AN OPTICAL PHYSICS INSPIRED CNN APPROACH FOR INTRINSIC IMAGE DECOMPOSITION

Harshana Weligampola, Gihan Jayatilaka, Suren Sritharan (presenter)
Parakrama Ekanayake, Roshan Ragel, Vijitha Herath, and Roshan Godaliyadda.
Intrinsic Image Decomposition (IID) is the problem of decomposing an image into its constituents.
IID - Background

• Supervised IID
  • Impractical due to the absence of large datasets with ground truth.

• Unsupervised IID
  • Not robust enough to decompose images with various scene types in different lighting conditions.
  • Techniques based on human vision system (e.g. retinex theory) do not exploit the existing physics understanding of light into account in improving the image decomposition.
Proposed Physics based IID

A novel **unsupervised IID** framework inspired by **optical physics** capable of decomposing a **wide variety** of images captured under different lighting conditions.
The perceived pixel intensity of an image is given by the **Phong’s model**.

The main components of an image are:

1) Ambient light intensity
2) Diffused light intensity
3) Specular light intensity
Phong’s model - Mathematical Formulation

The pixel intensity at point $p$ is,

$$ I_p = \int_\lambda k_a r_p(\lambda)i_a(\lambda) + \sum_{\hat{L}^{(n)} \in L} \{k_d r_p(\lambda)[\hat{L}^{(n)} \cdot \hat{N}_p]i_d^{(n)}(\lambda) + k_s s_p(\lambda)[\hat{R}^{(n)} \cdot \hat{V}]\gamma i_s^{(n)}(\lambda)\} d\lambda $$

- Ambient
- Diffused
- Specular
Phong’s model - Mathematical Formulation

First consider a narrow band ($\lambda_c$),

$$I_p(\lambda_c) = k_a r_p(\lambda_c)i_a(\lambda_c) + \sum_{\hat{L}(n) \in L} \{k_d r_p(\lambda_c)[\hat{L}^{(n)} . \hat{N}_p]i_d^{(n)}(\lambda_c) + k_s s_p(\lambda_c)[\hat{R}^{(n)} . \hat{V}]\gamma i_s^{(n)}(\lambda_c)\}$$

Then assuming:

1) The ambient illumination is constant
2) Only one light source exists
3) Specular term is negligible
4) Ignoring coefficient
Parameters derived from the image

1. **Reflectance Ratio Gradient (RRG)**
   Identify the boundaries of the uniform reflectance in an image.

2. **Reflectance Approximation Map (RAM)**
   Approximate reflectance of an image.

3. **Shading Gradient (SG)**
   Gradient of the shading map of an image.
Reflectance Ratio Gradient

The **log ratio** of pixel intensity between 2 narrow band wavelengths is,

\[
\mathcal{J}_p(\lambda_a, \lambda_b) = \log \left( \frac{I_p(\lambda_a)}{I_p(\lambda_b)} \right) = \log \left( \frac{r_p(\lambda_a)i_d(\lambda_a)}{r_p(\lambda_b)i_d(\lambda_b)} \right)
\]

Consider the gradient

Assuming that single wavelength intensity of adjacent pixels are equal

\[
\nabla \mathcal{J}(\lambda_a, \lambda_b) = \nabla \log \left( \frac{r(\lambda_a)}{r(\lambda_b)} \right) \quad \text{RRG}
\]
If $I_p(\lambda_a) = I_p(\lambda_b)$, then the log ratio $J_p(\lambda_a, \lambda_b) = 0$.

If $I_p(\lambda_a) \gg I_p(\lambda_b)$, then the log ratio $J_p(\lambda_a, \lambda_b)$ will be high and vise-versa.

Based on this the RAM is used to approximate the reflectance ($R$) as,

$$\text{RAM}_R = \frac{\overline{J}_p(\lambda_R, \lambda_G) + \overline{J}_p(\lambda_R, \lambda_B)}{2}$$

where $\overline{J}_p(\lambda_R, \lambda_G)$ is the value clipped in $[0, 1]$.

The RAM highlights the significant wavelength in the reflectance ($R$).
Shading Gradient Map

Consider $K_p(\lambda_c) = \log(I_p(\lambda_c))$. If 2 neighboring pixels have the same reflectance, then the gradient is independent of reflectance and illumination.

$$\nabla K(\lambda_c) = K_{p_1}(\lambda_c) - K_{p_2}(\lambda_c) = \nabla \log \left( [\hat{L}.\hat{N}] \right)$$

In such points, the gradient is independent or reflectance and illumination. So, SG is defined as,

$$\nabla K(\lambda_c) = \begin{cases} 
\nabla \log \left( [\hat{L}.\hat{N}] \right) & \text{if } M_{RRG}^{(c)} < 0.1 \\
0 & \text{otherwise}
\end{cases} ; c \in \{R, G, B\}$$

where $M_{RRG}^{(c)}$ is a filter to find pixels with same reflectance as their neighbors.

$$M_{RRG}^{(R)} = \frac{\nabla J(\lambda_R, \lambda_G) + \nabla J(\lambda_R, \lambda_B)}{2}$$
RRG, RAM and SG images
Network Architecture

[Diagram of a network architecture with layers including Conv, Batch norm, LeakyReLU, Dropout, ReLU, and Sigmoid functions connected in a series and parallel fashion.]
Proposed loss function

\[ \mathcal{L} = \alpha_1 \mathcal{L}_{recon} + \alpha_2 \mathcal{L}_{ss} + \alpha_3 \mathcal{L}_{rrg} + \alpha_4 \mathcal{L}_{sg} + \alpha_5 \mathcal{L}_{ram} \]

1. Reconstruction loss
2. Shading smoothness loss
3. Reflectance Ratio Gradient (RRG) loss
4. Reflectance Approximation Map (RAM) loss
5. Shading Gradient (SG) loss
Reflectance Ratio Gradient Loss (1/5)

Ensures that the RRG of reflectance \((R)\) is similar to the RRG of the image \((i)\)

\[
\mathcal{L}_{rrg} = \| f_{RRG}(R_i) - f_{RRG}(i) \|_1
\]
Reflectance Approximation Map Loss (2/5)

Ensures that the “significant wavelength” in the image and the reflectance are the same.

\[ L_{ram} = \| (R_i - f_{RAM}(i)) \times f_{RAM}(i) \|_1 \]
Shading Gradient Loss (3/5)

Ensures that the shading gradient of the image is similar to the natural logarithmic gradient of the shading component.

\[
\mathcal{L}_{sg} = \left\| \nabla \log(S_i) - f'_{SG}(i) \right\|_1 
\]
Reconstruction Loss (4/5)

Ensures that the image (I) can be reconstructed from the Reflectance (R) and Shading (S).
(assumption: all reflectance maps are invariable of the lighting condition)

\[ \mathcal{L}_{recon} = \| R_i S_i - I_i \|_1 \]
Shading Smoothness Loss (5/5)

Ensures that the shading map \((S)\) is smooth where the RRG is smooth.

\[
\mathcal{L}_{ss} = \| \nabla S_i \exp(-10f_{RRG}(i)) \|_1
\]
Training, Validation and Testing

● Training and Validation:
  ○ Dataset: LOL dataset. (485 training, 15 validation)
  ○ Epochs: 100 epochs using Adam optimizer.
  ○ Learning Rate: 0.002 initial with a exponential decay factor of 0.01

● Testing
  ○ Dataset: MIT dataset (20 images)
### Visual comparison of output (1/3)

<table>
<thead>
<tr>
<th>Input</th>
<th>Letry et.al</th>
<th>CGIntrinsic</th>
<th>RetinexNet</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Input Image" /></td>
<td><img src="image2.png" alt="Letry et.al" /></td>
<td><img src="image3.png" alt="CGIntrinsic" /></td>
<td><img src="image4.png" alt="RetinexNet" /></td>
<td><img src="image5.png" alt="Ours" /></td>
</tr>
<tr>
<td><img src="image6.png" alt="Input Image" /></td>
<td><img src="image7.png" alt="Letry et.al" /></td>
<td><img src="image8.png" alt="CGIntrinsic" /></td>
<td><img src="image9.png" alt="RetinexNet" /></td>
<td><img src="image10.png" alt="Ours" /></td>
</tr>
<tr>
<td><img src="image11.png" alt="Input Image" /></td>
<td><img src="image12.png" alt="Letry et.al" /></td>
<td><img src="image13.png" alt="CGIntrinsic" /></td>
<td><img src="image14.png" alt="RetinexNet" /></td>
<td><img src="image15.png" alt="Ours" /></td>
</tr>
</tbody>
</table>
# Visual comparison of output (2/3)

<table>
<thead>
<tr>
<th>Input</th>
<th>Letry et. al</th>
<th>CGIntrinsic</th>
<th>RetinexNet</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Input" /></td>
<td><img src="image2" alt="Letry et. al" /></td>
<td><img src="image3" alt="CGIntrinsic" /></td>
<td><img src="image4" alt="RetinexNet" /></td>
<td><img src="image5" alt="Ours" /></td>
</tr>
</tbody>
</table>

## Reflectance

## Shading
## Visual comparison of output (3/3)

<table>
<thead>
<tr>
<th>Input</th>
<th>Letry et.al</th>
<th>CGIntrinsic</th>
<th>RetinexNet</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Input Image" /></td>
<td><img src="image2.png" alt="Letry et.al" /></td>
<td><img src="image3.png" alt="CGIntrinsic" /></td>
<td><img src="image4.png" alt="RetinexNet" /></td>
<td><img src="image5.png" alt="Ours" /></td>
</tr>
<tr>
<td>Reflectance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Numerical comparison - Validation Set

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE ↓</th>
<th>PSNR ↑</th>
<th>SSIM ↑</th>
<th>NIQE ↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letry et.al</td>
<td>21.87</td>
<td>35.28</td>
<td>0.96</td>
<td>7.55</td>
</tr>
<tr>
<td>CGIntrinsic</td>
<td>63.28</td>
<td>18.95</td>
<td>0.36</td>
<td>14.78</td>
</tr>
<tr>
<td>Retinex-net</td>
<td>6.88</td>
<td>34.64</td>
<td>0.90</td>
<td>7.63</td>
</tr>
<tr>
<td>Ours</td>
<td>2.00</td>
<td>43.12</td>
<td>0.95</td>
<td>7.63</td>
</tr>
</tbody>
</table>

↓ : Lower is better
# Numerical comparison - Test Set

<table>
<thead>
<tr>
<th>Method</th>
<th>Metric</th>
<th>MIT : 20 Images</th>
<th>MIT (R)</th>
<th>MIT (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>PSNR</td>
<td>SSIM</td>
<td>NIQE</td>
</tr>
<tr>
<td>Letry et.al</td>
<td>6.67</td>
<td>39.26</td>
<td>0.99</td>
<td>12.06</td>
</tr>
<tr>
<td>CGIntrinsic</td>
<td>40.95</td>
<td>17.36</td>
<td>0.11</td>
<td>17.47</td>
</tr>
<tr>
<td>Retinex-net</td>
<td>3.77</td>
<td>37.85</td>
<td>0.95</td>
<td>14.02</td>
</tr>
<tr>
<td>Ours</td>
<td>1.04</td>
<td>41.66</td>
<td>0.96</td>
<td>14.02</td>
</tr>
</tbody>
</table>
Summary

- Intrinsic Image Decomposition (IID) with simplified Phong’s model:

\[ I_p(\lambda_c) = r_p(\lambda_c) [\hat{L} \cdot \hat{N}_p] i_d(\lambda_c) \]

- A set of maps (RRM, RAM, SG) to extract meaningful information from images.

- Optical physics inspired loss function and CNN model

\[ \mathcal{L} = \alpha_1 \mathcal{L}_{recon} + \alpha_2 \mathcal{L}_{ss} + \alpha_3 \mathcal{L}_{rrg} + \alpha_4 \mathcal{L}_{sg} + \alpha_5 \mathcal{L}_{ram} \]

- Evaluating the model using numerical (RMSE, PSNR, SSIM, NIQE) and visual results.