MITOSIS DETECTION FOR INVASIVE BREAST CANCER GRADING

Paul A., Mukherjee D. P.

Indian Statistical Institute, Kolkata, India

Abstract

Mitosis count in histologic slide plays the key role in breast cancer grading. We propose a fast and accurate solution for automatic mitosis detection. Cells are detected in a novel manner using area morphological scale space. The scale space is restricted by maximization of cross-entropy between cells and background for precise cell detection. The Random Forest Classifier classifies the detected cells in mitotic and non-mitotic class. Experiments show promising results on a variety of data.

Introduction

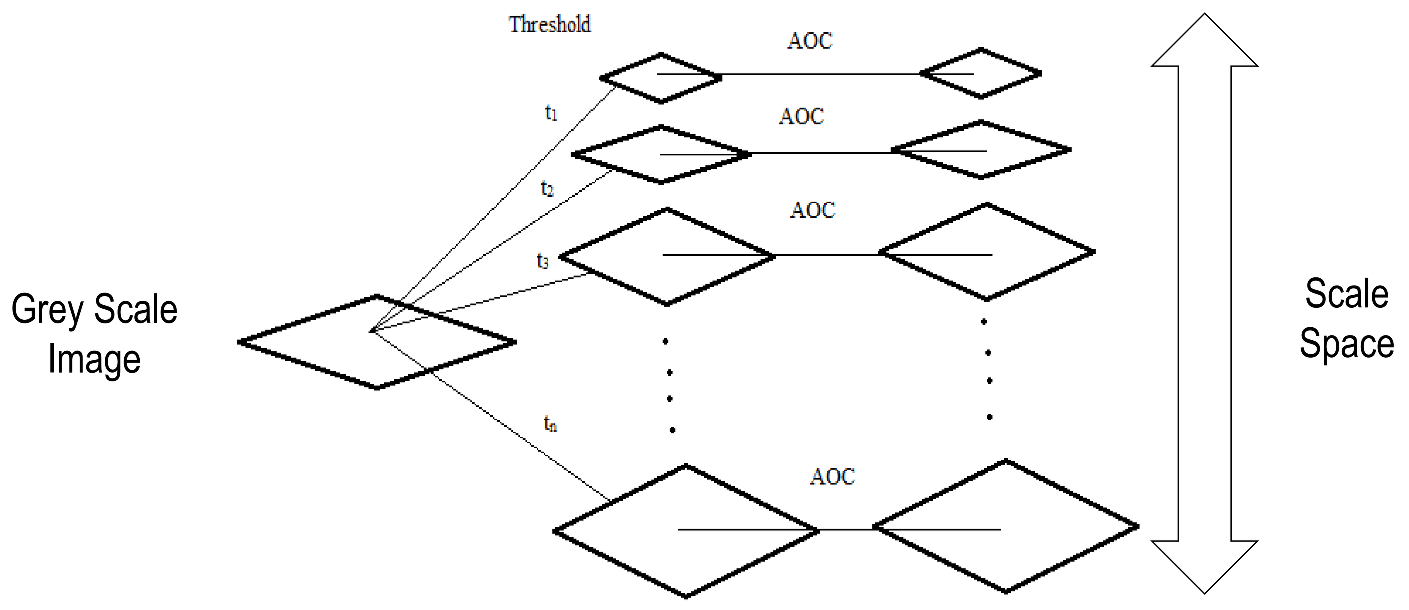
- Mitosis count : A crucial indicator in histologic grading of invasive breast cancer
- Manual Grading: not time efficient and prone to observer variability
- Goal: Automatic cancer grading from histopathological images
- Input: Regions extracted from whole-slide images
- Output: Number of mitotic figures in the region
- Mitotic cells: Various different forms of appearance at different phases
- Non mitotic objects like apoptotic nuclei may look similar to mitotic cells
- So, mitotic cells are difficult to classify precisely

Proposed Methodology

- Detection of the cells from the input image
- No shape or intensity pattern of the cell has been assumed unlike [1]
- No appearance model has been used for cell detection
- Extraction of the cell features from the detected cells
- Classification of the detected cells
- Two class problem: Mitotic and Non mitotic

Cell detection

- Accurate cell detection is very important for proper feature extraction
- This leads to good classification using simple features
- We employ area morphological scale space for cell detection
- Successive application of area morphology operator opening and closing (AOC)
- Scale space: generated by thresholding the grey scale image at different grey levels and applying AOC on them (see figure below)



- Use of the entire range of grey levels for generating the scale space often lead to diffusion of the object boundary [2], and hence inaccurate detection
- So, we propose to use only the grey levels, restricted by maximum cross entropy (between object and background) of the image to generate the scale space
- We define *Inter Class Entropy* between object (ob) and background (bg) as:

$$H_{int} = \sum_g p_g \log \frac{1}{p_g} \; : g = I(x,y) \; \forall \, (x,y) : [(x,y) \in \; ob \; and \; \exists i : \; (x'_i, y'_i) \in bg]$$

(x'_i, y'_i): i^{th} neighbour of (x,y); p_g: Occurrence probability of grey level g

- Clearly, H_{int} is the entropy contributed by the pixels on the edge region
- We define *Within Class Entropy* of object (ob) and background (bg)as:

$$\begin{cases} H_{ob} = \sum_g p_g \log \frac{1}{p_g} \; : g = I(x,y) \; \forall \, (x,y) : \left[(x,y) \in \; ob \; and \; \forall i : \; (x'_i, y'_i) \in ob\right] \\ H_{bg} = \sum_g p_g \log \frac{1}{p_g} \; : g = I(x,y) \; \forall \, (x,y) : \left[(x,y) \in \; bg \; and \; \forall i : \; (x'_i, y'_i) \in bg\right] \end{cases}$$

- H_{ob}, H_{bg} are the entropies, contributed by the pixels of object and background respectively
- For accurate object detection: maximize the separation information between *ob* and *bg*
- So, reduce H_{ob} and H_{bg} (Similarity information): Requires smoothing of *ob* and *bg* region
- Preserve H_{int}(Separation information): Requires preservation of edge region
- *ob*, *bg*: regions of low intensity variance; Edge: regions of high intensity variance
- Image update equation (I^t : image at time *t*):

$$I^{t+1}(x,y) = I^t(x,y) - \lambda \exp\left(-\beta\left\{var\left(L^t(x,y)\right)\right\}t\right)\left(I^t(x,y) - \mu_L^t\right)$$

$$\lambda, \beta : constant; \; L^t(x,y) : \; neighbouring \; region \; of \; (x,y); \; \mu_L^t : mean \; of \, L^t(x,y)$$

- After updating, we find the grey level that maximizes the total *Cross entropy* of the image:

$$i_t = \arg \underbrace{max}_{i} \left\{H_{pq} = \int p(i) \log\left(\frac{1}{q(i)}\right) di\right\}$$

$$p(i) = pdf \; of \; grey \; level \; i : i > i_t$$
$$q(i) = pdf \; of \; grey \; level \; i : i \leq i_t$$

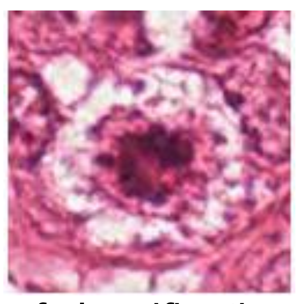
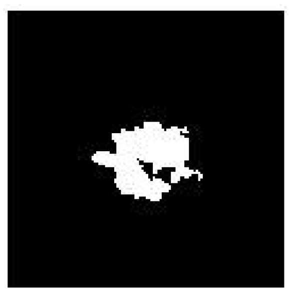
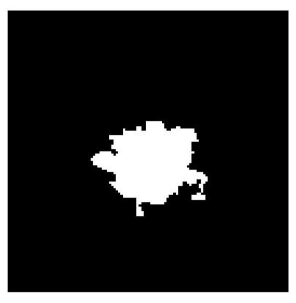
- Scale space is constructed with grey levels $i_t \pm \Delta i$, where Δi is the 3-dB point in entropy-grey level curve
- Objects (cells) are detected based on pixel classification in the scale space images using FCM algorithm

Cell Classification

- Nonlinear classification problem
- Random forest classifier: Nonlinear, less parameters, independent of data size [3]
- Features: colour histogram of the cells (red and blue channel histogram since green channel does not provide distinctive features here)
- Split points at nodes of random tree chosen using maximum information gain
- Features for growing each tree chosen randomly with replacement

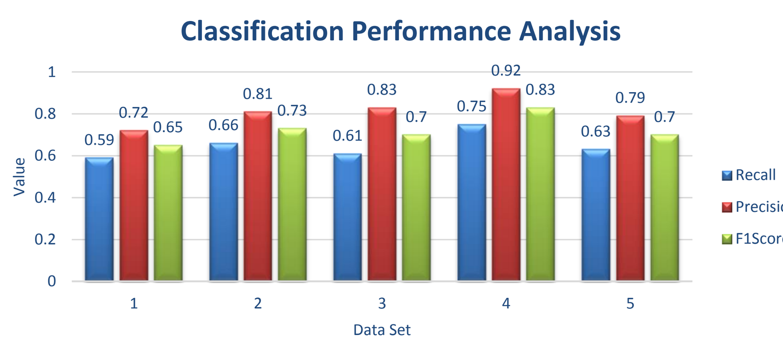
Experiments and Results

- Data : MITOS-ATYPIA 14 [4] data set (Aperio scanner, 40X magnification)
- 5 different data set; Total number of frames:480; Total number of mitosis: 614
- For cell detection, only red channel of the RGB image is used
- Results of cell detection using area morphological scale space with and without entropy restriction: original image (left), with (middle) and without (right) entropy restriction

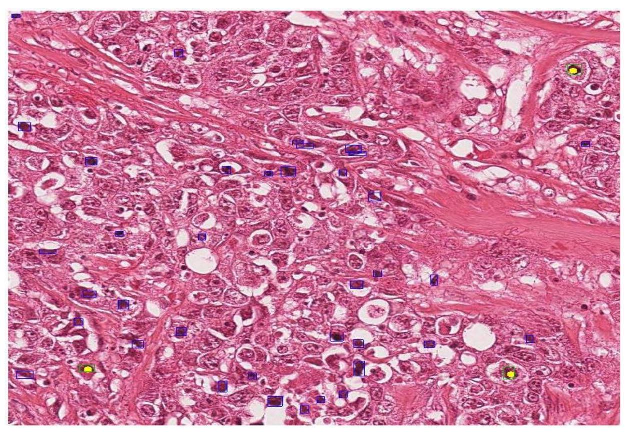
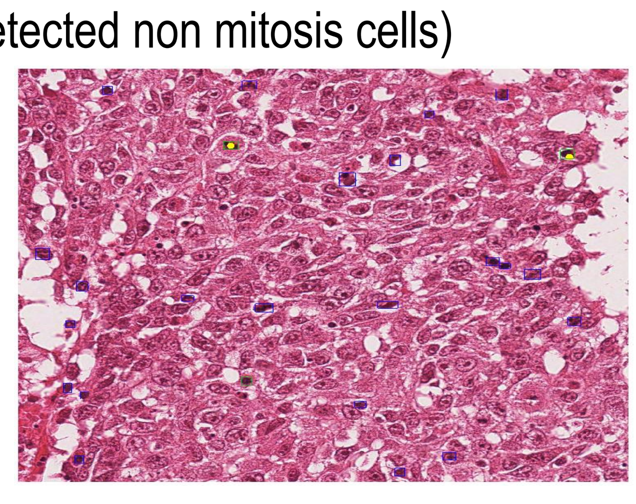
  

- Results of classification:

Data Set	Recall	Precision	F1 score
1	0.59	0.72	0.65
2	0.66	0.81	0.73
3	0.61	0.83	0.7
4	0.75	0.92	0.83
5	0.63	0.79	0.7
Average	0.648	0.814	0.722



- Classification results on the images: Solid yellow region (ground truth), green rectangle (detected mitosis cells), blue rectangle (detected non mitosis cells)

- The average processing time (detection and classification)for each image is 24 sec.

Conclusion

- Mitosis detection algorithm from the images of histopathological slides
- Accurate cell detection: Entropy restricted area morphological scale space
- Cell classification: Random forest classifier with color histogram features
- Fast algorithm and provides better F1 score compared to other existing algorithms

References

1. Pal, N. R. (1996). On minimum cross-entropy thresholding. *Pattern Recognition*, *29*(4), 575-580.
2. Acton, S. T., & Mukherjee, D. P. (2000). Scale space classification using area morphology. *Image Processing, IEEE Transactions on*, *9*(4), 623-635.
3. http://mitos-atypia-14.grand-challenge.org/dataset/
4. Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.