Graphics Separation and Skew Correction for Mobile Captured Documents and Comparative analysis with Existing Methods

H.K.Chethan
Research Scholar, Department of
Studies in Computer Science
University of Mysore, Mysore-570006, India

G.Hemantha Kumar
Professor, Department of
Studies in Computer Science
University of Mysore, Mysore-570006, India

ABSTRACT

CBDA is an emerging field in Computer Vision and Pattern Recognition .In recent technology camera are moulded with several equipments and are very interesting and playing a vital role by replacing scanner with hand held imaging devices(HIDs) like Digital Cameras, Mobile phones and gaming devices. Documents captured through Mobile are often prone to Skew Removal of graphics and Correction of skew for Mobile captured document is a major task and important factor in optical character recognition. The goal of the work is to remove graphics from the document and correct skew for the documents captured using cellular phone. In this paper we have proposed a novel method for separating or removal of graphics like logos, animations other than the text from the document and finally textual content skew is corrected and characters are recognized using commercial OCR. The basic process of our approach consists of three steps: First, a vertical and Horizontal projection is used to remove graphics from images secondly dilation operation is applied to the binary Images and the dilated Image is thinned; finally, the skew angle is detected using the Hough transform. The proposed approach with high precision can detect skew with large angle (-90 to +90) the experimental result reveal that the proposed method is efficient compared to well known existing methods. The experimental results show the efficacy compared to the result of well known existing methods.

Keywords

HIDs, CBDA, vertical and Horizontal Projection, Hough Transform, Skew, Computer Vision, Pattern Recognition

1. INTRODUCTION

Camera were developed to capture good and bad moments of life but as the technology is getting advanced more and more application are being developed for digital camera one such application is extraction of text regions in camera captured image/document, this is a well known problem in the area of computer vision, therefore Portable cameras are Ubiquitous in nature .Either in standalone versions, or incorporated in cell phones, the quality of the images has risen at a fast pace while their price has dropped drastically. Such pervasiveness has given rise to unforeseen application such as using portable cameras for digitalizing documents by user of many of different professional areas for instance, students and professionals are taking photos instead of taking notes.

This new research is evolving fast in many dimensions. Recent price performance has given birth to several new application[1][2].Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content retrieval from images and videos[3].With the help of digital camera we can capture characters and documents any-where in the 3D environment like signs and bill-boards ,color,texture,shape, as well as the relation-ship between them.CBDA is required because we are no longer constrained to traditional 2D image of As stated by Jung, Kim and Jain[4] text data is particularly interested because text can be used to easily and clearly describe the contents of an image since text data can be embedded in an image or video in different font styles ,sizes ,orientations ,colours

and against a complex background. Most of the algorithms are very sensitive to the distortion of document images. And skew may cause

serious problem for document analysis. Furthermore, it is very difficult, if not totally impossible, to prevent skew distortion produced in scanning procedure. It is therefore preferable to detect and correct the skew document image at the pre-processing stage in order to avoid the disturbance of skew to the further processing. Several attempts have been made for skew detection. And the methods can be mainly categorized into five groups. The one based on Hough transform, Cross Correlation, Projection profile, Fourier transformation and K nearest neighbor (K-NN) clustering. Hough transform has been used by [5] for skew detection. The basic method consists of mapping points in the Cartesian space (x, y) to sinusoidal curves in (ρ, θ) space via the Transformation

$$\rho = x \cos \theta + y \sin \theta \qquad (1)$$

Each time a sinusoidal curve intersects another particular value of ρ and θ, the likelihood increases that a line corresponding to that (ρ, θ) coordinates value is present in the original image. An accumulator array is used to count the number of intersect the various ρ and θ values. The skew is then determined by the θ values corresponding to the highest number of counts in the accumulator array. In [6] use bottom pixels of the candidate objects within a selected region for Hough transformation. The hierarchical Hough transformation technique is also adopted in another paper [7]. The main idea of the above methods is to reduce the amount of Input data, but their computational complexities are still very high. In [8] they proposed an improved method to overcome the drawback of the method [6].The cross-correlation method proposed by [9] is based on the correlation between two vertical lines in a document image. Since the pixels in the two parallel lines are translated due to skew, correlation matrix can be produced, it is defined as

$$R(X_0, S) = \sum L_1(X_0, Y) L_2(X_0+d, y+s) \qquad (2)$$

(here, L_1 (x, y) and L_2 (x, y) denote the two parallel vertical lines, respectively, d is the space between L1 and L2and s is the maximum translation. though this technique can result in accuracy, it is time consuming to calculate the correlation matrix and its projection profile. Furthermore, in certain situations parameter d should be changed and backtracked, which may increase the computing cost. The horizontal (vertical) projection profile [10] is a histogram of the number of black pixels along horizontal (vertical) scan lines. In [1] they have described an approach where the document is partitioned into vertical strips. The horizontal projection profiles are calculated for each strip and from the correlation of the profiles of the neighbouring strips, the skew angle is determined. Although the proposed method is computationally inexpensive, it works well if the document is skewed within ±100. Method proposed by [10] is based on vertical projection profile of horizontal strips which works well if the skew angle is small. The method proposed by [11] belongs to Fourier Transformation approach. According to their method, the skew angle of a document image corresponding to the direction where the density of Fourier space becomes the largest. However its computing complexity is very high. In [5] proposed nearest neighbor clustering to skew detection. He found all the connected components in the documents and for each component computed the direction of

its nearest neighbor. A histogram of the direction angle are computed, the peak of which indicates the document skew angle. In [8] the K nearest centres of the successive connected components is selected to calculate the vector directions between random pairs. The histogram peak corresponding to the skew angle of the whole document image will be generated afterwards. This approach achieves high accuracy, yet with high computing complexity O (N2) (here N is the number of connected component). Method by [12] proposed a nearest neighbor chain based approach to skew estimation in document images. Size restriction is introduced to the detection of nearest neighbour pairs. Then the chains with a largest possible number of nearest neighbor pairs are selected and their slopes are computed to give the skew angle of document image. Method by [13] proposed skew detection and correction in document images based on straight–line fitting. The bottom centre of the bounding box of a connected component is regarded as an eigen point. According to the relations between the successive eigen-points in every text line, the eigen points laid on the baseline are extracted as sample points. Then these samples are adopted by the least squares method to calculate the baseline direction. In order to reduce the computing cost and to gain high accuracy of the above Hough transform based approaches, a fast skew angle detection method is proposed in this paper. The proposed method considers some of the selected character present in the document. The selected component is blocked and thinning is performed to the region. The obtained thinned points of the document region are then applied to Hough transform to estimate skew angle accurately. Experiments prove that through the proposed approaches the speed can be improved and the higher accuracy can be achieved. Some key challenges in the development of OCR for mobile cam based are:

1.1 Data Quality

Images captured using Mobile often suffers from a number of distortions and degradation. A major distortion includes uneven lighting and skews which we deal in the paper. Skew occurs when the camera axis is not properly aligned with that of the object and illumination occurs due to uncontrolled environment like shadows reflection and other. Quality of the image is greatly affected due to improper lighting condition of the surroundings.

1.2 Nature of the Text

Text appearing on the document may be cursive, underlined, and artistic and of various fonts and styles text can either be embedded on the animations or it can be included with in the document anywhere.

1.3 Light weight algorithms for mobile devices

The ultimate goal will be to embed document analysis processing directly into the devices but computation for mobile devices is very lethargic and suffers from three major computing constraints viz.poor computing power, low working memory and no floating point unit which is very difficult to handle floating point and should be avoided or else results in the larger computational time which is not efficient.

In order to recognize the text information from the captured document, text regions need to be separated from the graphic content .pre-processing steps are performed to reduce illumination and skew correction and finally text are extracted and recognized for further process based on the application.

2 PROPOSED METHODOLOGY

The current work carried out in the paper is divided into five modules Pre-processing, Graphics removal, Dilation, Thinning and finally Hough Transform to correct skewed document, graphic removal from mobile captured images,
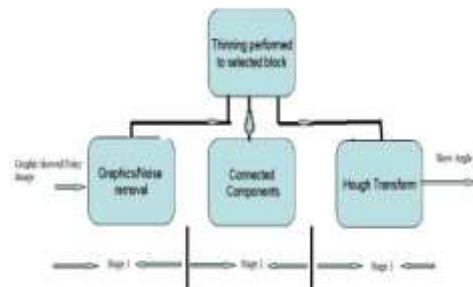

Fig. 1 Block Diagram of the Proposed Method

finally skew detection and correction which are very essential to improve the data quality of the image which will be fruitful final stages
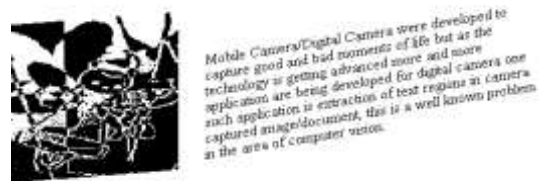

Fig.2 Sample Images

2.1 Pre-processing

It is the initial stage where captured document is pre-processed before skew detection. We convert input image to binary image. In Common, the histogram of the document image with 256 gray scales is apparently bimodal, the one is text pixels, and the other is background pixels. By Thresholding, 1 represents a dark pixel in a character and 0 represents a bright pixel in the background. Image binarization converts an image of up to 256 gray levels to a black and white image. Survey [13] have showed that global thresholding is not ideal for Camera-captured images due to lightning variations so We proposed locally adaptive thresholding method that is robust to variation of llumination.The simplest way to use image binarization is to choose a threshold value, and classify all pixels with values above this threshold as white, and all other pixels as black.

$$T_{rxr} = T * g_{high} - g_{low} \qquad (3)$$

Where g_{high} and g_{low} are maximum and minimum Intensity value of pixels in rxr sub-window. Some samples include noise. Removal of that noise from documents which is a common problem Noise is a common problem in most of the image understanding problems. These involve white noise that is introduced by interferences in the sensors .The common sources of noise include white noise, salt and pepper noise, quantization artefacts, etc. These are well-known noise sources that are compensated for by using techniques such as median filtering, dithering, low-pass filtering, etc. In this section, we look at one of the artefacts that
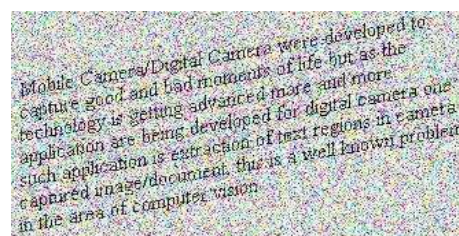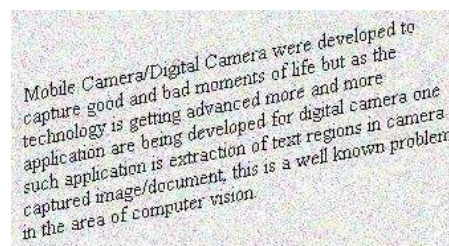
Fig. 3 Sample of Noise images are specific to document images, called halftones. Using low pass filters and Median filter Noise are reduced.

2.2 Graphics Separation

After pre-processing Mobile captured document next step is separation of the text regions from background. Texture and graphics is an important step of any optical character recognition system for the images containing both texts and graphics
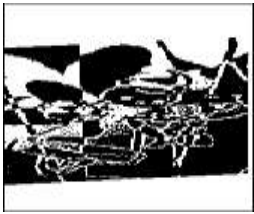

Fig. 4 Graphics eliminated

Based on the connected components Horizontal and vertical projection are applied to images to separate graphic and text from the image based on the height, width, width to height ratio, gray pixel density, black pixel density, number of vertical and horizontal segments and change in threshold values[14]
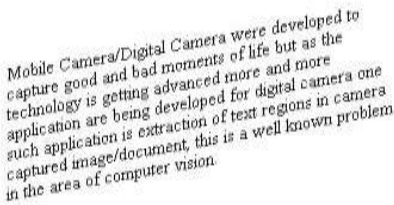

Fig. 5 Text document separated from original document

2.3 Dilation

In this method first selected characters from the document images are blocked and dilated in order to obtain word blocks and thinning is performed over the word block regions .A horizontal line structuring is used for dilation In mathematic morphology, the dilation operator that the image A is dilated by the structuring element B is defined as:

A©B=U {A+b: b€B} = U {B+a: a€A}   (4)

 It is well known that the interval between two text lines is wider than that between two characters. By selecting appropriate quadrate structuring element, the text lines will be dilated into thick lines, while the intervals between two adjacent lines are still background. An example is showed in Fig. 6.
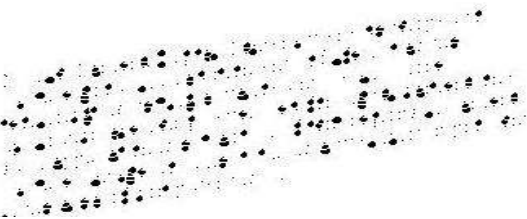

Fig. 6 Dilation performed over Textual Content

2.4 Thinning

In order to extract the skeleton of the dilated image, a thinning operation is performed. In the process of thinning, two demands need to be satisfied: 1) The image is dwindled regularly; 2) The connected property should be kept during the thinning process. Considering a 3x3 pixels region in the image, its central pixel is marked as p_1, and the others are marked orderly as $p_2 \ldots p_9$. If $p_1 =1$(dark pixel), and all

of the following conditions are satisfied, then the pixel Pl ($p_1 =0$) can be deleted:

1) $2 \leq N (p_1) \leq 6$;
2) $S (p_1) =l$;
3) P2. p_4. $p_6=0$;
4) P4. p_6. $p_8 = 0$;

Where N (p_1) is the number of the neighbour pixels of Pl which value is not 0, S (p_1) is the transition counts from 0 to 1 in accordance with the order from p_1 to p_9.
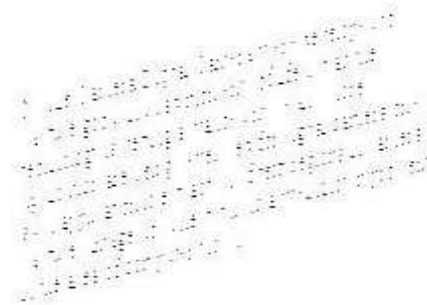

Fig. 7 Thinned result Obtained over the dilated image

Every pixel is performed with this operation until no pixel can be deleted. The text lines in the dilated image will become the thin line that is showed in Fig. 7 after thinning operation.

3 SKEW CORRECTION

Correction of skew is an important factor in optical character recognition. Documents captured through Mobile are often prone to Skew. Very few works have been reported on skew estimation for mobile cam based documents. In our work we employ Hough transform to detect and correct skew. When a document is scanned or Captured at an angle with respect to the reference axes skew results. The problem of skew correction plays an important role in the effectiveness of many document analysis algorithms, such as text line estimation, region boundary detection, etc. For example, algorithms based on projection profiles assume an axis-aligned scan. The primary challenge in skew correction is the estimation of the exact skew angle of a document image. A variety of techniques are used for the detection of skew. Most of them assume the presence of some text component in the document and estimate the orientation of text lines using different methods. The commonly used techniques are projection profiles and line fitting. In the projection profile-based approach, we compute the histogram of text pixels on the vertical axis (vertical projection) and try to maximize the distance between consecutive line profiles, called line spacing, while rotating the document image.

4 HOUGH TRANSFORM

Hough transform is widely used in image analysis and computer vision and digital image processing
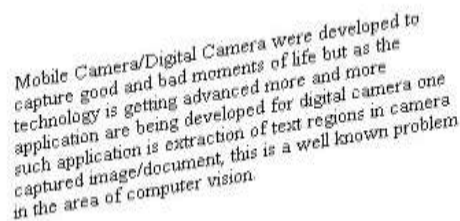

Fig. 8 Original Skewed document

Hough Transform technique is an approach used for fitting lines and curves. This approach is preferred when the objective is to find lines or curves formed by groups of individual points on an image plane. The method involves a transformation from an image plane to a parameter space Consider the case in which lines is the objects of interest. The line is expressed as

$$\rho = X\cos\theta + Y\sin\theta. \qquad (5)$$

There are two line parameters namely, the distance (ρ) and the angle (θ) which defines transformation space. Each coordinate (x, y) of ON pixel in the image plane is mapped onto the locations in the transformed plane for all possible straight lines. This is depicted in Fig. 9 and 10. For all possible values of ρ and θ the transformations intersect at the same point on the transformed plane when multiple points are collinear. Therefore, the point (ρ, θ), which has the greatest accumulation of mapped points, indicates lines with these parameters. In practice, due to discrimination error and noise, points mapped will not be exactly collinear. Thus the points do not map on to exactly the same location on the transformed plane. For connected lines or positions of lines, computations can be reduced greatly by considering not all (ρ, θ) points but only those (ρ, θ) points that are in one orientation as indicated by the angle.

The HT has a limitation that it does not give the coordinates of end points of the line, and further the long lines are favoured over short lines. For these reasons, spatial domain method is considered to be often faster and more effective for skew detection and estimation.
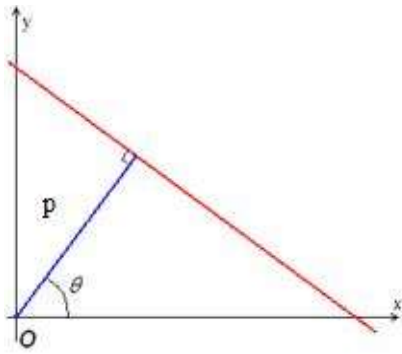

Fig. 9 X, Y plane for line detection

Mobile Camera/Digital Camera were developed to capture good and bad moments of life but as the technology is getting advanced more and more application are being developed for digital camera one such application is extraction of text regions in camera captured image/document, this is a well known problem in the area of computer vision.

Fig. 10 Skew Corrected document using Hough transform

4.1 ALGORITHM SKEW

Step1: Find connected components in the document image and compute average bounding height (AH).
Step2: Select those connected components whose height is less than AH and remove very small connected components so that the dots of the character i, j punctuation marks like full stop, comma, hyphen etc. are deleted.
Step3: Block the selected component present in the document.
Step4: Perform thinning operation over the selected block region.
Step5: Remove the parallel straight lines using prespecified threshold.
Step6: The obtained points are then subjected to Hough transform to estimate skew angle accurately.
Step7: Stop.

5 EXPERIMENTAL RESULTS

Experiments were conducted on large dataset captured using mobile camera of 2 Mega pixel with 640* 480 resolutions with varying distance from 10 to 50 cm in uncontrolled environment .Proposed method has been implemented using Matlab language on Core 2 Duo 2.66 GHz machine. Dataset were captured from various sources like newspaper, journals, text books, sign board and on walls .all the samples were captured in 3D environment with skew ranging from 1 degree to 60 degree.

For experimental purpose 500 document have been used as shown in figure 1 .Mean skew angle and standard deviation have been calculated as shown in Table 1.

Table 1 Estimated Skew angle, Mean (M) and Standard Deviations (SD)

True Angle in degree	Proposed Method	
	M	**SD**
3	2.89	0.340
5	4.90	0.357
10	10.0	0.328
15	14.99	0.293
20	20.00	0.349
25	24.19	0.396
30	30.15	0.431
40	39.5	0.220
50	51.5	0.420
60	59.8	0.465

Experiments were conducted on dataset where data quality of the document is degraded by noise, illumination and perspective distortion Result obtained for noisy images are shown in Table 2.

However the proposed method fails if the density of noise is more than 0.05 as shown in Table 2 from the table we can see that performance decays as the noise density increases.

Table 2 Proposed Method with different Noise densities

Noise Density	Known angle(degree)	Calculated angle (degree)	Performance (%)
0.01	45	43.9	97.55
0.02	45	45.12	99.73
0.03	45	44.94	99.86
0.04	45	44.85	99.66
0.05	45	45.20	99.55
0.06	45	48.64	92.51
0.07	45	36.78	81.73

5.1 Comparative study

Proposed method is compared with the well known existing methods for accuracy and efficacy .Comparative analysis is performed on 100 samples captured from different content Sample image includes skew angle varying from 3, 5, 10, 15, 20, 30, 40, 50 and 60 degrees. Mean, standard deviation and computing time obtained for the proposed method and well known existing methods have been reported in Table 3. By looking at the Table 3 we can come to a conclusion that the proposed method is suitable to correct skew and is accurate and precise compared with other existing methods.

6 CONCLUSIONS

In summary the paper reveals that only few work have been reported towards Graphic separation and Skew correction for Mobile captured documents .In our work we have presented vertical and horizontal projection methods to remove graphics from documents and finally skew correction is performed using Hough transform. Experiments have been performed for documents containing graphics and without graphics and also for noisy images. Experimental results reveal that it

works better for both types with graphics and without graphics and therefore proposed method is efficient, novel and accurate for mobile captured documents. Where as in the case of noisy images the proposed method performance degrades as the noise density increases. A comparative analysis is performed for the existing and proposed method. However proposed method as well as the existing method fails to work for document of low resolution and illumination images.

7 REFERENCES

1. Jian Liang, David Doermann and Huiping Li: Camera-based analysis of text and documents: a survey, Springer-Verlag 2004.

2. Majid Mirmehdi: Special issue on camera-based text and document recognition, Springer-Verlag .2005.

3. P.Shivkumara, Weihua Huang and Chew Lim Tan ,Efficient Video Text Detection using Edge Features, IEEE 978-1-4244-2175,2008.

4. Keechul Jung, Kwang In Kim and Anil K. Jain:Text Information Extraction in Images and Video: a Survey, Pattern Recognition, 37 PP.977-997, 2004.

5. Srihari SN and Govindaraju V, Analysis of textual images using the Hough Transform, Machine Vision and Applications, vol 2, 1989, pp. 141-153.

6. Le D S, Thoma G R and Wechsler H, Automatic page orientation and skew angle detection for binary document images. Pattern Recognition 27, 1994, pp 1325-1344.

7. Yu, B., Jain, A.K., A robust and fast skew detection algorithm for generic documents, Pattern Recognition 29 (10), pp 1599-1629, 1996.

8. U and Chaudhuri B. B, An Improved document skew angle estimation technique, Pattern Recognition Letters, Vol. 17, 1996, pp 899-904.

9. Yan, H. Skew correction of document images using using interline cross correlation, Computer Vision, Graphics, and Image Processing 55, 1993, pp 538-543.

10. Hou H.S., Digital Document Processing, Wisely New York, 1983, Computer Vision, Graphics

11. Postl W, Detection of linear oblique structures and skew scan in digitized documents. Proceedings 8th International Conference on Pattern Recognition, 1986, pp. 687-689.

12. Yue Lu and Chew Lim Tan, A nearest neighbor chain based approach to skew estimation in document images, Pattern Recognition Letters 24, 2003, pp 2315-2323.

13. Cao Yang, Shuhua Wang, Li Heng., Skew detection and correction in document images based on straight-line fitting, Pattern Recognition Letters, 24, pp 1871-1879, 2003.

14. A.F.Mollah, S.Basu, M.Nasipuri and D.K.Basu,"text/graphics separation for business card images for mobile devices", Proceedings of the Eight IAPR International Workshop on Graphics Recognition (GREC09), July, 2009, France,pp263-270.

Table 3
Comparative analysis of Existing Methods with the proposed methods

True Angle in degree	Proposed Method using Hough Transform			Projection Profile Method			Cross Correlation Method			KNN Clustering Method			Base point Method		
	M	SD	CT (s)	M	SD	CT (s)	M	SD	CT (s)	M	SD	CT (s)	M	SD	CT (s)
3	2.89	0.340	**1.78**	5.66	0.784	**2.78**	3.12	1.20	**2.40**	3.60	1.240	**2.10**	3.10	0.740	**1.90**
5	4.90	0.357	**1.66**	6.40	1.30	**1.89**	5.6	1.750	**1.70**	4.60	0.750	**1.80**	5.00	1.570	**1.80**
10	10.0	0.328	**1.70**	11.25	2.36	**2.20**	10.60	2.754	**2.00**	10.30	0.854	**1.80**	10.10	0.850	**1.95**
15	14.99	0.293	**1.77**	15.20	1.20	**2.77**	15.20	0.950	**2.20**	14.20	0.450	**2.00**	15.20	0.756	**2.00**
20	20.00	0.349	**1.80**	20.40	2.40	**2.60**	21.20	0.850	**2.10**	20.40	0.540	**1.90**	21.50	0.450	**2.10**
25	24.19	0.396	**1.72**	25.62	3.396	**2.42**	24.50	3.396	**2.25**	25.80	1.210	**2.42**	25.75	1.950	**1.95**
30	30.15	0.431	**1.42**	32.14	2.40	**1.80**	32.50	2.420	**2.00**	31.40	1.450	**1.90**	31.50	0.850	**1.80**
40	39.5	0.220	**1.74**	39.5	1.10	**1.96**	41.50	1.754	**1.80**	40.60	0.750	**1.90**	40.35	1.654	**1.80**
50	51.5	0.420	**1.78**	54.12	1.30	**1.98**	52.50	1.112	**1.78**	50.50	0.650	**1.70**	52.00	0.895	**1.85**
60	59.8	0.465	**1.66**	60.12	0.675	**2.10**	61.50	1.256	**2.40**	60.50	0.550	**2.05**	62.50	0.575	**2.00**

Computed Mean, Standard Deviation and Computing Time using proposed method and existing methods to measure Accuracy, Consistency and Efficacy