A Region Based Predictor For Lossless Compression of RNAi Images

N. Karimi, S. Samavi, S. Shirani, S. Amraee, Z. Safaryazdi, E. Mahmoodzadeh

Department of Electrical and Computer Engineering
Isfahan University of Technology, Isfahan, Iran
McMaster University, Hamilton, Canada
Outline

- Introduction about HCS and RNAi
- Prediction and compression
- Characteristics of RNAi images
- The proposed prediction method
- Implementation and results
- Conclusion
High Content Screening (HCS)

• A powerful tool for the analysis of complex cellular biologic structures

• Triggered advances in the medical diagnosis methods, prognosis techniques and development of new medications

RNA interference (RNAi) using HCS

• Application in molecular biology and further study of the genes

• Scrutinize the behavior of a certain gene by analyzing the absence of that gene in a biological process
RNAi Screening

• fluorescent dyes
• three different channels.

• DNA (The nucleuses of the cells)
• Actin (The cytoplasm)
• Rac (Some auxiliary information)
Sample RNAi Image (Channels)

DNA channel

Actin channel

Rac channel
Challenges in Analysis and Compression

- 400,000 images in one RNAi study (Drosophila flies)
- Automated analysis and large storage space
- Processing of RNAi images (per channel):
  - Segmentation of the cells and identification of the components
    - No grid formation
    - No pre-specified locations
    - No pre-specified number of components
    - Cell overlapping
Compression of the HCS images: Lossy vs Lossless

• **Lossy Methods:**
  - Usually based on psycho-visual characteristics
  - Inexact reconstruction of the original image
  - Better compression
  - Omission of some minor details

• **Lossless Methods:**
  - Exact reconstruction
  - Information sensitive applications such as RNAi.

• Need for efficient compression methods for these images

• Need for the advent of customized compression methods
Prediction and Compression

- Prediction the intensity of each pixel based on the intensity of some previously coded neighboring pixels and compression the residuals

- Use of Prediction methods in most successful lossless image compression algorithms in compression of general images

  - Predictors used in Lossless mode of JPEG (LJPEG)

    - MED (Median Edge Detector) in JPEGLS

    - GAP (Gradient Adjusted Predictor) in CALIC
Predictors used in LJPEG

• Uses 7 different linear predictors

\[(a + b - c), a + (b - c)/2, b + (a - c)/2, (a + b)/2, a, b, c\]

• Selection of predictor with lowest prediction error entropy for each image

MED Predictor

\[\hat{x} = \begin{cases} 
\min(a, b) & \text{if } c \geq \max(a, b) \\
\max(a, b) & \text{if } c \leq \min(a, b) \\
a + b - c & \text{otherwise}
\end{cases}\]
GAP Predictor

- Vertical and horizontal edge detection and determination of the sharpness of an edge:

\[
\begin{align*}
    d_h &= |WW - W| + |NW - N| + |N - NE| \\
    d_v &= |NW - W| + |NN - N| + |NNE - NE|
\end{align*}
\]

If \((d_v - d_h) > 80\) Then \(\{\text{Sharp Horizontal edge}\}\)

\[\hat{X} = W\]

Elseif \((d_h - d_v) > 80\) Then \(\{\text{Sharp Vertical edge}\}\)

\[\hat{X} = N\]

Else

\[\hat{X} = (N + W)/2 + (NE - NW)/4\]

If \((d_v - d_h) > 32\) Then \(\{\text{Horizontal edge}\}\)

\[\hat{X} = (\hat{X} + W)/2\]

Elseif \((d_v - d_h) > 8\) Then \(\{\text{Weak Horizontal edge}\}\)

\[\hat{X} = (3\hat{X} + W)/4\]

Elseif \((d_h - d_v) > 32\) Then \(\{\text{Vertical edge}\}\)

\[\hat{X} = (\hat{X} + W)/2\]

Elseif \((d_h - d_v) > 8\) Then \(\{\text{Weak Vertical edge}\}\)

\[\hat{X} = (3\hat{X} + W)/4\]

End if

End if
Some characteristics of RNAi images

1. **Categorize each Pixel**: foreground, background, and boundary
2. **Background** region in all channels has low intensity values with low fluctuations.
3. **DNA and Rac**: foreground higher intensities and fluctuations than background
4. **Actin foreground**: in places where the nucleus resides in the cytoplasm, a flat region with lower intensities than surroundings.
5. **Wide variety of intensities in boundaries** of all channels.
The proposed prediction method

1. In all three channel images MED predictor produced large errors especially in the boundary regions.

2. In non-boundary regions, the accuracy of MED was not significant.

Goal of proposed prediction method:

1. Preserving the simplicity of MED

2. Using the characteristics of the RNAi images to outperform MED for these specific images.
The proposed prediction method

- Use Otsu instead of complex segmentation methods

\[
\text{Region}(x) = \begin{cases} 
\text{Boundary} & \text{if } T_B \leq a, b, c \leq T_F \\
\text{Non-Boundary} & \text{otherwise}
\end{cases}
\]

\[
\text{Max_flux}(x) = \max (|c - a|, |c - b|, |a - b|)
\]

\[
\text{Min_flux}(x) = \min (|c - a|, |c - b|, |a - b|)
\]
If $\text{Max}_\text{flux}(x) < m$ Then
\[ \tilde{x} = (a + b + c)/3 \]
Elseif $c \geq \max(a, b)$ Then
\[ \text{If Region}(x) = \text{Boundary} \text{ Then} \]
\[ \tilde{x} = \min(a, b) - \alpha \text{Min}_\text{flux}(x) \]
Else
\[ \tilde{x} = \min(a, b) \]
End if
Elseif $c \leq \min(a, b)$ Then
\[ \text{If Region}(x) = \text{Boundary} \text{ Then} \]
\[ \tilde{x} = \max(a, b) + \alpha \text{Min}_\text{flux}(x) \]
Else
\[ \tilde{x} = \max(a, b) \]
End if
Else
\[ \tilde{x} = a + b - c \]
End if
Implementation and results

- 2304 images for each RNAi channel (6912 images in total) human cells
- 512×512, 8 bits image in each channel
- $\alpha = \frac{2}{3}$ \quad $m = 4$

### Bit Per Pixel (BPP).

<table>
<thead>
<tr>
<th>Channel Type</th>
<th>Original Images</th>
<th>GAP</th>
<th>MED</th>
<th>Lossless JPEG Predictors</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>3.45</td>
<td>2.70</td>
<td>2.84</td>
<td>2.92</td>
<td>2.62</td>
</tr>
<tr>
<td>Actin</td>
<td>3.62</td>
<td>2.56</td>
<td>2.67</td>
<td>2.77</td>
<td>2.53</td>
</tr>
<tr>
<td>Rac</td>
<td>1.74</td>
<td>2.06</td>
<td>2.19</td>
<td>2.12</td>
<td>1.91</td>
</tr>
</tbody>
</table>
Implementation and results

Mean and Standard Deviation (STDV) for thresholds of different channels

<table>
<thead>
<tr>
<th>Channel Name</th>
<th>Mean of TB</th>
<th>STDV of TB</th>
<th>Mean of TF</th>
<th>STDV of TF</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>27.32</td>
<td>4.20</td>
<td>131.54</td>
<td>26.20</td>
</tr>
<tr>
<td>Actin</td>
<td>17.67</td>
<td>3.09</td>
<td>53.61</td>
<td>18.37</td>
</tr>
</tbody>
</table>

• Finding the universal thresholds base on a small number of images (such as 50 randomly picked images)

<table>
<thead>
<tr>
<th>Channel Type</th>
<th>Original Images</th>
<th>GAP</th>
<th>MED</th>
<th>Lossless JPEG Predictors</th>
<th>Proposed Method</th>
<th>Proposed Method (universal thresholds for all images)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>3.45</td>
<td>2.70</td>
<td>2.84</td>
<td>2.92</td>
<td>2.62</td>
<td>2.63</td>
</tr>
<tr>
<td>Actin</td>
<td>3.62</td>
<td>2.56</td>
<td>2.67</td>
<td>2.77</td>
<td>2.53</td>
<td>2.55</td>
</tr>
<tr>
<td>Rac</td>
<td>1.74</td>
<td>2.06</td>
<td>2.19</td>
<td>2.12</td>
<td>1.91</td>
<td>1.90</td>
</tr>
</tbody>
</table>
# Implementation and results

<table>
<thead>
<tr>
<th>Channel Type</th>
<th>Original Images</th>
<th>GAP</th>
<th>MED</th>
<th>Lossless JPEG Predictors</th>
<th>Proposed Method</th>
<th>Proposed Method (universal thresholds for all images)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>3.45</td>
<td>2.70</td>
<td>2.84</td>
<td>2.92</td>
<td>2.62</td>
<td>2.63 (BPP)</td>
</tr>
<tr>
<td>Actin</td>
<td>3.62</td>
<td>2.56</td>
<td>2.67</td>
<td>2.77</td>
<td>2.53</td>
<td>2.55 (BPP)</td>
</tr>
<tr>
<td>Rac</td>
<td>1.74</td>
<td>2.06</td>
<td>2.19</td>
<td>2.12</td>
<td>1.91</td>
<td>1.90 (BPP)</td>
</tr>
<tr>
<td>Average Time (DNA + Actin)</td>
<td>0.48</td>
<td>0.23</td>
<td>1.63</td>
<td>0.39</td>
<td>0.30 (sec)</td>
<td></td>
</tr>
</tbody>
</table>

Execution time comparison of the proposed method with MED and GAP

![Execution time comparison graph]
Implementation and results

Residuals for MED

Residuals for proposed method
Conclusion

1. The importance of the HCS and RNAi
2. Critical Need for segmentation and compression of RNAi images
3. Weakness of the current prediction
4. Develop a customized predictor based on specification of RNAi images
5. Thresholding vs Segmentation (Boundary/Smooth region)
6. Incorporate the steep slope of the pixel changes in the boundary regions into the predicted value
7. Comparable complexity but better performance than MED
8. Lower complexity but comparable performance with respect to GAP
Questions ?