HDR from Images and Events

Motivation:
• Image-based HDR methods suffer from limited dynamic range per exposure bracket
• Motion related artefacts (ghosting)

Event-based Cameras:
• Capture motion in the scene, no motion → no signal
• Asynchronous independent pixels
• Event $E = \{x, y, t, p\}$ if brightness change > threshold

Advantages:
• High temporal resolution (~1μs)
• High dynamic range (140 dB vs 60 dB)
• Ultra-low power (1mW vs 1W)
• No motion blur

Limitations:
• Information only in moving regions
• Grayscale
• Low-resolution
• Fine detail difficult to reconstruct (SoTA results are poor quality)

Idea: Combine the best of both worlds
• A multi-modal HDR model that leverages both bracketed exposures from a conventional frame-based camera and events from an event-based camera
• Event cameras provide high-frequency visual information, images provide rich RGB colour and fine details at lower frequency

Our Method
$$\hat{I} = g \left( A(L), P^E(E), P^E(D, E); \theta \right)$$

Event-to-Image Distillation:
• Leverage complementary information between events and images
• Translate sub-sampled event features into image features
• Self-supervising loss on the event features and corresponding image features:
$$L_D = \sum_{s=1}^{S} \sum_{i=1}^{3} (f^E_{s,i} - sg(f^I_{s,i}))^2$$

Ablations

<table>
<thead>
<tr>
<th>Method</th>
<th>Imbalanced Params</th>
<th>Balanced Params</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Params(M)</td>
<td>PSNR-L</td>
</tr>
<tr>
<td>Images-only</td>
<td>2.81</td>
<td>39.18</td>
</tr>
<tr>
<td>+ Event alignment</td>
<td>4.38</td>
<td>40.97</td>
</tr>
<tr>
<td>+ Event sub-sampling</td>
<td>4.67</td>
<td>41.32</td>
</tr>
<tr>
<td>+ Event-to-image distill</td>
<td>6.13</td>
<td>41.81</td>
</tr>
</tbody>
</table>

Results

Conclusions
• Multi-model HDR architecture combining events and low dynamic range images
• Leverages high-frequency events to better align and reconstruct details
• Significant improvements over other SoTA event- or image-only methods