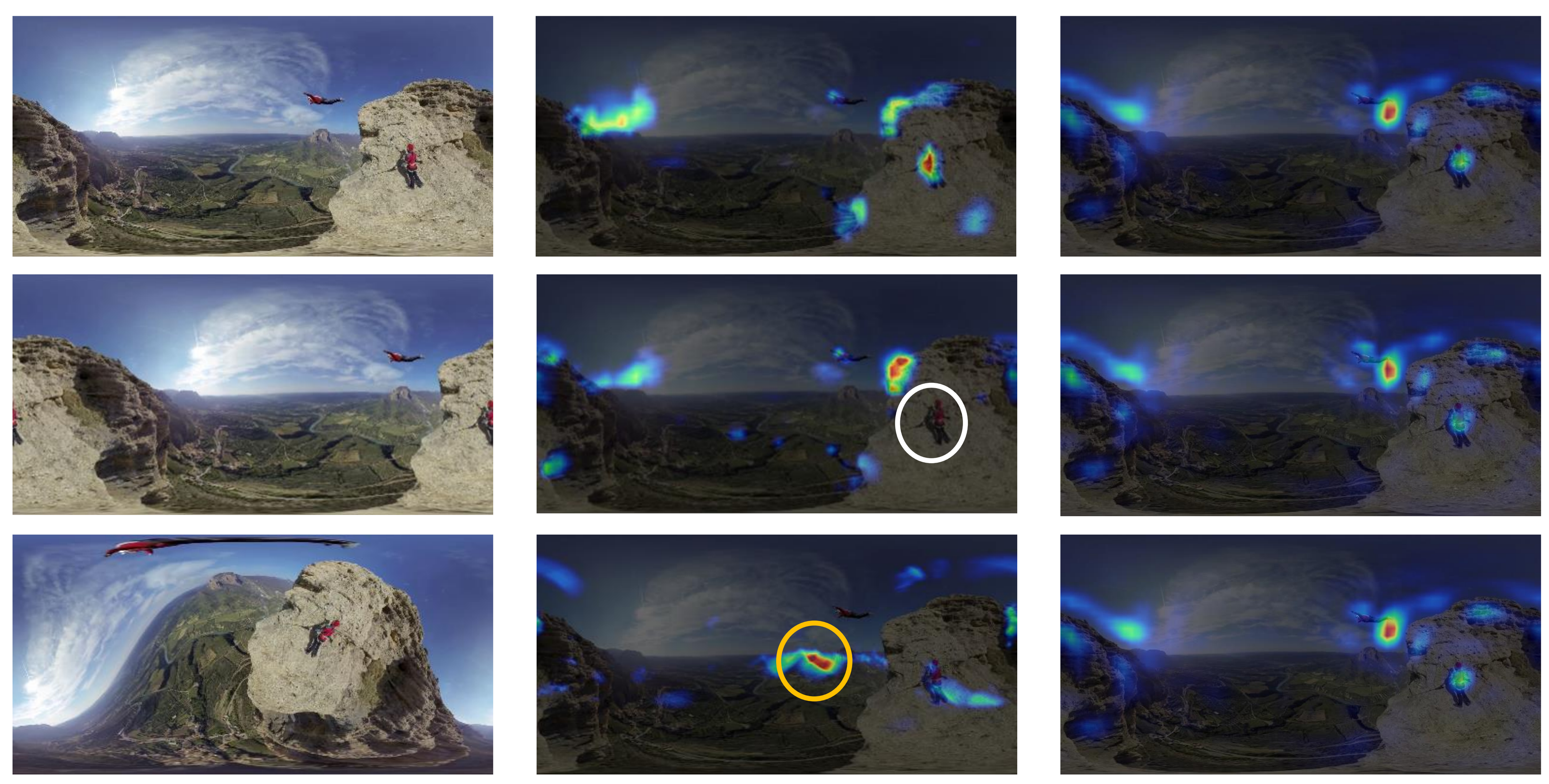
Spherical Harmonics

for Saliency Computation and Navigation in 360° Videos

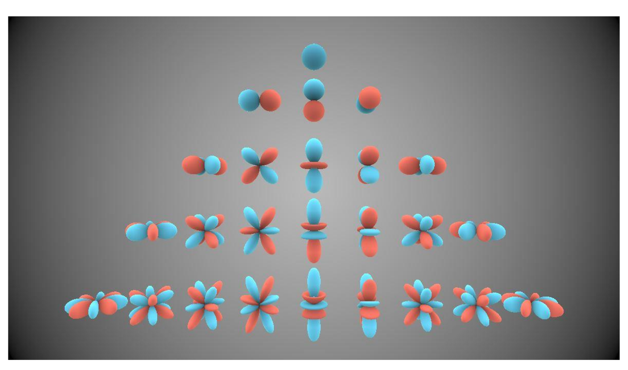
Ruofei Du and Amitabh Varshney

INTRODUCTION

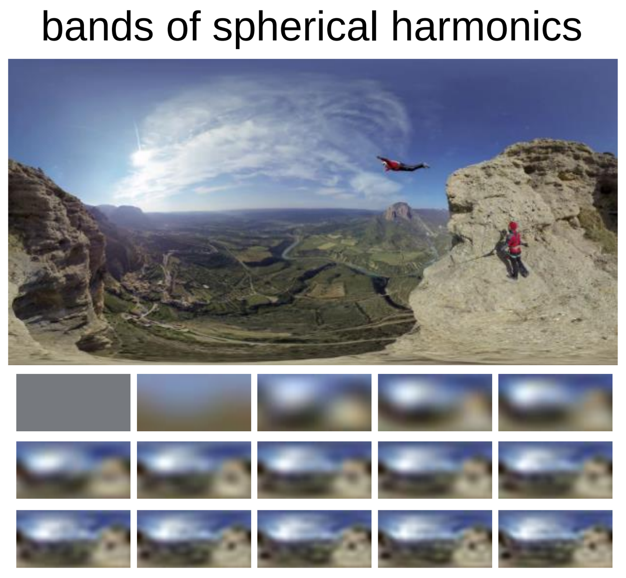
Omnidirectional videos, or 360° videos, have exploded in popularity due to the recent advances in virtual reality head-mounted displays (HMDs) and cameras. Despite the 360° field of regard (FoR), almost **90%** of the pixels are outside a typical HMD's field of view (FoV). Hence, understanding where users are more likely to look at plays a vital role in efficiently processing and rendering 360° videos. While conventional saliency models have shown robust performance over rectilinear images, they are not formulated to handle equator biases, horizontal clipping, and spherical rotations in 360° videos.



We present a novel GPU-driven pipeline for saliency computation and navigation in 360° videos, based upon spherical harmonics (SH). By analyzing the spherical harmonics spectrum of the 360° video, we extract the spectral residual by accumulating the SH coefficients between a low band and a high band. Our model outperforms the classic Itti et al.'s model in timings by over 5 to 13 times and runs at over 60 FPS for 4K videos. Furthermore, we use the interactive computation of saliency to devise a saliency-guided system for virtual cinematography in 360° videos. We formulate a spatiotemporal model to ensure large saliency coverage while reducing the camera movement jitter. We envision that our pipeline will be used in processing, navigating, and rendering 360° videos in real time.



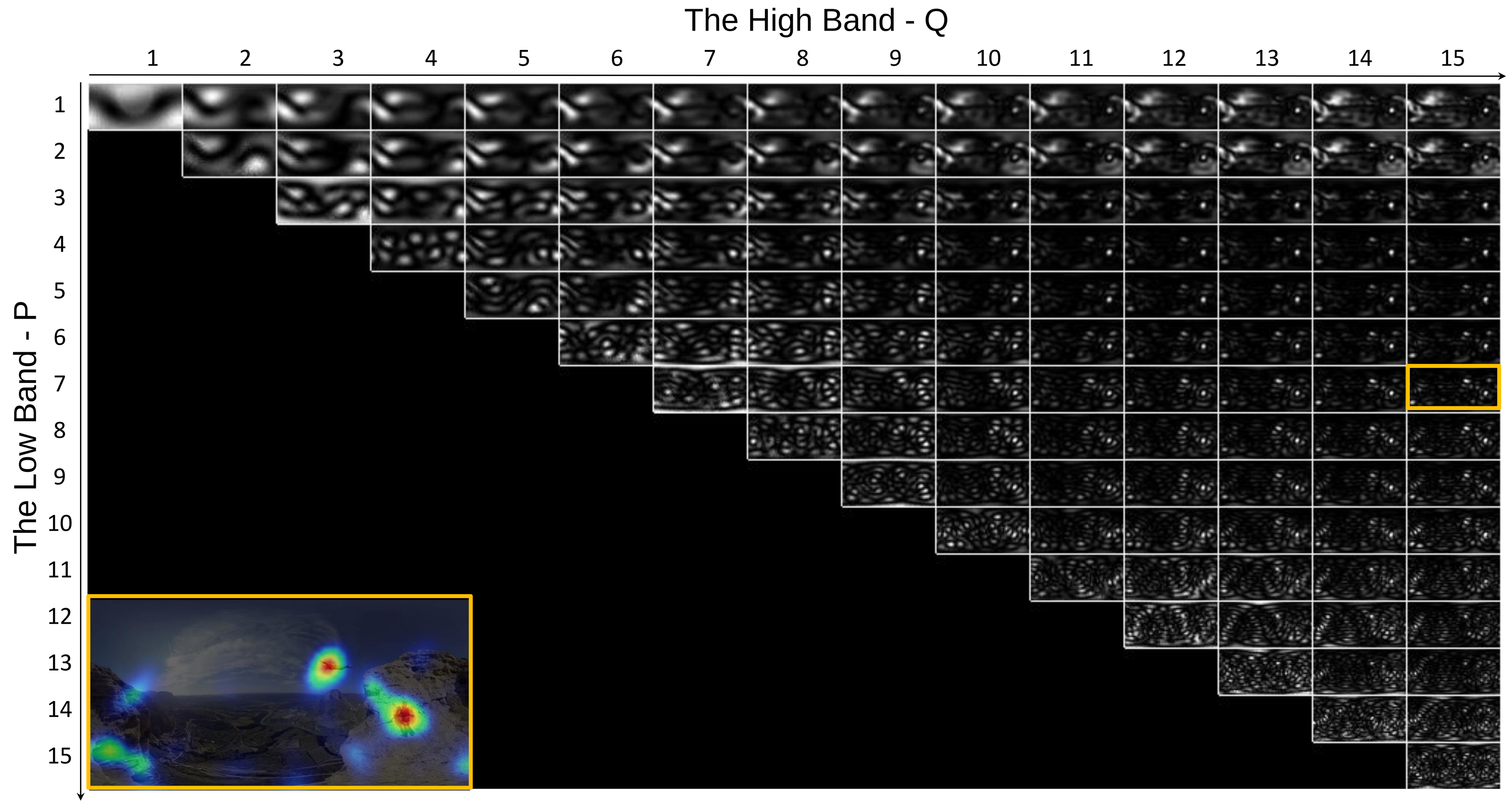
Visualization of the first five bands of spherical harmonics



Reconstructed images from the first 15 bands of Spherical Harmonics coefficients

SALIENCY COMPUTATION

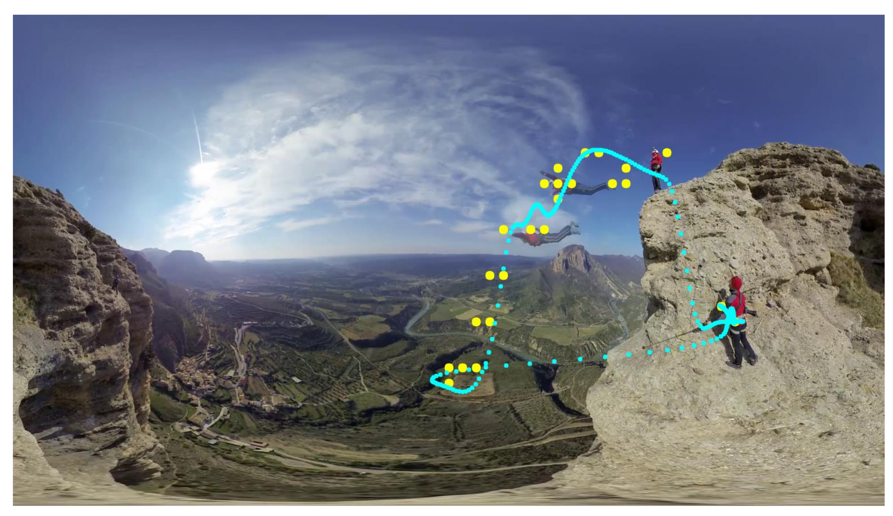
In the space of $\mathbb{SO}(2)$, we define the spherical spectral residual as the subtraction between higher bands (up to Q) of spherical harmonics coefficients and the lower bands (up to P) of spherical harmonics coefficients.



AUTOMATIC NAVIGATION

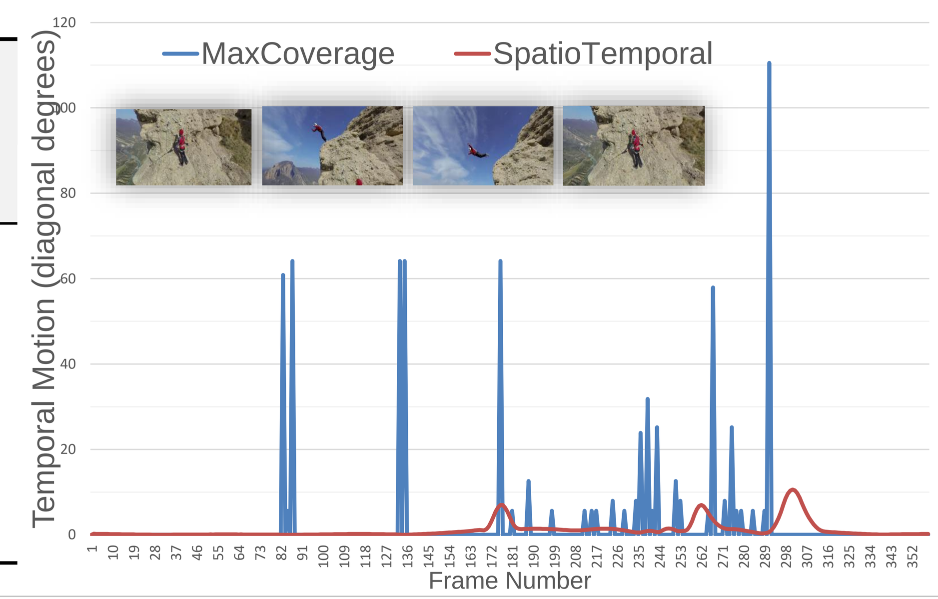
We propose a spatiotemporal optimization model of the virtual camera's discrete control points and further employ a spline interpolation amongst the control points to achieve smooth camera navigation.

$$E(L) = \lambda_{saliency} \cdot E_{saliency}(L) + \lambda_{temporal} \cdot E_{temporal}(L)$$



EXPERIMENTAL RESULTS

Resolution	The Average Timings Per Frame			
	Itti (CPU)	SCS (CPU)	SSR (CPU)	SSR (GPU)
1920x1080	104.46 ms	47.93 ms	21.34 ms	10.81 ms
4096x2048	314.94 ms	85.29 ms	48.18 ms	13.20 ms
7680x3840	934.26 ms	152.62 ms	69.53 ms	26.58 ms



THE AUGMENTARIUM
VIRTUAL AND AUGMENTED REALITY LABORATORY
AT THE UNIVERSITY OF MARYLAND
GVIL
COMPUTER SCIENCE
UNIVERSITY OF MARYLAND