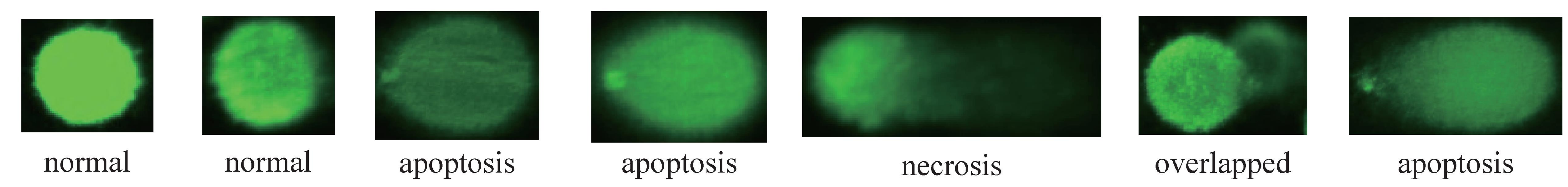
Robust Classification of DNA Damage Patterns in Single Cell Gel Electrophoresis

T. Lee[1], S. Lee[1], W. Sim[2], Y. Jung[2], S. Han[2], C. Chung[2], J. Chang[2], H. Min[3], S. Yoon[1,4,*]

[1]:Electrical and Computer Engineering, Seoul National University, Seoul 151-744, Korea
[2]:Research Division, NanoEnTek Inc., Guro-gu, Seoul 152-740, Korea
[3]:College of Pharmacy, Chung-Ang University, Seoul, 156-756, Korea
[4]:Bioinformatics Institute, Seoul National University, Seoul 151-747, Korea
*: To whom correspondence should be addressed

Introduction

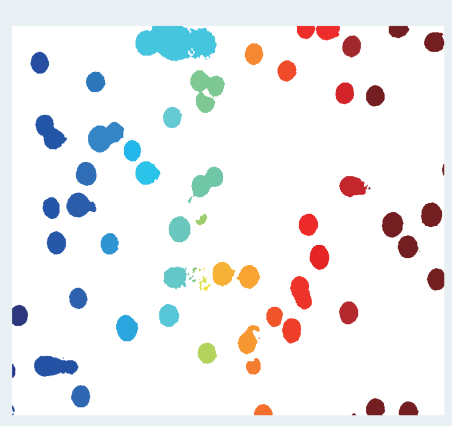


- The **single cell gel electrophoresis** (SCGE) is a method developed for assessing the single cell DNA breakage [1] [2], and each single cell **appear as 'comets.'**
- Test cells can be classified into two types according to their shapes: **normal and abnormal.**
- We propose a novel procedure for analyzing comet assay images, which considers various DNA damage patterns and **classifies them** in a robust manner.

Overview of the proposed methodology

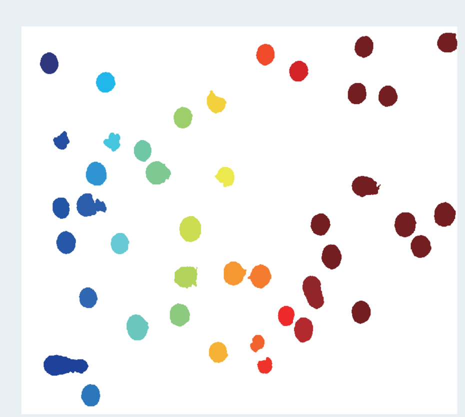
Detection

Removal of scalebar (Cleaning) → Median filtering (Denoising) → K-means clustering (Separation) → Connection of adjacent pixels (Numbering)



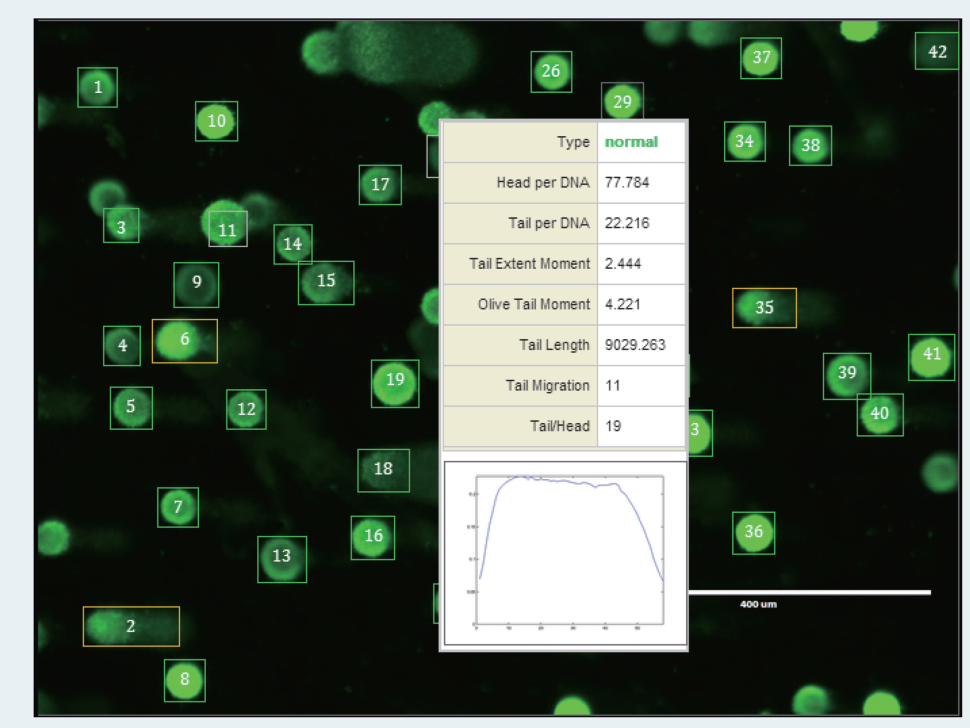
Adjustment

Removal of comets on boundaries → Exclusion of small size comets → Removal of overlaps → Reinclusion of small comets which are close to large ones



Analysis

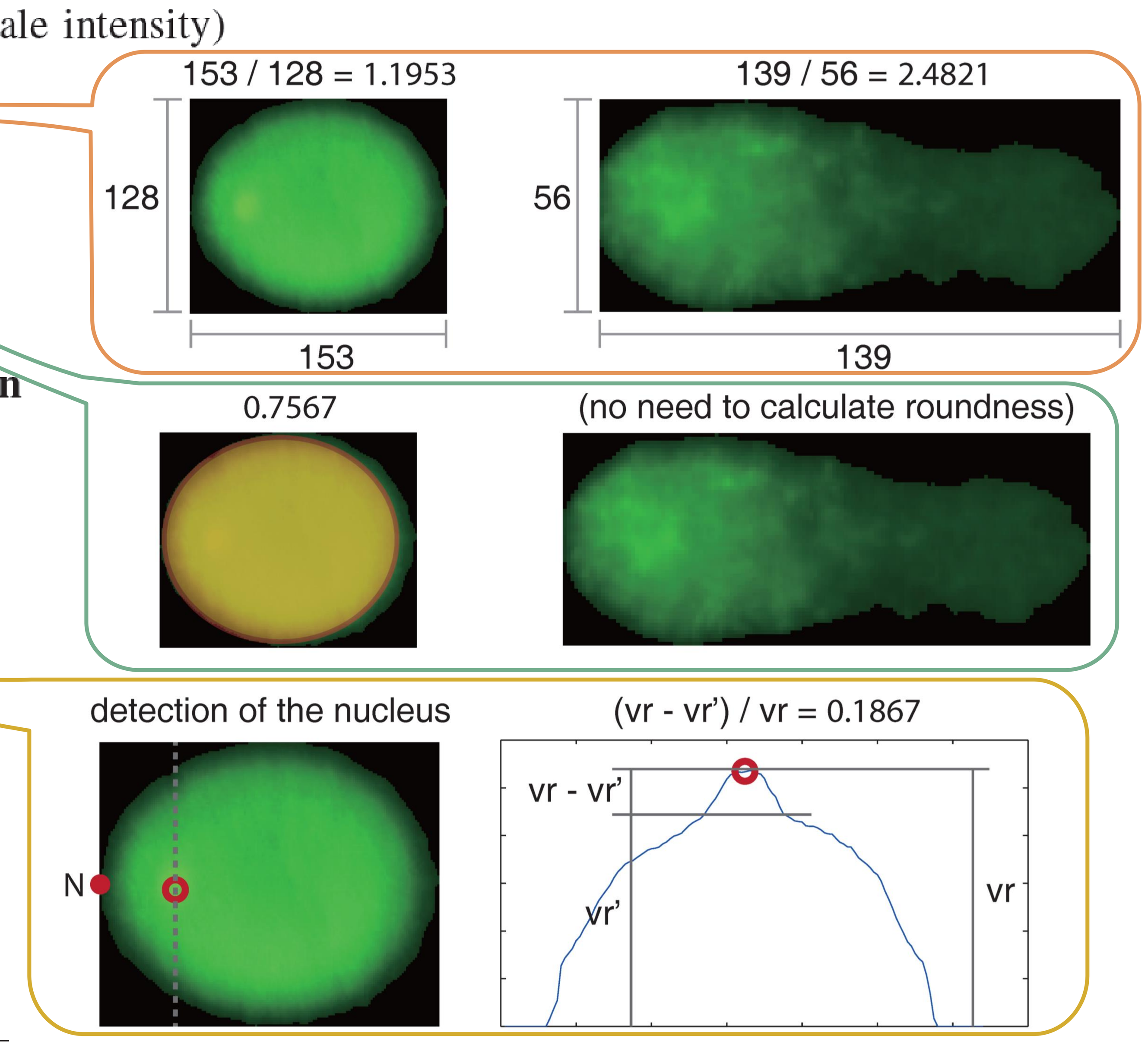
Parameters returned :
Intensity profile
Head/tail per DNA
Surface rate of head/tail
Extent tail moment
Olive tail moment
Tail length
Tail inertia
Tail distance
Center of head
Radius of head
Size and Location
Type

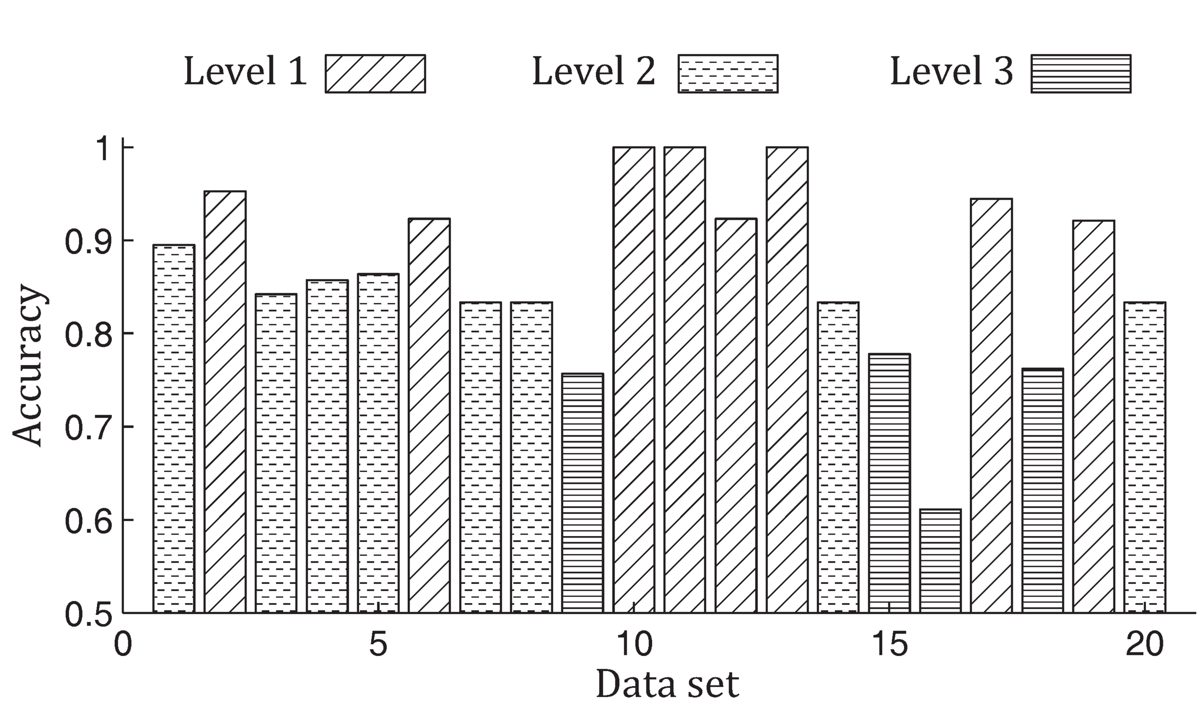


Methods

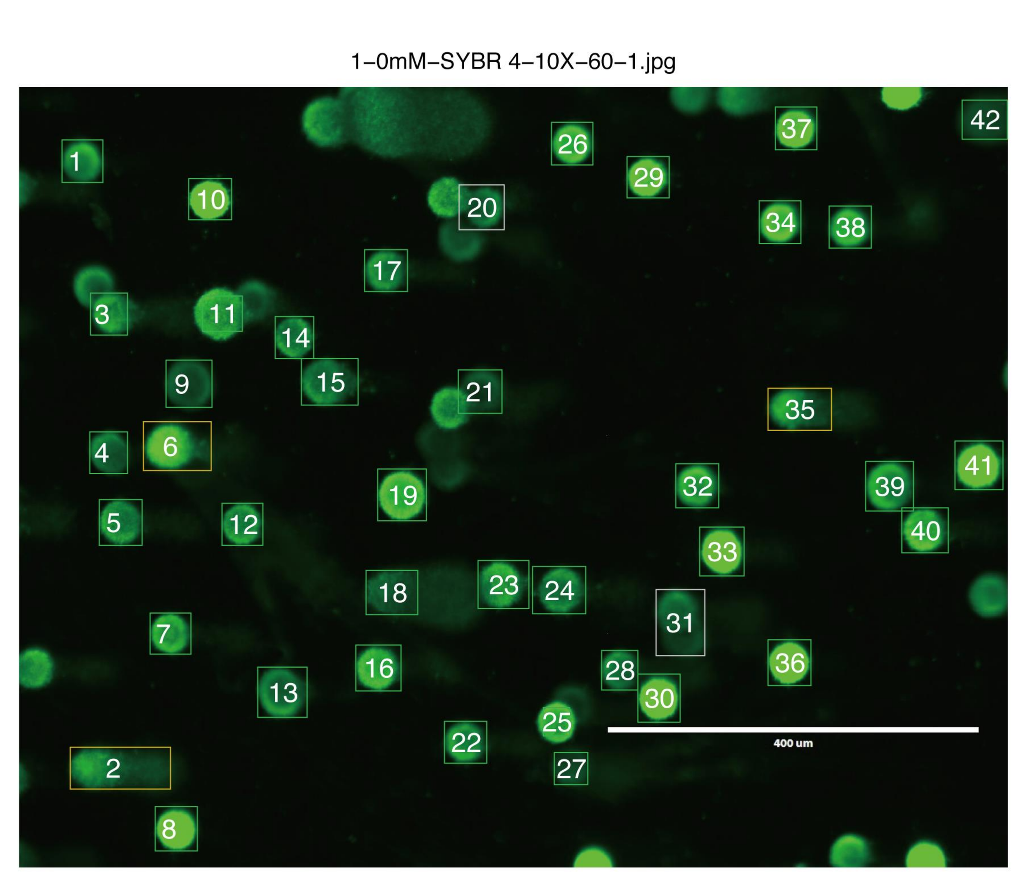
- In the **detection** phase, **image segmentation** is applied to detect comet pixels and identify comets by merging contiguous comet pixels.
- In the second **adjustment** phase, we eliminate **overlapped comets** and the objects on the image boundaries.
- The proposed method characterizes individual cells and then classifies them into two groups in the analysis step.
- To achieve above objectives, we define three parameters, and **the decision-tree-based algorithm** can be described as shown in the left figure.
- **The range of ratio is divided into four intervals:** (A) [0:85, 1.3], (B) [1.3, 4.5], and (C) the others ([0, 0.85] and [4.5, 1]).
- The comets in case of **(B)** are classified as **'abnormal'** cells, because the comets are distributed widely over the x-axis.
- The comets in group **(C)** are classified as **'fail'** comets because their ratio are unrealistic.
- In addition, we classify a comet in **(A)** as **'fail'** if its shape is not a circle, **'abnormal'** if the nucleus is presented on the image significantly, and **'normal'** if a nucleus does not exist.

```
procedure type = GETTYPE(img)
    img : rows × cols matrix (grayscale intensity)
    if ratio < 0.85 then
        type ← 'fail'
    else if ratio < 1.3 then
        if roundness < 0.6 then
            type ← 'fail'
        else
            if peakHeight > 0.12 then
                type ← 'abnormal'
            else
                type ← 'normal'
            end if
        end if
    else if ratio < 4.5 then
        type ← 'abnormal'
    else
        type ← 'fail'
    end if
    return type
end procedure
```





	Level 1	Level 2	Level 3
Average of Accuracy (Standard Deviation)	95.8% (3.41)	84.9% (2.06)	72.7% (6.73)



Results and Conclusions

- We tested with **20 golden data sets,** which were generated by a micro comet-assay system. These comet assay images contain 140 normal and 229 abnormal cells in total.
- Domain experts marked the labels of individual comets, and also categorized the golden data sets into **three groups** according to the **difficulty** of image processing.
- The average classification accuracy achieved was **86.8% for 20 test data sets (over 300 comets)** with varying difficulty levels.
- The proposed procedure aims to handle comet assay images and consists of **three phases: detection, adjustment, and analysis.**
- Our approach is the **first attempt to organize a series of established methods suitable for comet assays.**

References

1. D. W. Fairbairn, P. L. Olive, and K. L. O'Neill, "The comet assay: a comprehensive review," Mutation Research/Reviews in Genetic Toxicology, vol. 339, no. 1, pp. 37–59, 1995.
2. J. H. Hoeijmakers, "Dna damage, aging, and cancer," New England Journal of Medicine, vol. 361, no. 15, pp. 1475–1485, 2009.
3. A. Collins, "The comet assay for dna damage and repair," Molecular Biotechnology, vol. 26, pp. 249–261, 2004.
4. B. Hellman, H. Vaghef, and B. Bostr´om, "The concepts of tail moment and tail inertia in the single cell gel electrophoresis assay," Mutation Research/DNA Repair, vol. 336, no. 2, pp. 123–131, 1995.
5. A. Vlahou, J. O. Schorge, B. W. Gregory, and R. L. Coleman, "Diagnosis of ovarian cancer using decision tree classification of mass spectral data," J. of Biom. and Biot., vol. 2003, no. 5, pp. 308–314, 2003.
6. K. Fu and J. Mui, "A survey on image segmentation," Pattern Recognition, vol. 13, no. 1, pp. 3–16, 1981.
7. S. R. Vantaram and E. Saber, "Survey of contemporary trends in color image segmentation," Journal of Electronic Imaging, vol. 21, no. 4, 2012.
8. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," IEEE T-PAMI, vol. 24, no. 7, pp. 881–892, 2002.
9. X. Jiang and H. Bunke, "Edge detection in range images based on scan line approximation," Computer Vision and Image Understanding, vol. 73, no. 2, pp. 183–199, 1999.
10. J. Canny, "A computational approach to edge detection," IEEE T-PAMI, vol. PAMI-8, no. 6, pp. 679–698, 1986.