze[10]. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s)[10].

In this paper Region growing methods used to segment images, because the region growing methods can correctly separate the regions that have the same properties we define, region growing methods can provide the original images which have clear edges the good segmentation results, the concept is simple. We only need a small numbers of seed point to represent the property we want, then grow the region, we can determine the seed points and the criteria we want to make, we can choose the multiple criteria at the same time and it performs well with respect to noise. Region growing is one of the simplest region-based image segmentation methods and it can also be classified as one of the pixel-based image segmentations because it involves the selection of initial seed points[11,12,13,14].

This approach to segmentation examines the neighboring pixels of the initial “seed points” and determines if the pixel should be added to the seed point or not. The process is iterated as same as data clustering. We describe the algorithm as below. The main goal of segmentation is to partition an image into regions. Some segmentation methods such as "Thresholding", achieve the goal by looking for the boundaries between regions based on discontinuities in gray levels or color properties. Region-based segmentation is a technique finding the region directly. Here are the basic formulation for Region-Based Segmentation:

$$(a) \bigcup_{i=1}^n R_i = R$$

$$(b) R_i \text{ is a connected region, } i = 1, 2, \dots, n$$

$$(c) R_i \cap R_j = \emptyset \text{ for all } i = 1, 2, \dots, n.$$

(d)  $P(R_i) = \text{TRUE}$  for  $i = 1, 2, \dots, n$ .

(e)  $P(R_i \cup R_j) = \text{FALSE}$  for any adjacent region  $R_i$  and  $R_j$ .

Where  $P(R_i)$  is a logical predicate defined over the points in set  $P(R_k)$  and  $\varnothing$  is the null set.

(a) indicates that the segmentation must be complete; that is, every pixel must be in a region.

(b) requires that points in a region must be connected in some predefined sense.

(c) indicates that the regions must be disjoint.

(d) deals with the properties that must be satisfied by the pixels in a segmented region-for example  $R_i = \text{TRUE}$  if all pixels in  $R_i$  have the same gray level.

And the condition (e) indicates that region  $R_i$  and  $R_j$  are different in the sense of predicate  $P$  [2].

**2.1.3 Normalization :** The image must be normalized after segmentation to size (256 , 256) pixels, so that it has prespecified mean and variance. This results in a maximum span of the grayscale variation in the image, with the help of spreading the histogram of the image across the entire spectrum. This is done by analyzing minimum and maximum values of the image [15].

$$I_{\text{prep}}(x, y) = \frac{I(x, y) - \min(I)}{\max(I) - \min(I)} \dots\dots\dots (3)$$

With  $I$  being the input image pixel values and  $I_{\text{prep}} \in [0, 1]$

## 2.2 Feature Extraction Module

In pattern classification and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is

called features extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Best results are achieved when an expert constructs a set of application-dependent features. Nevertheless, if no such expert knowledge is available general dimensionality reduction techniques may help. These include Principal components analysis and Kernel principal component analysis [16] .

Kernel principal component analysis (kernel PCA) are used to reduced segmented image (256 , 256 ) pixels to 3 kernel PCA values. kernel PCA is an extension of principal component analysis (PCA) using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are done in a reproducing kernel Hilbert space with a non-linear mapping.

Recall that conventional PCA operates on zero-centered data; that is,

$$\sum_{i=1}^N \mathbf{x}_i = 0 \quad \text{..... (4)}$$

It operates by diagonalizing the covariance matrix,

$$\frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \quad \text{..... (5)}$$

in other words, it gives an eigen decomposition of the covariance matrix[17]:

$$\lambda \mathbf{v} = \mathbf{C} \mathbf{v} \quad \text{..... (6)}$$

which can be rewritten as

$$\lambda \mathbf{x}_i^T \mathbf{v} = \mathbf{x}_i^T \mathbf{C} \mathbf{v} \quad \forall i \in [1, N] \quad \text{..... (7)}$$

(See also: Covariance matrix as a linear operator)

To understand the utility of kernel PCA, particularly for clustering, observe that, while  $N$  points cannot in general be linearly separated in  $d < N$  dimensions, but can almost always be linearly separated in  $d \geq N$  dimensions. That is, if we have  $N$  points,  $\mathbf{x}_i$ , if we can map them to an  $N$ -dimensional space with



$$\Phi(x_i) = \delta_{ij} \quad \dots\dots\dots (8)$$

where  $\Phi: \mathbb{R}^d \longrightarrow \mathbb{R}^N$  and  $\delta_{ij}$  is the Kronecker delta.

it is easy to construct a hyperplane that divides the points into arbitrary clusters. Of course, this  $\Phi$  creates linearly independent vectors, so there is no covariance.

In kernel PCA, a non-trivial  $\Phi$  function is chosen so that the points  $\Phi(x_i)$  are not independent in  $\mathbb{R}^N$ . And instead of choosing  $\Phi$  explicitly, we choose .

$$K = k(x, y) = (\Phi(x), \Phi(y)) \quad \dots\dots\dots (9)$$

where  $K$  is the Gramian matrix in the high-dimensional space.

Kernel PCA allows us to operate in such a space without explicitly mapping the data into the high-dimensional space. Because PCA can be cast as an optimization problem in terms of inner products with the transposed data matrix,

$$v_1 = \arg \max_{\|v\|=1} \text{var} \{v^T x\} = \arg \max_{\|v\|=1} E\{(v^T x)^2\} \quad \dots\dots\dots (10)$$

etc. we just need to compute inner products in the high-dimensional space. This is the purpose of the kernel [17].

### 2.3 Linear Artificial Neural Network Module

The result of feature extraction (kernel PCA) are 3 kernel PCA values. These three kernel PCA values are used as input to linear network to classify image gender as male or female.

The linear network provides a good benchmark against which to compare the performance of neural networks. It is quite possible that a problem that is thought to be highly complex can actually be solved as well by linear techniques as by neural networks. If you have only a small number of training cases, you are probably anyway not justified in using a more complex model [18].

A linear neuron with R inputs is shown below .

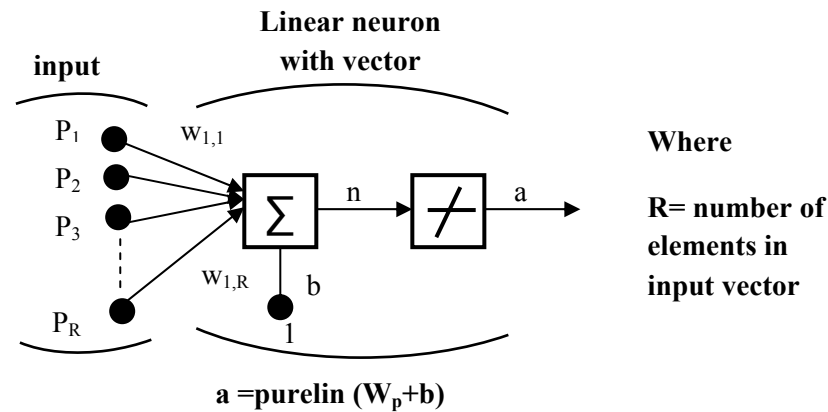


Figure (4) linear neural network

This network has the same basic structure as the perceptron. The only difference is that the linear neuron uses a linear transfer, named purelin. This network is sometimes called a MADALINE for Many ADALINEs. [18].

The network output is

$$a = \text{purelin}(n) = \text{purelin}(W_p + b) = W_p + b \dots\dots\dots (11)$$

or


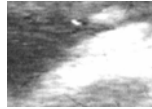

$$a = w_{1,1}p_1 + w_{1,2}p_2 + b \dots\dots\dots (12)$$

However, linear network can classify objects in this way only when the objects are linearly separable. Thus, linear network has the same limitation as the perceptron [18].

### 3.- Experiment results

We have used the linear neural network for training and testing for classification of the ultrasound images of human gender. 75 Ultrasound Human Gender images (Male or female with ages 16 – 36 weeks) taken from Ibn Sena Teaching hospital / Mosul, and from internet [19] divided these images as 50 images for training on linear neural network, and 25 images for testing. The purpose of the experiment is to evaluate the performance of the gender classification system by applying the noise remove techniques (canny filtering) and segmentation using region growing to the human gender images. Kernel principal component analysis are used to reduced segmented image after normalized from (256, 256) pixels to 3 kernel PCA values, See table (1)

Table(1) 3 kernel PCA values for gender image

No.	Images	3 Kernel PCA values		
1		-0.0491	0.0281	-0.0301
2		-0.0466	0.1243	0.0665
3		-0.1324	0.0396	0.0363

Kernel principal component analysis dose extract features which are more useful for classification human gender .A one layer linear neural network was built as shown in figure (5) using Matlab (R2009a) , with the kernel PCA values taken as input values . Here we use purelin as the activation function. The structure of linear neural network used in human gender classification system consists of, see figure (5), one layer feed-forward networks , 3 input nodes ( $x_1, x_2, x_3$ ), 1 output nodes ( $Y_1$ ), number of epoch for training is 4000, bias = 1, learning rate = 0.9.

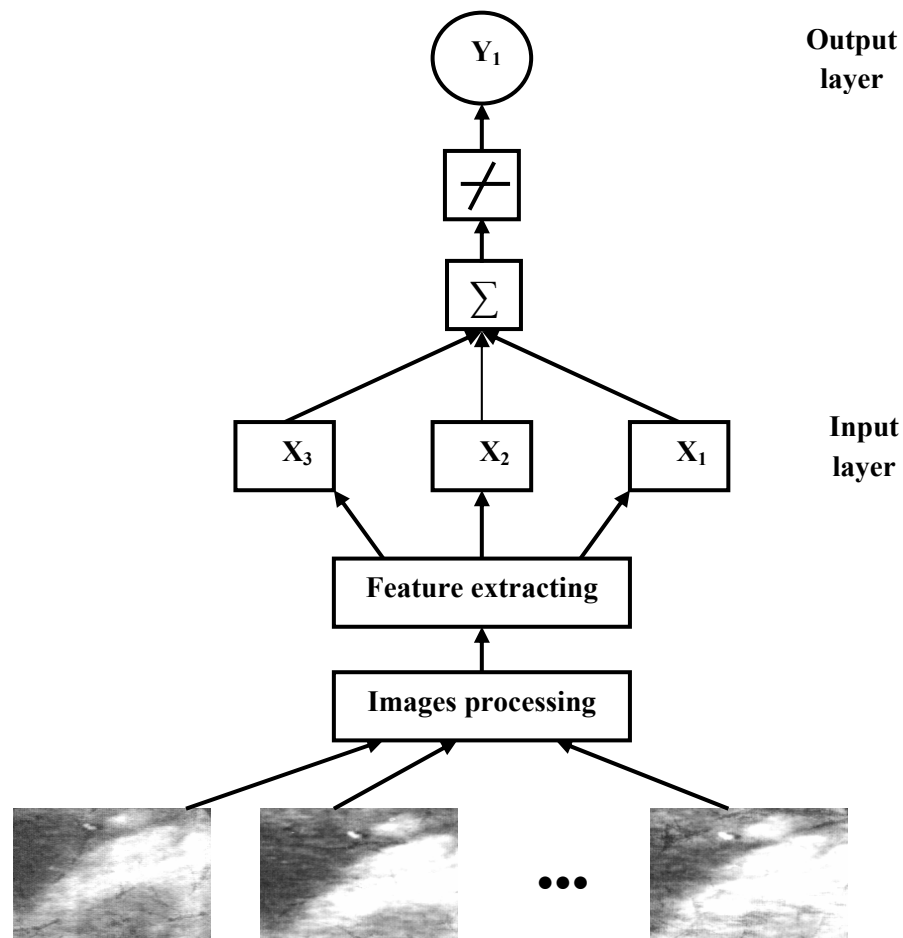


Figure (5) Automatic Ultrasound Human Gender linear neural network Classification System

The result of classification of training on 50 Ultrasound Human Gender images are 99.78% and result of classification of testing on 25 Ultrasound Human Gender images are 92%.

#### 4.Conclusion :

In this paper, a novel human gender classification system using artificial neural networks is presented. The experimental results 99.78% classification of training on 50 Ultrasound Human Gender and 92% classification of testing on 25 Ultrasound Human Gender images are 92%.by using Kernel PAC and linear neural network classifier gives the highest accuracy.

## References :

- 1- Gaetano ,P. (2004) . "*Artificial neural network in the assessment of respiratory mechanism*" . Unpublished doctoral dissertation , Faculty of Medicine , University of Upsaliensis , Uppsala .
- 2- R. C. Gonzalez and R.E. Woods, *Digital Image Processing 2nd edition*, Prentice Hall, New Jersey, 2002.
- 3- J. C. Dainty and R. Shaw, *Image Science*. New York: Academic, 1974.
- 4- R. A. Schowengerdt, "*Noise Models in Remote Sensing. Models and Methods for Image Processing*", 2nd ed. New York: Academic, 1997,ch. 4, pp. 127–131.
- 5- I. N. Bankman, T. S. Spisz, and S. Pavlopoulos, "Two-dimensional shape and texture quantification," in *Handbook of Medical Imaging, Processing and Analysis*. New York: Academic, 2000, pp. 215–230.
- 6- A. Wismüller, F. Vietze, and D. R. Dersch, "Segmentation with neural networks," in *Handbook of Medical Imaging, Processing and Analysis*. New York: Academic, 2000, pp. 107–126.
- 7- Canny, J., *A Computational Approach To Edge Detection*, IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714, 1986.
- 8- R. Deriche, *Using Canny's criteria to derive a recursively implemented optimal edge detector*, Int. J. Computer Vision, Vol. 1, pp. 167-187, April 1987.
- 9- Lindeberg, Tony "*Edge detection and ridge detection with automatic scale selection*", International Journal of Computer Vision, 30, 2, pp 117--154, 1998. (Includes the differential approach to non-maximum suppression.)
- 10- Linda G. Shapiro and George C. Stockman (2001): "*Computer Vision*", pp 279-325, New Jersey, Prentice-Hall

11- Jian-Jiun Ding, The class of "*Time-Frequency Analysis and Wavelet Transform*", the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan, 2007.

12- Jian-Jiun Ding, The class of "*Advanced Digital Signal Processing*", the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan, 2008.

13- W. K. Pratt, "*Digital Image Processing 4th Edition* ", John Wiley & Sons, Inc., Los Altos, California, 2007

14- M. Petrou and P. Bosdogianni, " *Image Processing the Fundamentals* ", Wiley, UK, 2004.

15- P. Hymër , " *Extraction and Application of Secondary Crease Information in Fingerprint Recognition* " , department of Science and Techonlogy, Linköpings Universitet , SE-60174 Norrköping , Sweden , 2005.

16- Kernel principal components analysis – wikipedia , the free encyclopedia , [en.wikipedia .org / wiki / kernel- principal- components- analysis](http://en.wikipedia.org/wiki/kernel_principal_components_analysis) .

17- B. Schölkopf , A . Smola and K . Müller , "*Nonlinear Component Analysis as a Kernel Eigenvalue Problem*" Max-Plank- Institut für biologisch Kybernetik Arbeitsgruppe Bülthoff , Spemannstraße 38.72076 Tübingen – Germany , Technical Report No.44 , December 1996 .

18- Matlab Language (R2009a) (Toolbox) , ver 7.8

19- Ultrasound Photos of Girls and Boys in Pregnancy By [Robin Elise Weiss](#), [LCCE](#), [About.com](#)