An Fuzzy Based Cervical Overlapping Detection using RFC

Abirami S
Mrs.A.Shiyamala Me..(Ph.D)
Mohamed Sathak Engineering College

Abstract

The automated detection and segmentation of overlapping cells remains one of the most challenging problems in the analysis of microscopic images. In particular, these automated systems must detect and accurately segment both the nucleus and cytoplasm of each cell, even when they are clumped together and hence partially occluded. However, this is an unsolved problem due to the poor contrast of cytoplasm boundaries; the large variation in size and shape of cells. Our approach consists of three main steps 1)preprocessing 2)fuzzy c means clustering for segmentation 3) feature extraction and SVM classification. We test our implementation on both real and simulated cervical cell images, containing an assortment of cells and configurations that often present occlusion and/or poor contrast. Our results show both qualitative and quantitative assessment of the datasets, using a completely automated computer program. MATLAB tool has been used to evaluate a performance

I INTRODUCTION

Cervical cancer is a preventable disease, and, unlike most other cancers, it can be easily detected by a routine screening
test. Currently, cervical smear screening is the most popular method used to detect the presence of abnormal cells developing in the cervix. The method is as follows: Using a small brush, a cotton wad stick, or a wooden stick, a specimen is taken from the uterine cervix, smeared onto a thin rectangular glass plate—a slide, and colored, making it easier to examine the cells under a microscope. The purpose of the smear screening is to diagnose premalignant cell changes before they progress to a cancer.

Cell deposition techniques, such as mono-layer preparations, remove a large portion of blood, mucus and other debris, reduce cell overlap and produce cells that are more likely to occur in a single focal plane. This makes both manual and automated slide analysis faster and easier. Automated slide analysis techniques attempt to improve both the sensitivity and specificity of screening by automatically detecting, segmenting and then classifying individual cells present on a slide.
Therefore, automated slide analysis techniques must be capable of analyzing both free-lying and overlapping cells. In the conventional approach to cell analysis, where the cells are first segmented before features are extracted, this implies that the segmentation technique must be capable of both detecting and segmenting the nucleus and cytoplasm from overlapping cells.

The detection and segmentation of the nuclei and cytoplasm from cervical cells is a well studied problem. Current systems can segment the nucleus and cytoplasm of cervical cells in isolation (i.e., cells without any overlap with other cells), segment overlapping nuclei and segment overlapping nuclei plus the whole region representing the cellular clump. However, only recently has the complete segmentation of overlapping cells been addressed in cervical cytology.
It is worth noting that the segmentation of overlapping cells in cytological images is challenging due to the same issues that affect the manual analysis mentioned previously (number of cells, their variability, occlusion and poor contrast). Therefore, it is important that further research be undertaken on this problem to identify and validate methodologies that can automatically produce precise segmentations of the large number of both isolated and overlapping cells typically present in cytology images.

II RELATED WORKS

This multi-stage algorithm, illustrated in Fig.3, benefits from the dark staining of the nuclear material to identify the nuclei, followed by its use as seeds to detect cytoplasm boundaries based on geometrical constructs.

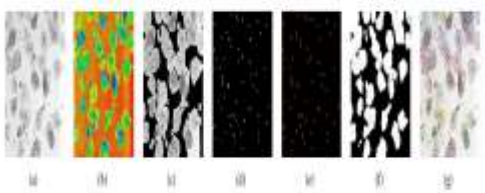


Fig.1. Ushizima, Bianchi and Carneiro's approach. (a) EDF image; (b) Contrast enhancement; (c) Cell clumps; (d) Nuclei candidates; (e) Cytoplasm bands;(f) Voronoi diagram for cytoplasm spliting; (g) Final results.

The proposed method consists of four stages: (1) pre-processing: this step includes the application of the bilateral filter for minimizing intensity noise, followed by brightness improvement using the contrast limited adaptive histogram equalization for improving contrast, which enhances nuclei that often appears brighter due to cytoplasm overlap or staining artifacts (Fig.1(b)); (2) identification of cellular clumps through super pixel definition: this code merges regions based on pixel adjacency and intensity similarity using a graph-based linear-time algorithm , followed by a global search cut-off algorithm on the oversegmented (superpixel) map as shown in Fig.1(c); (3) splitting cell clumps into nuclei (Fig.1(d)) and a rough approximation of cytoplasmastic regions, using a local thresholding algorithm , based on the properties within a sub-window of radius 15 pixels or 5:7_m, which is approximately the diameter of the superficial squamous epithelium cell nucleus; (4) finally, a narrow-band (Fig.1(e)) around each candidate nuclei is used as the seeds for a region growing process that considers both geometric and photometric information about the pixels to identify cytoplasm clumps, followed by an image partitioning into convex polygons through Voronoi diagrams (Fig.1(f)), so that the boundaries dividing neighboring cells have the equal distance to its nearest nucleus. Fig.1(g) shows final result of segmentation, highlighting all nuclei in red, and individual cytoplasm boundaries in different colours. Notice that Fig.1(b) is an intensity-based image, but it appears with pseudocolor to emphasize intensity variations.

Variational Approach for Overlapping Cell Segmentation

The first step consists of the nuclei detection based on the union of the output of MSER and a random decision forest (RF)with non-elliptical connected components excluded (Fig.(b)).Each cytoplasm and its corresponding nucleus are represented as two signed distance maps (SDM), ϕ_i^c and ϕ_i^n. Then an energy functional is optimized with respect to cytoplasms and nuclei level set functions, where this energy functional is a linear sum of the following five terms: (i) the regional term that measures the agreement of an image pixel with background, cytoplasm and nucleus statistical models, where the probability of a given pixel belonging to the background is estimated by training a random forest classifier using the provided ground truth segmentation in the training set (Fig. 4(c)- 4(e)); (ii) the distance prior between the boundary of the cytoplasm and its corresponding nucleus (defined using a spatially varying weight w ensures that the nucleus ϕ_i^n is contained within the cytoplasm ϕ_i^c, , while maintaining a distance of d pixels between them with w being used due to the large variation in the size of the cells and the distance between the nucleus and the cytoplasm (note that in regions with large nuclei density, the distance prior is enforced more strongly to prevent the cytoplasm's contour from growing too far from its nucleus, while in regions with no or sparse nuclei the distance prior is relaxed and the regional term steers the contour, so w = eSDM(all nuclei)=20); (iii) the shape prior term which is defined; (iv) the overlap constraint that limits the overlapping between two neighbouringcytoplasmsandpenalizes the common area between two neighbouringcells,where the cytoplasms that belong to the same clump (Fig.4(f)) are considered as neighbouring cells, and the clumps are detected by thresholding the cytoplasm probability obtainedfrom RF; and (v) the regularization term which ensures smooth boundaries.

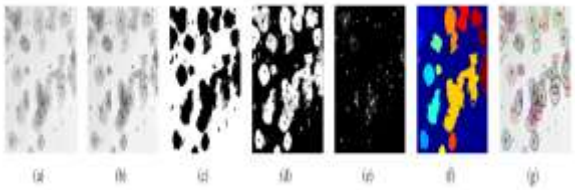


Fig. Nosrati and Hamarneh's approach. (a) EDF image; (b) Detected nuclei; (c) Background probability (pbg(xjI(x)); (d) Cytoplasm probability (pc(xjI(x));
(e) Nucleous probability (pn(xjI(x)); (f) Clumps identification; (g) Final results.

III PROPOSED SYSTEM

The classic K-means clustering algorithm is a deterministic search and may terminate in a locally optimal clustering.

In this project, a genetic K-means clustering algorithm, called GKMCA, for clustering in gene expression datasets is described. GKMCA is a hybridization of a genetic algorithm (GA) and the iterative optimal K-means algorithm (IOKMA).

In GKMCA, each individual is encoded by a partition table which uniquely determines a clustering, and three genetic operators (selection, crossover, mutation) and an IOKM operator derived from IOKMA are employed

1. Choose a number of clusters k

2. Initialize cluster centers 1,… k based on mode

3. For each data point, compute the cluster center it is closest to (using some distance measure) and assign the data point to this cluster.

4. Re-compute cluster centers (mean of data points in cluster)

5. Stop when there are no new re-assignments.

6. GA based refinement

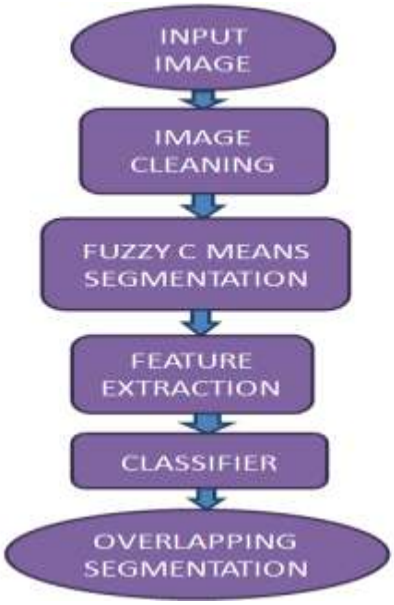
a. Construct the initial population (p1)

b. Calculate the global minimum (Gmin)

c. For i = 1 to N do i. Perform reproduction ii. Apply the crossover operator between each parent. iii. Perform mutation and get the new population. (p2) iv. Calculate the local minimum (Lmin). v. If Gmin<Lmin then a. Gmin = Lmin; b. p1 = p2;

d. Repeat

FUZZY C-MEAN



The fuzzy c-mean algorithm is one of the common algorithms that used to image segmentation by dividing thespace of image into various cluster regions with similarimage's pixels values. For medical images segmentation, thesuitable clustering type is fuzzy clustering. The Fuzzy c-means(FCM) can be seen as the fuzzified version of the k-means algorithm. It is a clustering algorithm which enables data item to have a degree of belonging to each cluster by degree of membership. It's developed by Dunn and changed by Bezdek. The algorithm is an iterative clustering method that produces an optimal c partition by minimizing the
weighted within group sum of squared error objective function. Is widely used in image segmentation and patternrecognition.

In the following section we provide the improved fuzzy cmean algorithm:
Step1: Let H represent the frequency of each item in Data.
Step2:create vector I=min(Data):max (Data)
Step3:Choose random centroid at least 2.
Step4:Compute membership matrix:

$$U_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{|I_i - c_j|}{|I_i - c_k|}\right]^{\frac{2}{m-1}}} \qquad (3)$$

Step5: calculate the cluster center:

$$C = \frac{\sum_{i=1}^{n} U^m * H * I}{\sum_{i=1}^{n} U^m * H} \qquad (4)$$

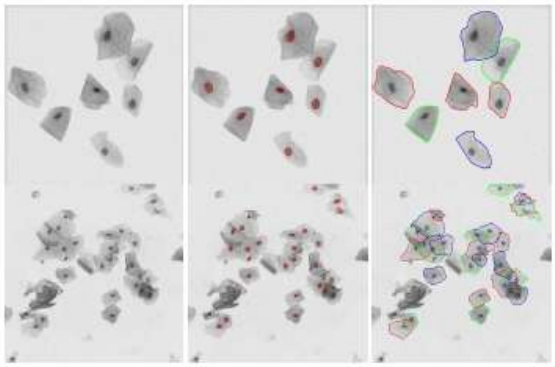
Step6: if $C^{(k-1)} - C^k < \varepsilon$ then Stop else go to Step4.

The improved fuzzy c-mean use values that represent the frequency of items instead of actual values, in gray images the number of values of it may be reached to 256*256=65,536 and that is will take more time in processing, but in improved algorithm will take, at worst case, 256 item to process it. The proposed algorithm does not depend on whole data of image, it actually depends on data that represent the frequency of each data item in original image's data. A number of frequencies at most is 256.

NUCLEUS EXTRACTION

In nucleus extraction, the nuclei of cells are extracted to locate individual cells in the cell mass. Since the cell nuclei are usually darker than the cytoplasm, as shown in Fig. 1, we extract nuclei by local

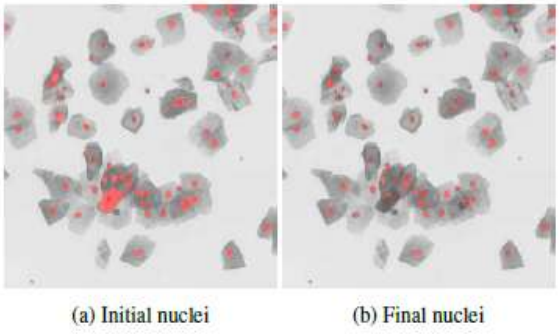
thresholding . In local thresholding, the threshold T for each superpixel is determined as $T = \mu(1+p \exp(-q\mu)+k(\sigma/s-1))$, where μ, σ are mean and standard deviation of local window intensities, respectively, and $p = 2$, $q = 10$, $k = 0.25$, $s = 1$ are weighting variables. The local window for each superpixel is defined as a group of superpixels whose distances between the centers are less than $d = 0.1 \times l$ where l is image width. With the locally defined threshold T, the superpixels which mean intensities are less than T, are classified as initial nucleus candidates. Initial nucleus candidates regularly consist of relatively dark spots, however, include also non-nucleus outliers.



(a) Input images (b) Nuclei (c) Cytoplasm

Figure 1: Examples of microscopic images for overlapping cervical cells, and nucleus and cytoplasm segmentation.

Thus, From the nucleus candidates, outlier removal is performed by (1) breaking superpixel clusters, (2) rejecting tiny superpixels, and (3) rejecting nucleus candidates with low circularity. First, clustered superpixels as shown in Fig. 5 (a), are reduced by rejecting the brightest superpixel iteratively. Since the nucleus often contains at most two superpixels, a superpixel cluster containing more than three superpixels can be considered to have outlier candidates in it. Superpixel clusters are extracted by connected component labeling and for each cluster containing more than three superpixels, a superpixel with the highest mean intensity is rejected. The cluster extraction and superpixel rejection are iteratively repeated until no clusters are extracted. Second, the superpixel candidates with too small sizes are rejected. In experiments, the superpixels with less than 30 pixels are rejected from the candidates. Finally, the candidates with low circularity are rejected since the nucleus has circular shape. The circularity ρ of a candidate is computed as $\rho = 4\pi A/P2$ where A, P are the area and the perimeter of a candidate region, respectively. ρ is known to be 1 for perfect circle and much less than 1 for starfish shape. In our approach, the candidate superpixels with $\rho < 0.5$ are rejected. Throughout the outlier removal, fi- nally extracted nuclei consist of most of cell nuclei without outlier regions as shown in Fig. 5 (b). The extracted nuclei are used as cell indicators in cytoplasm segmentation.



(a) Initial nuclei (b) Final nuclei

Figure 5: Nucleus extraction; (a) Initial nucleus candidates by local thresholding, and (b) finally extracted nuclei by outlier removal.

FEATURE EXTRACTION

In image processing terminology, a feature refers to information chosen from the image, pertaining to the application. It may be a structural element or a quantifiable attribute of image. Morphological features like shape, size and textural features as first order, second order and third order moments of intensity, Gray Level Co-occurrence Matrix (GLCM) are extracted in this paper. Feature extraction involves reducing the amount of details required to describe a large set of data usually and especially an image in this context. Morphological and location features are calculated from the images of nucleus and cytoplasm.

RANDOM FOREST:

Random forest (or **random forests**) is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees. The term came from **random decision forests** that was first proposed by Tin Kam Ho of Bell Labs in 1995.The method combines Breiman's "bagging" idea and the random selection of features.

Random forests (RF) is very efficient statistical classifier introduced by Breiman which is based on ensemble learning for solving classification and regression problems. This classifier is based on bagging method in which each successive tree is constructed independently by taking bootstrap sample of the dataset. Random forests add a new layer of

randomness to bagging where predictors are chosen randomly using the best among a subset strategy at given node to split a node of the tree. The classifier is robust against over-fitting and lack of generalization problems in the datasets and performs well as compared to other classifiers. Another advantage of Random Forests is that it requires only two parameters, i.e. number of variables to be chosen randomly for each subset at given node and the number of trees in the forest.

The advantages of random forest are:

- It is one of the most accurate learning algorithms available. For many data sets, it produces a highly accurate classifier.
- It runs efficiently on large databases.
- It can handle thousands of input variables without variable deletion.
- It gives estimates of what variables are important in the classification.
- It generates an internal unbiased estimate of the generalization error as the forest building progresses.
- It has an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing.

ALGORITHM:

Each tree is constructed using the following algorithm:

1. Let the number of training cases be *N*, and the number of variables in the classifier be *M*.
2. We are told the number *m* of input variables to be used to determine the decision at a node of the tree; *m* should be much less than *M*.
3. Choose a training set for this tree by choosing *n* times with replacement from all *N* available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.
4. For each node of the tree, randomly choose *m* variables on which to base the decision at that node. Calculate the best split based on these *m* variables in the training set.
5. Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier).

For prediction a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is iterated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction.

VI RESULTS AND DISCUSSIONS

Our proposed results are coded and simulated by using the tool MATLAB R2013a which is shown below

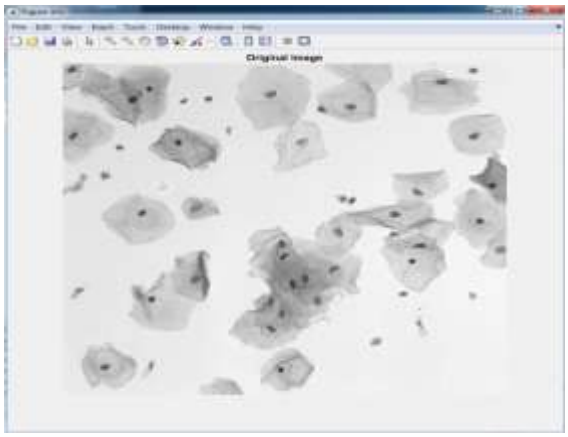


Fig input image

Above fig shows input image for cervical segmentation with rgb color space

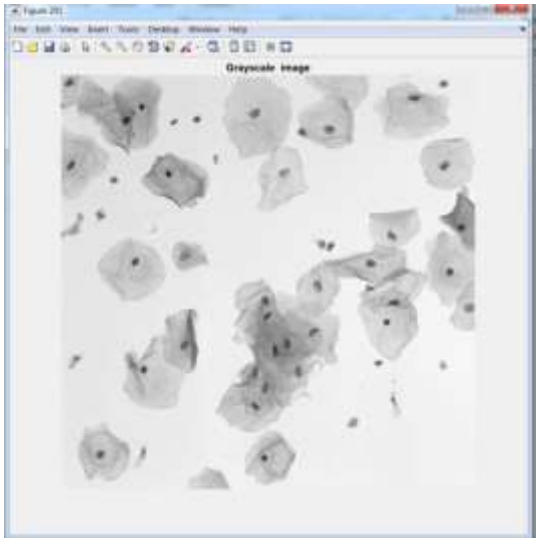


Fig gray scale image

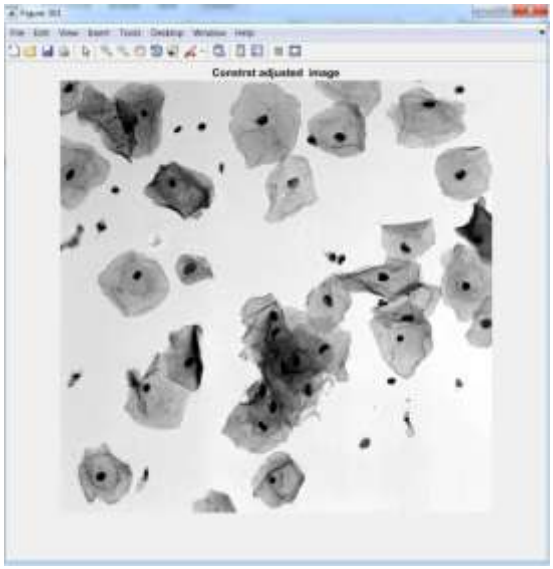


Fig contrast adjusted image

Above fig shows gray scale conversion and contrast adjusted image as preprocessing

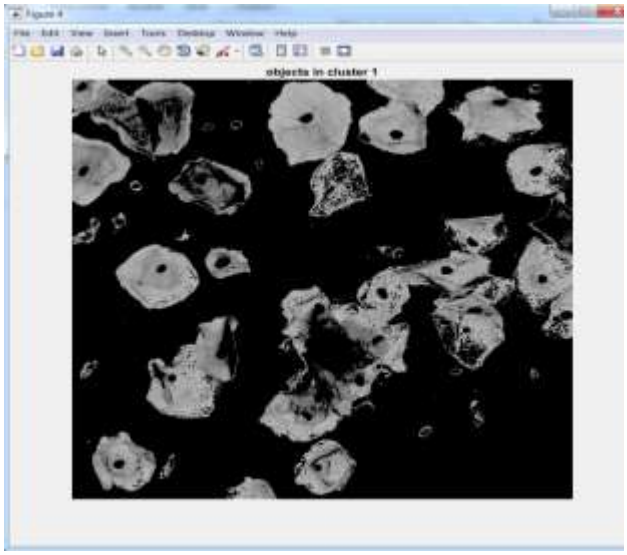


Fig fuzzy cluster 1

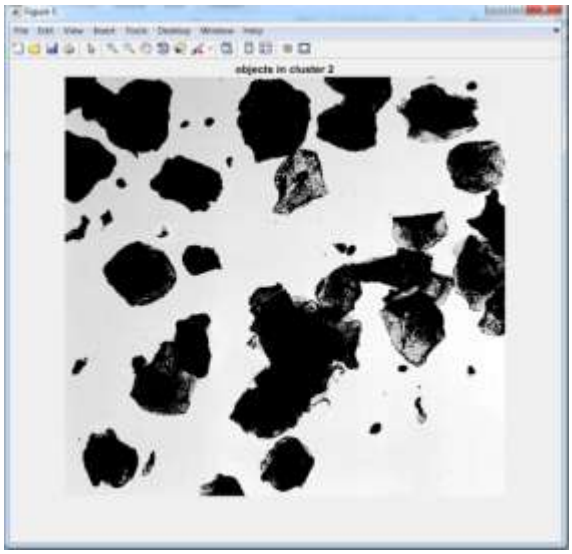


Fig fuzzy cluster 2

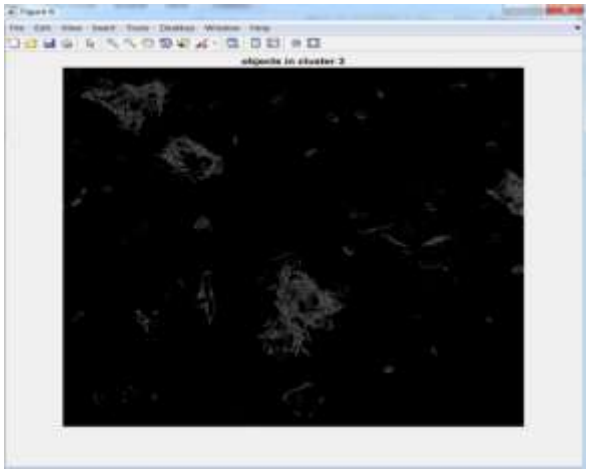


Fig fuzzy cluster -3

Above fig shows fuzzy cluster output images applied in different threshold values

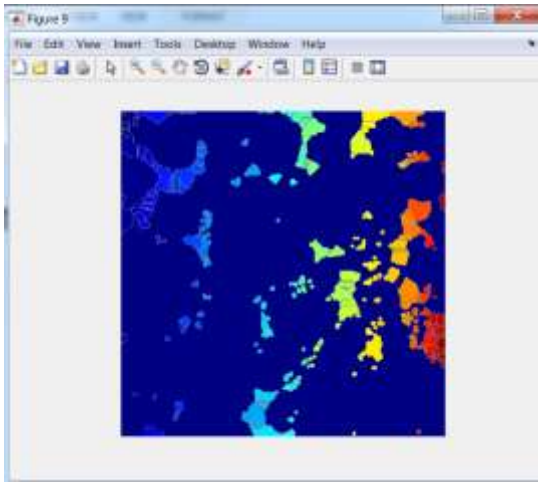


Fig clump identification

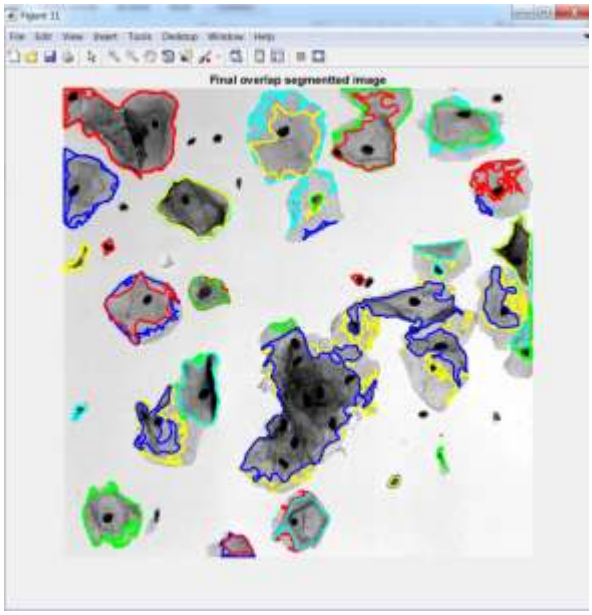


Fig overlap identification

Above fig shows final segmented image by classification

V CONCLUSION

The task of identifying the cell nuclei in conventional Pap smear images is a challenging issue. We have developed a robust and accurate method for the automated segmentation of cells using fuzzy c means clustering and Random forest classifier . The major advantage of the proposed method is that it is fully automated and it is suitable for images with high degree of cell overlapping, as it can successfully detect not only the nuclei of isolated cells, but also the nuclei in cell clusters

REFERENCES

[1] WHO, 2015. http://www.who.int/mediacentre/factsheets/fs380/en/, assessed on October 2015.
[2] Cancer Australia, 2014. http://canceraustralia.gov.au/affectedcancer/cancertypes/gynaecological-cancers/cervical-cancer/cervicalcancer-statistics, assessed on December 2014
[3] Papanicolaou, G. A new procedure for staining vaginal smears. Science1942; 95:438-439.
[4] Genc¸tav, A., Aksoy, S., O¨ nde, S. Unsupervised segmentation and classification of cervical cell images. Pattern Recognition 2012; 45:4151-4168.
[5] Kale, A., Aksoy, S. Segmentation of cervical cell images. International Conference on Pattern Recognition - ICPR 2010; 2399-2402.
[6] Plissiti, M., Nikou, C. A Review of Automated Techniques for Cervical Cell Image Analysis and Classification. Biomedical Imaging and Computational Modelling in Biomechanics. Lecture Notes in Computational Vision and Biomechanics 2013;4:1-18.
[7] Li, K., Lu, Z., Liu, W., Yin, J. Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake. Pattern Recognition2012;45:1255-1264.
[8] Guan, T., Zhou, D., Liu, Y. Accurate Segmentation of Partially Overlapping Cervical Cells Based on Dynamic Sparse Contour Searching and GVF Snake Model. IEEE Journal of Biomedical and Health Informatics2015;19:1494-1504. 38
[9] Yang-Mao, S.-F., Chan, Y.-K., Chu, Y.-P. Edge enhancement nucleus and cytoplast contour detector of cervical smear images. IEEE Trans. On Systems, Man, and Cybernetics, Part B: Cybernetics 2008;38:353-366.
[10] Plissiti, M., Nikou, C., Church anti, A. Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering. IEEE Transactions on Information Technology in Biomedicine2011;15:233-241.
[11] Plissiti, M., Nikou, C. Overlapping cell nuclei segmentation using as partially adaptive active physical model. IEEE Transactions on Image Processing 2012;21:4568-4580.
[12] Vedaldi, A., Soatto, S. Quick Shift and Kernel Methods for Mode Seeking. In: Forsyth, David and Torr, Philip and Zisserman, Andrew, editors. Computer Vision ECCV 2008, Heidelberg: Springer; 2008, p.705 - 718.
[13] Wu, H.-S., Gil, J., Barba, J. Optimal segmentation of cell images. IEEProc. on Vision, Image and Signal Processing 1998;145:50-56.
[14] Matas, J., Chum, O., Urban, M., Pajdla, T. Robust wide baseline stereo from maximally stable extremal regions. Image and Vision Computing -British Machine Vision Computing 2002 2004;22:761767.
[15] Bishop J., Bigner S., Colgan T., Husain M., Howell L., McIntoshK., Taylor D., Sadeghi M. Multicenter Masked Evaluation of Auto Cyte PREP Thin Layers with Matched Conventional Smears. ActaCytolo