VR Sickness Assessment with Perception Prior and Hybrid Temporal Features

Po-Chen Kuo, Li-Chung Chuang, Dong-Yi Lin and Ming-Sui Lee

Department of Computer Science and Information Engineering, Graduate Institute of Networking and Multimedia, National Taiwan University, Taiwan

Speaker: Li-Chung Chuang
VR sickness

Dizziness
Accelerated heartbeat
Weakness
Unsteadiness
nausea
Related work


Goal

Contributions:
1. A newly created dataset of VR content with per-minute Discomfort Score
2. A novel hybrid temporal feature with perception prior
3. Performance is on par with the state-of-the-art method
• 20 videos, each is 5 minutes
• Every video is watched by 5 testers and report their Discomfort Score every minute
• The average discomfort scores of 5 testers is the DS of that video
Perception prior features

<table>
<thead>
<tr>
<th>width:height</th>
<th>PLCC</th>
<th>SROCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1</td>
<td>0.82</td>
<td>0.81</td>
</tr>
<tr>
<td>1:2</td>
<td>0.87</td>
<td>0.83</td>
</tr>
<tr>
<td>1:3</td>
<td>0.90</td>
<td>0.76</td>
</tr>
<tr>
<td>2:1</td>
<td>0.73</td>
<td>0.75</td>
</tr>
<tr>
<td>3:1</td>
<td>0.67</td>
<td>0.58</td>
</tr>
</tbody>
</table>

- Optical flow based features
- Image area outside 110° FOV are culled out
- Image area inside the FOV are weighted by a Gaussian kernel
Hybrid temporal features

- **Horizontal motion strength**
  \[ f_x = \frac{1}{t} \sum_{k=1}^{t} \sum_{p(i,j) \in V} x^k(i,j) \odot g(i,j); \]

- **Vertical motion strength**
  \[ f_y = \frac{1}{t} \sum_{k=1}^{t} \sum_{p(i,j) \in V} y^k(i,j) \odot g(i,j); \]

- **Motion anisotropy**
  \[ f_h = \text{median}(h(i,j)), p(i,j) \in V \]

\[ h(i,j) = - \sum_{l=1}^{16} P\left(z^l(i,j)\right) \log_b P\left(z^l(i,j)\right) \]
Training model

1 minute
\{I_t, I_{t+1}, \cdots, I_{t+N}\} 

2 minute
\{I_{t+N+1}, I_{t+N+2}, \cdots, I_{t+2N}\} 

\ldots

5 minute
\{I_{t+4N+1}, I_{t+4N+2}, \cdots, I_{t+5N}\} 

\begin{align*}
\text{calculate optical flow} & \quad \Rightarrow \\
\text{feature extract} & \quad \Rightarrow \\
\text{Random Forest} & \quad \Rightarrow \text{Predict Score } 1 \\
\end{align*}

\begin{align*}
f_1 &= (f^1_x, f^1_y, f^1_h) \\
\text{calculate optical flow} & \quad \Rightarrow \\
\text{feature extract} & \quad \Rightarrow \\
\text{Random Forest} & \quad \Rightarrow \text{Predict Score } 2 \\
\end{align*}

\begin{align*}
f_2 &= (f^2_x, f^2_y, f^2_h) \\
\text{calculate optical flow} & \quad \Rightarrow \\
\text{feature extract} & \quad \Rightarrow \\
\text{Random Forest} & \quad \Rightarrow \text{Predict Score } 5 \\
\end{align*}
Experimental results on KAIST dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Perception Prior</th>
<th>Hybrid temporal feature</th>
<th>PLCC</th>
<th>SROCC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Horizontal Motion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vertical Motion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Motion Anisotropy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee’s [1]</td>
<td>N/A</td>
<td>N/A</td>
<td>0.75</td>
<td>0.80</td>
</tr>
<tr>
<td>VRSA [9]</td>
<td>N/A</td>
<td>N/A</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>Proposed</td>
<td></td>
<td>✓</td>
<td>0.57</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓</td>
<td>0.69</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>0.79</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>0.87</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>0.90</td>
<td>0.90</td>
</tr>
</tbody>
</table>

- The proposed method outperforms VRSA and Lee’s method on the KAIST dataset.
- The ablation test demonstrated that the proposed perception prior and hybrid temporal features are effective.
## Execution time comparison on KAIST dataset

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.00</td>
<td>3.81</td>
<td>249.79</td>
<td>4.57</td>
</tr>
<tr>
<td>2</td>
<td>29.97</td>
<td>3.56</td>
<td>249.24</td>
<td>3.98</td>
</tr>
<tr>
<td>3</td>
<td>29.97</td>
<td>3.65</td>
<td>249.62</td>
<td>4.10</td>
</tr>
<tr>
<td>4</td>
<td>30.00</td>
<td>3.58</td>
<td>249.46</td>
<td>4.13</td>
</tr>
<tr>
<td>5</td>
<td>29.97</td>
<td>3.57</td>
<td>249.95</td>
<td>4.07</td>
</tr>
<tr>
<td>6</td>
<td>29.97</td>
<td>3.70</td>
<td>250.03</td>
<td>4.34</td>
</tr>
<tr>
<td>7</td>
<td>25.00</td>
<td>3.63</td>
<td>208.13</td>
<td>4.07</td>
</tr>
<tr>
<td>8</td>
<td>59.94</td>
<td>3.72</td>
<td>500.23</td>
<td>4.25</td>
</tr>
<tr>
<td>9</td>
<td>29.97</td>
<td>3.67</td>
<td>248.98</td>
<td>4.23</td>
</tr>
</tbody>
</table>
Conclusion

- We collected a dataset of twenty 360 degree videos with per-minute Discomfort Score
- A novel hybrid temporal feature with perception prior was proposed
- The experiment results show that the proposed method is comparable to the state-of-the-art methods