Journal of Telematics and Informatics (JTI)
Vol.3, No.2, September 2015, pp. 49~59
ISSN: 2303-3703

Automated Erythrocytes Counting in Microscopic Thin Blood Smear Digital Images

Naveed Abbas*, Dzulkifli Mohamad, Abdual Hanan Abdullah
Faculty of Computing, Universiti Teknologi Malaysia
UTM Skudai, 81310 Johar Bahru, Malaysia

*Corresponding author, e-mail: naveed.abbas@icp.edu.pk

Abstract

The Erythrocytes counting is part of the complete blood count test and is frequently suggested by the Physician to know the number of Erythrocytes in the patient's body. At present mostly the counting process is performed manually which is laborious, error prone and time consuming. The main purpose of this study is to use the digital image processing techniques to automate the counting process of the Erythrocytes or Red Blood Cells in Microscopic thin Blood smear digital images. The automated diagnosing gain the attention of the researchers from the last two decades because it assist the experts to reduce the burden of errors, labour and time of examination. In this regard, too much research has been performed on the automation of the counting process of the Erythrocytes but still the test demands to be done in a proper, efficient, accurate and realistic way. The proposed method achieved an average True Positive Rate (TPR) of 95%, True Negative Rate (TNR) of 5%, average accuracy of 97% and average error of 3%.

Keywords*: Counting Erythrocytes; Red Blood Cells; Health care Applications; Clustered RBC splitting; Complete Blood Count (CBC).*

1. Introduction

The counting of Erythrocytes is demanding process in various blood tests because the deviation of number of Erythrocytes from normal range in both cases (Low and High) is an important indicator about any disorder in the body. The normal range of the Erythrocytes in male is 4.7-6.1million cells /mcl and in female is 4.2- 5.4 million cells/mcl. The number of Erythrocytes high then normal range indicates Kidney tumour, Heart diseases, Low Blood oxygen level etc. while the low number of Erythrocytes from its normal range indicates, Anaemia, Haemorrhage, Leukaemia, Mal-nutrition, Nutritional deficiencies like iron, foliate, copper etc [1]. Due to consumption of too much time, jeopardy of errors and much physical and mental labour on the part of haematologist increases the demand of automatic counting techniques to combat the mentioned problems by assisting the haematologists [2, 3]. In this connection, many researchers did much work but still the work needs to be more efficient, robust, accurate and realistic. This study considered the proposed technique in the context that it will be efficient, accurate, robust and realistic. Counting Erythrocytes through image processing techniques is not difficult task but for high accuracy it involves several other problems i.e. image pre-processing, separation of single and clustered Erythrocytes. If theses mentioned problems are not addressed in proper way then the accuracy will be on compromise because the occluded or clustered Erythrocytes are appeared as a single area and in reality it is combination of more than one Erythrocyte. Further, the clustered Erythrocytes are divided into Clumped and Overlapped Erythrocytes. Clumps of Erythrocytes occurred in the case when iron deficiency exists in the blood, the Erythrocytes glued each other and formed long chains while overlapped Erythrocytes are formed due to improper slide preparation and is also considered as big problem in the manual microscopy as also mentioned by[4] because it leads to discard the slide and prepare another one. This study considered all these problems and after solving the given problems then quantify the Erythrocytes.

Recently, too much effort have been made by researchers to develop algorithms for the quantification of Erythrocytes addressing the problems of splitting the clustered Erythrocytes

and show a high degree of success but still needs improvements to address the mentioned hurdles in proper way. The study made by [5], the authors mentioned that quantifying Erythrocytes is not a big issue in image processing but the hurdles like clustered Erythrocytes splitting is too important because they will affect the accuracy that's why they did it through concavity points finding and splitting. However, they did not mention how to separate the single and clustered Erythrocytes and identification of clusters existence while the Erythrocytes are counted using boundaries tracing and labeling. In the study, of [6], the authors did not consider the separation and clustered Erythrocytes splitting but did the counting. Erythrocytes counting without solving the problem of cluster Erythrocytes Splitting compromise on the accuracy. Some studies while counting the Erythrocytes do not consider the clumps and overlaps of Erythrocytes for splitting but they rely on guessing Area based estimation approaches as mentioned in the work of [7-8]. The problem in this approach is that in some cases we want to note the disorder as well in the Erythrocyte in such case this approach will fails while also the areas of Erythrocytes by most of the studies considered as circular, which is not true as because morphology of the Erythrocytes highly changes due to any disorder. Circular Hough Transform based approaches for counting and splitting as mentioned by [9-12] mainly considered the Erythrocytes as circles which is not true because Erythrocytes morphology is not static and changed by other diseases.

The approaches adopted by previous studies to combat the problem of clumped and overlapped Erythrocytes splitting are divided into the following categories i.e. Morphological operation based includes erosion, dilation or opening closing to split the clusters of Erythrocytes [13-15],. However, the main problem in morphological based approach is that it works well in overlap of Erythrocytes not more than two cells but in reality we have some clumps of Erythrocytes which are very long chains. Concavity based approaches deal the problems in the way to find out the concavity regions and some cases the concavity points and split the clustered Erythrocytes through lines cuts or circles drawing or ellipses drawing as stated in the studies of [16-24]. The concavity based approaches gives good results but in some cases they are computationally very expensive. Watershed based techniques includes all form of watershed algorithm based etc as presented by the studies of [25-30]. Watershed based approach have certain degree of success but in dense clumps it results in over segmentation while in some cases also suffered from the problem of under segmentation. Edges or contour based techniques can gives solution in the form of analyzing split edges and linkages of contours etc as mentioned in the works of [31-32]. This approach working well but required model based on some templates and complex both in execution as well as in implementation. Model based approach gives various models in the form of circles through various theories like Gestalt, geometrical theories etc as presented in the work of [35-37]. The problem in this approach seems to be unrealistic as due to its highly complex nature and implementation. Also it is too much expensive computationally.

2. Research Method

We performed the experimentations on Microscopic thin blood smear digital images set of 40 images, which are obtained from [38], which are freely available for research purposes. The proposed methodology started with image pre-processing by following some of the steps in the study of [39], then the slide image is checked for clustered Erythrocytes if existed then passed from splitting the clumps and overlapped Erythrocytes because without splitting the accuracy is compromised on the other hand if clustered Erythrocytes not existed then the control is directly transferred to counting the Erythrocytes. This whole process is presented as overall methodology of this study in Figure 1 and its simulated diagram in the form of images is depicted in Figure 2.

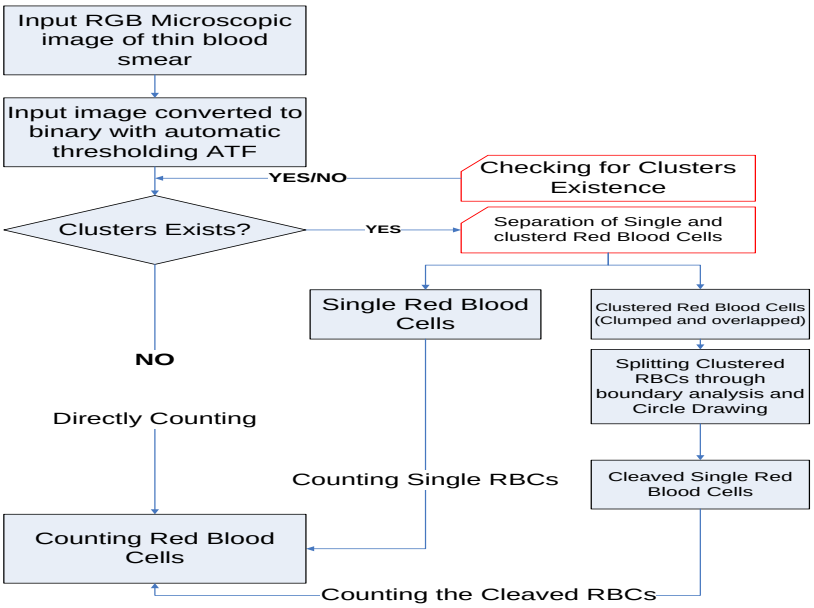

Figure 1. Overall Methodology

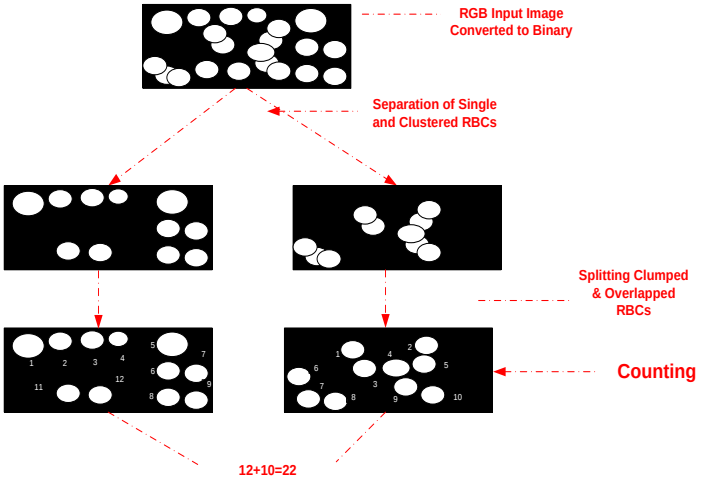

Figure 2. Simulated Images of the whole process

2.1. Image Pre-processing

As image pre-processing we only convert the input RGB image to binary image through Global thresholding OTSU for the purpose to reduce the processing time.[39] After conversion small areas are identified as noise and removed from the binary image and holes in the centres of the Erythrocytes, formed due to haemoglobin in the centres of the Erythrocytes and its similarity to the background are filled and we get the image presented in Figure 3 which is ready for further processing.

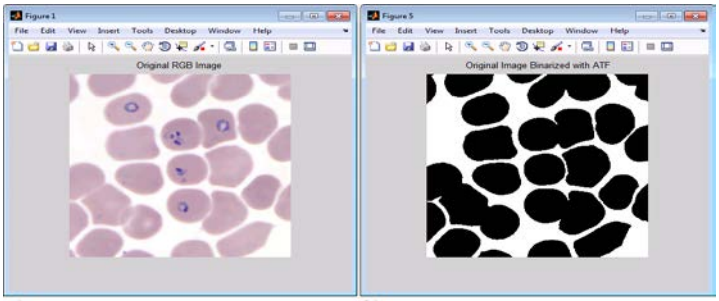

Figure 3. Matlab Results a) Original RGB Image b) Pre-processed Image

2.2. Checking for Clustered Erythrocytes Existence

In checking for clusters of Erythrocytes existence we applied a double check on the convex hulls (through equation 1) of all the Red Blood Cells. We find the areas and elongation of the convex hulls of the Erythrocytes as mentioned in equations1 and 2 respectively. Next between these two measures we find a normalize variance among all the Erythrocytes and empirically through much experimentation we found that if the variance is high 0.2 in case of area and high than 0.5 in case of elongation will be considered as clustered Erythrocytes existed and non-existed otherwise as mentioned in equation 3.

$$\sum_{i=1}^{|X|} \alpha_i x_i | (\forall_i : \alpha_i \geq 0) \& \sum_{i=1}^{|X|} \alpha_i = 1 \qquad (1)$$

where , |X| = finite set of points, x_i is point |X| while $\alpha_i$is weight assigned to x_i, the sum of the weights must be equal to 1 mean normalized.

$$\text{Area= No. of Pixels} \qquad (2)$$

where, No.of Pixels= Pixels defining the convex hull object of Erythrocytes.

$$Elongation = \frac{Length}{Breadth} \qquad (3)$$

where, Length =Major Axis and Breadth = Minor Axis

$$\sigma^2 = \frac{(X-\mu)^2}{N} \qquad (4)$$

where, X represents the area in one case while elongation in the other case, N is the number of terms in distribution.

2.3 Separation of Single and Clustered Erythrocytes

Once this is decided that clusters of Erythrocytes existed then the next step is to separate them from Single Erythrocytes for the purpose to improve efficiency. In the same way in the separation we again applied the double check mentioned in equations 2 and 3, while here the measure of central tendency chosen is “median” in both the cases mentioned in equation 4. We consider median among many central tendency measure for the purpose that the median is the best central tendency measure in case when the data values are irregular and having some small while some large values. We divide the area of every convex hull of Erythrocytes with the median area, the result obtained if equals to 1 or near to 1 are considered as single Erythrocytes and are considered for mask of single Erythrocyte while the negation of the single Erythrocytes resulted in multi- Erythrocytes mask. Then we pass the single Erythrocytes mask to the pixel IDX list of the input image and obtained the image of single Erythrocytes while on passing the multi-mask we obtained the image of clustered Erythrocytes. In the same way we performed for the second check but instead of area we used elongation here.

2.4 Splitting Clustered Erythrocytes

After separation of single and clustered Erythrocytes the actual work of splitting the clustered Erythrocytes starts. In this regard we adopted the approach of boundary analysis. For increasing the accuracy we consider the convex hulls of the clustered Erythrocytes and we calculated the convex hulls of all the clustered Erythrocytes whether, it is clumped or overlapped with equation 1 and the conceptual design of this approach is presented in the simulated diagram depicted in Figure 4.

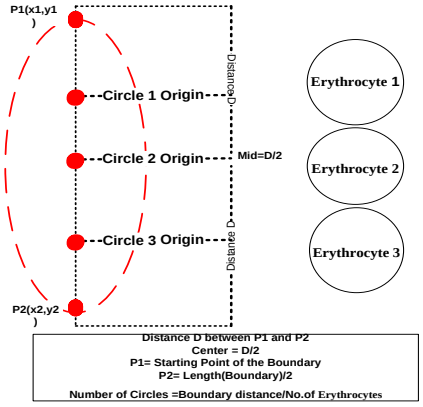

Figure 4. Coceptual Simulated Design

After finding the convex hulls next we find out the boundaries of every clustered Erythrocyte as shown in Figure 4. We divide the boundaries of every clustered Erythrocyte into two halves using equation 5 and measure the distance between the points P1 and the mid-point P2 with the equation 6 and divide the boundary into parts using equation 7.

$$P2 = \frac{Length(Boundary)}{2} \quad (5)$$

where, boundary is the boundary of clumped or overlapped Erythrocytes and index is the index of boundary containing its points.

$$D = \sqrt{\frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2}} \quad (6)$$

$$No.of.Parts = \frac{D}{No.ofRBCs} \quad (7)$$

where, Number of Erythrocytes, we can found while dividing the convex hull area by the median area of single Erythrocyte. After division of boundary into parts next we draw the circle by finding a mid-point in the two consecutive points and using the mid-point circle algorithm with 4-way symmetry separated from each other shown Figure 5.

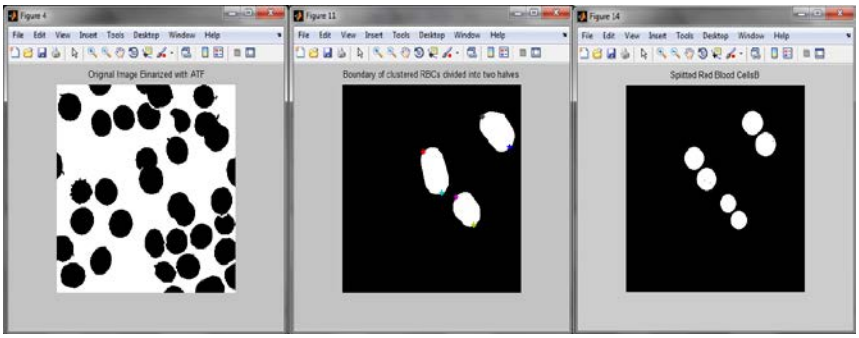

Figure 5. Matlab Results of the process

In Figure 5 a) Presents the original binary image having clustered Erythrocytes b) Presents the Convex hulls of the clustered Erythrocytes and also highlights the Points P1 and P2 further division is based on the number of Erythrocytes in the cluster and distance calculated between P1and P2. Finally, c) presents the circles which cleaved clustered Erythrocytes.

2.5 Counting of Erythrocytes

Once the Clustered Erythrocytes are cleaved into single Erythrocytes then it is not difficult to count them. Thus for counting we consider the Matlab built-in function bwlabel, which uses a binary image and produces a label matrix L having value 0 for the background pixels while gives greater integer values than 0 according to the number of objects in a fashion that assign 1 to the first object, assign 2 to the second object and in this way increase the number according to the number of objects in an arbitrary order and for numbering the Red Blood Cells we followed the study of [40].

3 Results and Analysis

In this section we analyzed the results through visual inspection and statisticall compared with ground truth made by experts, on microscopic thin blood smear digital image dataset of 40 images obtained from [37].

3.1 Analysis through Visual Insecption

In this analysis we divide the images into two groups i.e. images having clustered Erythrocytes and images without clustered Erythrocytes for presentation through visual inspection with ground reality.

3.1.1 Results without Clustered Erythrocytes

The thin blood smear digital images having no clustered Erythrocytes are directly counted after conversion to binary and are by passed from clumps and overlapped or clustered Erythrocytes Splitting as the results are presented in Figures 6,7and 8.

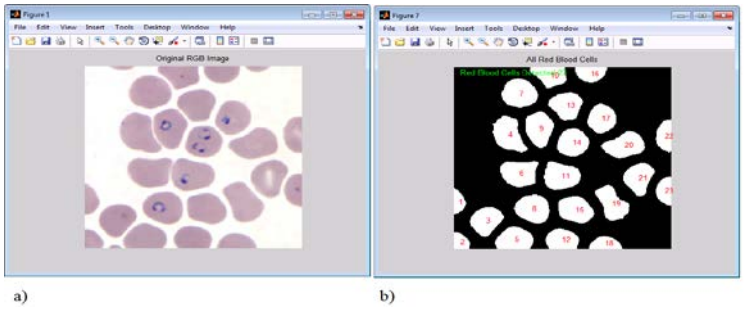

Figure 6. a) Presents Original RGB image b) Presents Counted result of Erythrocytes

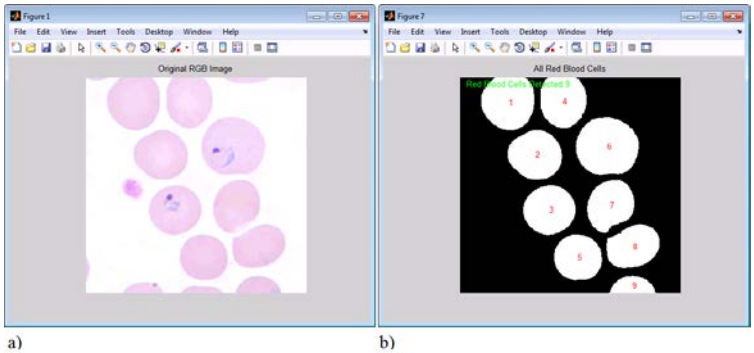

Figure 7. a) Presents Original RGB image b) Presents counted result of Erythrocytes

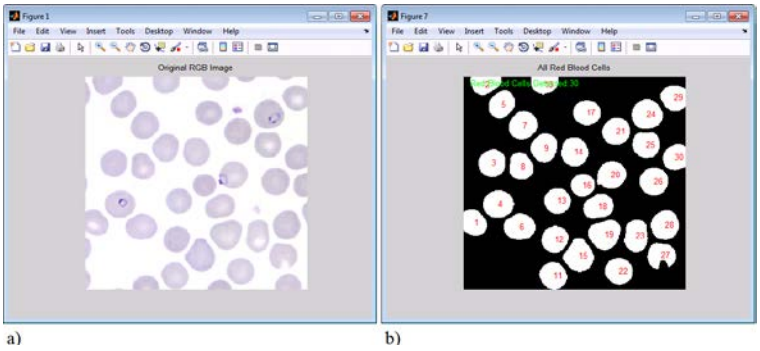

Figure 8. a) Presents Original RGB image b) Presents counted result of Erythrocytes

All the results presented in this category are on the images without clustered Erythrocytes. The last figure i.e. Figure 9, is presented that how the accuracy is affected by the clustered Erythrocytes.

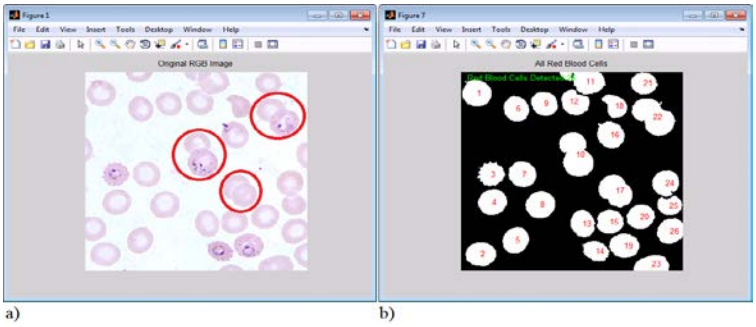

Figure 9. a) Presents Original RGB image with clustered Erythrocytes b) Presents the binary image having total number of Erythrocytes is equal to 26 while the actual number is 31.

3.1.2 Results with Clustered Erythrocytes Splitting

In this category we performed the experimentation on the images having clustered Erythrocytes and we successfully cleaved the clustered Erythrocytes to single Erythrocytes and then performed the counting which will increase the accuracy as shown in “Fig. 10” and “Fig. 11”.

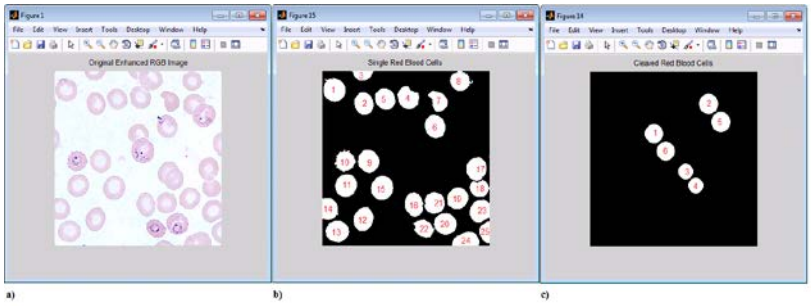

Figure 10. a) Presents Input Original Image, (b) Prsents the counting of single Erythrocytes c) Presents splitting and counting of Clustered Erythrocytes by the proposed method

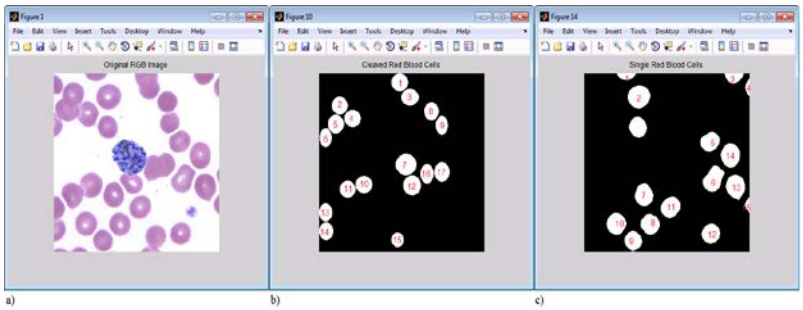

Figure 10. a) Presents Input Original Image, (b) Prsents the counting of single Erythrocytes c) Presents splitting and counting of Clustered Erythrocytes by the proposed method

3.2 Analysis through Statistical Metrics between Automatic and Manual Counting

In this category we compared the results statistically of each slide obtained through manually counted erythrocytes by medical expert using visual inspection and the automatically counted erythrocytes using the proposed method as listed in "Tab. II". Using the confusion matrix presented as "Tab. I" we calculate the sensitivity or True Positive Rate (TPR) or Recall, Accuracy (AC), Error Rate (Er.R) and Specificity or True Negative Rate (TNR) with "(8)", "(9)", "(10)" and "(11)" respectively.

Table 1 Confusion Matrix

Confusion Matrix		Detected	
		Positive	Negative
Actual	Positive	A: True +ve	B: False –ve
	Negative	C: False +ve	D: True –ve

$$TPR = \frac{A}{A+B} \quad (8) \qquad AC = \frac{A+D}{A+B+C+D} \quad (9)$$

$$Er.R = 1 - AC \quad (10) \qquad TNR = \frac{D}{C+D} \quad (11)$$

Table 2. Estimated Results of Manual and Automatic Counting

			Quantitative Analysis			
Slide No.	**Manual**	**Automatic**	**TPR**	**AC**	**Er.R**	**TNR**
1	37	35	0.946	0.972	0.028	0.054
2	55	52	0.945	0.972	0.028	0.055
3	47	43	0.915	0.956	0.044	0.085
4	63	60	0.952	0.976	0.024	0.048
5	55	53	0.964	0.981	0.019	0.036
6	45	43	0.956	0.977	0.023	0.044
7	20	18	0.9	0.947	0.053	0.1
8	105	99	0.943	0.971	0.029	0.057
9	35	33	0.943	0.971	0.029	0.057
10	26	25	0.962	0.98	0.02	0.038
11	19	19	1	1	0	0
12	23	21	0.913	0.955	0.045	0.087

	Quantitative Analysis					
Slide No.	**Manual**	**Automatic**	**TPR**	**AC**	**Er.R**	**TNR**
13	29	23	0.793	0.885	0.115	0.207
14	59	55	0.932	0.965	0.035	0.068
15	20	17	0.85	0.919	0.081	0.15
16	51	46	0.902	0.948	0.052	0.098
17	40	39	0.975	0.987	0.013	0.025
18	29	27	0.931	0.964	0.036	0.069
19	97	90	0.928	0.963	0.037	0.072
20	81	70	0.864	0.927	0.073	0.136
21	48	46	0.958	0.979	0.021	0.042
22	77	72	0.935	0.966	0.034	0.065
23	73	73	1	1	0	0
24	43	41	0.953	0.976	0.024	0.047
25	22	22	1	1	0	0

3.3 Analysis based on Correlation between Manual and Automatic Counting

Here we study the realiablity of the proposed technique through cheking the strength of its relationship with manual methods made by the expert. We presented the graphical comparison shown in Figure 11 on the basis of Linear Pearson's correlation co-efficient which shows strong correlation as R^2=0.992.

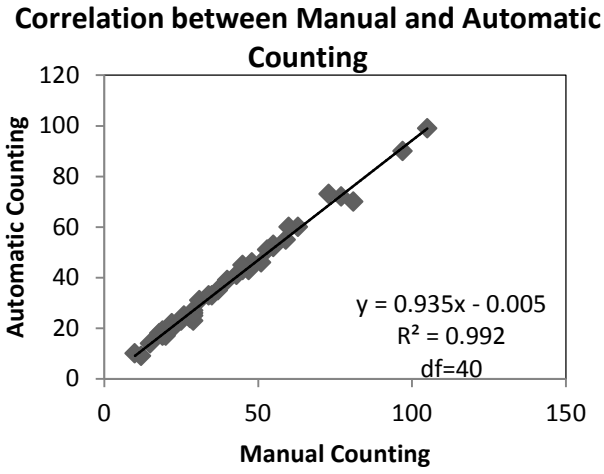

Figure 11 Correlation between Manual and Automatic Counitng

4. Conclusion

Counting of erythrocytes is challenging job in the case when the slides have clumped and overlapped or clustered erythrocytes which the proposed method deal in robust and efficient way by first checking for clustered erythrocytes if they exist then trimming the time by separation of clustered erythrocytes from single erythrocytes and passed only the clustered erythrocytes from splitting algorithm, in simple, effective and efficient way. On the other hand if no clustered erythrocytes existed, directly counting has been started. Moreover, the proposed method

achieved the overall average True Positive Rate of 95% and accuracy of 97% while in contrast to TPR and AC the proposed method achieved the overall True Negative Rate of 5% and Error Rate of 3% that are valuable achievements in the field. As a future suggestion there is still space to increase the accuracy by smoothing the boundaries of the erythrocytes when the convex hulls of the erythrocytes are obtained.

References

[1] National Institute of Health.Red Blood Cells Count. Retrieved on 27-4-2014 from: http://www.nlm.nih.gov/medlineplus/ency/article/003644.htm.

[2] Abbas N, Mohamad D. Automatic Color Nuclei Segmentation of Leukocytes for Acute Leukemia. *Research Journal of Applied Sciences, Engineering and Technology*. 2014 12-04-2014;7(14):2987-93.

[3] Abbas N, Muhamad D. Accurate Red Blood Cells Automatic Counting In Microscopic Thin Blood Smear Digital Images. *Science International Journal.* 2014;26(3):1119-24.

[4] Hartati S, Harjoko A, Supardi TW. The Digital Microscope and Its Image Processing Utility. *TELKOMNIKA Telecommunication, Computing, Electronics and Control*. 2013;9(3):565-74.

[5] Berge H, Taylor D, editors. *Improved red blood cell counting in thin blood smears*. IEEE International Symposium on Biomedical Imaging: From Nano to Macro,Chicago, 2011:204-207.

[6] Buttarello M, Plebani M. Automated blood cell counts state of the art. *American Journal of Clinical Pathology*. 2008;130(1):104-16.

[7] SHAIKH O, BHAT N, Automated Red Blood Cells Count [MS Thesis]. Mumbai, University of Mumbai; 2013-2014.

[8] Nguyen N-T, Duong A-D. Cell splitting with high degree of overlapping in peripheral blood smear. *Int J Comp Theory Eng.* 2011;3(3): 473-478.

[9] Mahmood NH, Lim PC, Blood Cells Extraction Using Color Based Segmentation Technique. *International Journal of Life Sciences Biotechnology and Pharma Research.* 2013; 2(2):78-83.

[10] Grietoinfo; [Cited 12-4-2014] Available from: http://grietinfo.in/projects/MAIN/BME2013/cd-8-project%20report_1_.pdf.

[11] Mahmood NH, Mansor MA. Red blood cells estimation using Hough transform technique. *Signal & Image Processing: An International Journal (SIPIJ)*. 2012;3(2):53-64.

[12] Ramesh N, Salama ME. *Segmentation of haematopoeitic cells in bone marrow using circle detection and splitting techniques.* IEEE International Symposium on Biomedical Imaging (ISBI). Barcelona. 2012: 206-209.

[13] Buggenthin F, Marr C. An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. *BMC bioinformatics*. 2013;14(1):297-308.

[14] Prasad K, Winter J.Image analysis approach for development of a decision support system for detection of malaria parasites in thin blood smear images. *Journal of digital imaging*. 2012;25(4):542-9.

[15] Kumar A, Tembhare P, Enhanced Identification of Malarial Infected Objects using Otsu Algorithm from Thin Smear Digital Images. *International Journal of Latest Research in Science and Technology.* 2012; 1(2):59-63

[16] LaTorre A, Nanclares L. Segmentation of neuronal nuclei based on clump splitting and a two-step binarization of images. *Expert Systems with Applications.* 2013;40(16):6521-30.

[17] Tafavogh S, Navarro KF. *Segmenting Neuroblastoma Tumor Images and Splitting Overlapping Cells Using Shortest Paths between Cell Contour Convex Regions*. Artificial Intelligence in Medicine: Springer Berlin. 2013; 171-5.

[18] Zhang C, Sun C. editors. *Segmentation of clustered nuclei based on curvature weighting*. Proceedings of the ACM 27th Intl. Conference on Image and Vision Computing. New Zealand. 2012: 49-54.

[19] Wang H, Zhang H, editors. *Clump splitting via bottleneck detection*. 18th IEEE International Conference on Image Processing (ICIP). Burssels.2011: 61-64.

[20] Kumarasamy SK, Ong S. Robust contour reconstruction of red blood cells and parasites in the automated identification of the stages of malarial infection. *Machine Vision and Applications*. 2011;22(3):461-9.

[21] Wen Q, Chang H. *A Delaunay triangulation approach for segmenting clumps of nuclei*. 09th IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI. Bostan. 2009:9-12.

[22] Makkapati VV, Naik SK. Clump splitting based on detection of dominant points from contours. IEEE International Conference on Automation Science and Engineering (CASE).Banglore. 2009:197-201.

[23] Gurcan MN, Boucheron LE. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering.* 2009; 2(2):147-71.

[24] Cloppet F, Boucher A, editors. *Segmentation of overlapping/aggregating nuclei cells in biological images*. IEEE 19th Intl. Conference on Pattern Recognition ICPR. Florida. 2008.: 1-4.

[25] Tulsani H. Segmentation using morphological watershed transformation for counting blood cells. *IJCAIT*. 2013;2(3):28-36.
[26] Ferro L, Leal P. Multinuclear Cell Analysis Using Laplacian of Gaussian and Delaunay Graphs. *Pattern Recognition and Image Analysis: Springer*; 2013: p. 441-9.
[27] Hodneland E, Kögel T. CellSegma MATLAB toolbox for high-throughput 3D cell segmentation. *Source code for biology and medicine*. 2013;8(1):1-24.
[28] Schmitt O, Reetz S. On the decomposition of cell clusters. *Journal of Mathematical Imaging and Vision.* 2009;33(1):85-103.
[29] Schmitt O, Hasse M. Morphological multiscale decomposition of connected regions with emphasis on cell clusters. *Computer Vision and Image Understanding.* 2009;113(2):188-201.
[30] Špringl V. Automatic Malaria Diagnosis through Microscopy Imaging [MS Thesis]. Prague. Czech Technical University In Prague.2009.
[31] Kong H, Gurcan M, editors. *Splitting touching-cell clusters on histopathological images*. IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Chaicgo; 2011: .
[32] Kong H, Gurcan M. Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. *IEEE Transactions on Medical Imaging.* 2011;30(9):1661-77.
[33] Köppen M, Yoshida K. *Gestalt theory in image processing: A discussion paper*. Proceedings of Three-Rivers Workshop on Soft Computing in Industrial Applications. University of Passau. 2007: 1-5.
[34] Jiang H, Ngo C-W. Gestalt-based feature similarity measure in trademark database. *Pattern recognition.* 2006;39(5):988-1001.
[35] Airsang U, Ghorpade V. *Contour feature-point tagging as a mechanism for automated splitting of highly-occluded and dissimilar-sized cells in blood smear images*. IEEE Second International Conference on Image Information Processing (ICIIP);Delhi. 2013: 470-475.
[36] DPDx. Image Databank retrieved on 4-4-2014 from url: http://www.dpd.cdc.gov/dpdx/default.htm
[37] Fan Y, Sang Y. An Improved Image Contrast Assessment Method. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014;12(3):2398-405.
[38] ABBAS N, MOHAMAD D. Microscopic Rgb Color Images Enhancement For Blood Cells Segmentation In Ycbcr Color Space For K-Means Clustering. *Journal of Theoretical & Applied Information Technology*. 2013;54(1):117-25.
[39] Yuniarti A. Classification and numbering of dental radiographs for an automated human identification system. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012;10(1):137-46.