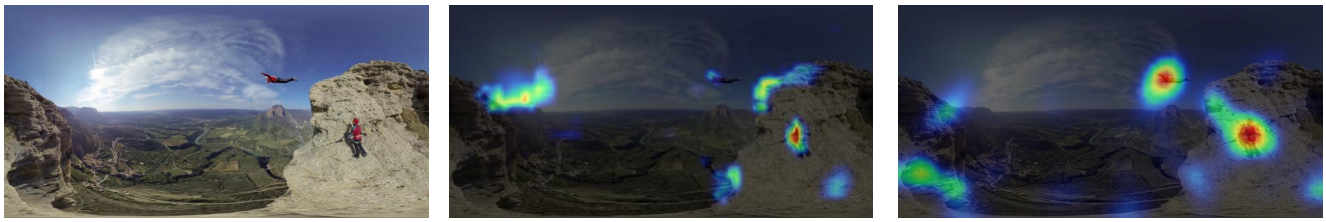
# A Pilot Study of Spherical Harmonics for Saliency Computation and Navigation in 360° Videos

Ruofei Du and Amitabh Varshney
Augmentarium, Department of Computer Science and the Institute for Advanced Computer Studies (UMIACS)
University of Maryland, College Park
{ruofei,varshney}@umiacs.umd.edu



**(A)** The input 360° video frame **(B)** Saliency map by Itti *et al.*'s model **(C)** Saliency map by our SSR model

**Figure 1: This paper presents an efficient GPU-driven pipeline of computing saliency maps of 360° videos using spherical harmonics (SH). (A) shows an input frame from a 360° video. (B) shows the saliency maps computed by the Itti model in** 104.46 **ms on the CPU. (C) show the saliency maps computed by our spherical spectral residual (SSR) model in** 21.34 **ms on the CPU and** 10.81 **ms on the GPU. In contrast to the Itti model, our model is formulated in the** $\mathbb{SO}(2)$ **space and handles challenging cases such as horizontal clipping, spherical rotations, and equator biases in 360° videos.**

## ABSTRACT

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. In this paper, we present a novel GPU-driven pipeline for saliency computation and navigation in 360° videos, based upon spherical harmonics (SH). We introduce the Spherical Spectral Residual (SSR) model. In this approach, we define the saliency maps as the accumulation of the SH coefficients between a low band and a high band. Our model outperforms the Itti *et al.*'s model in timings by over 5 to 13 times and runs at over 60 FPS for 4*K* videos. We envision that our pipeline will be used in processing, navigating, and rendering 360° videos in real time.

## CCS CONCEPTS

• **Computing methodologies** → **Interest point and salient region detections**; *Perception*; Virtual reality;

## KEYWORDS

spherical harmonics, virtual reality, visual saliency, 360° videos, omnidirectional videos, perception, Itti model, spectral residual, GPGPU, CUDA

**ACM Reference Format:**
Ruofei Du and Amitabh Varshney. 2018. A Pilot Study of Spherical Harmonics for Saliency Computation and Navigation in 360° Videos. In *Proceedings of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (ACM I3D'18),* Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.475/123_4


*ACM I3D'18, May 2018, Montreal, Quebec, Canada*

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

## 1 INTRODUCTION

With recent advances in consumer-level virtual reality (VR) head-mounted displays (HMD) and panoramic cameras, more and more scenes are captured as omnidirectional videos. These 360° videos are becoming a crucial medium for news reports, live concerts, remote education, and social media. Despite the rich omnidirectional visual information in the panoramas, most of the content is out of the field of view (FoV) of the head-mounted displays, as well as human eyes. Therefore, predicting where humans will look at, *i.e.*, saliency detection, has great potential over a wide range of applications, such as efficiently compressing and streaming high-resolution panoramic videos under poor network conditions and salient object detection in panoramic images or videos.

In the pilot study, we present a GPU-driven pipeline to formulate saliency *natively and directly* in the special orthogonal group $\mathbb{SO}(2)$ space using the spherical harmonics coefficients, without converting the image to $\mathbb{R}_2$. Furthermore, we reduce the computational cost and formulate the spherical saliency using the spectral residual model with spherical harmonic.

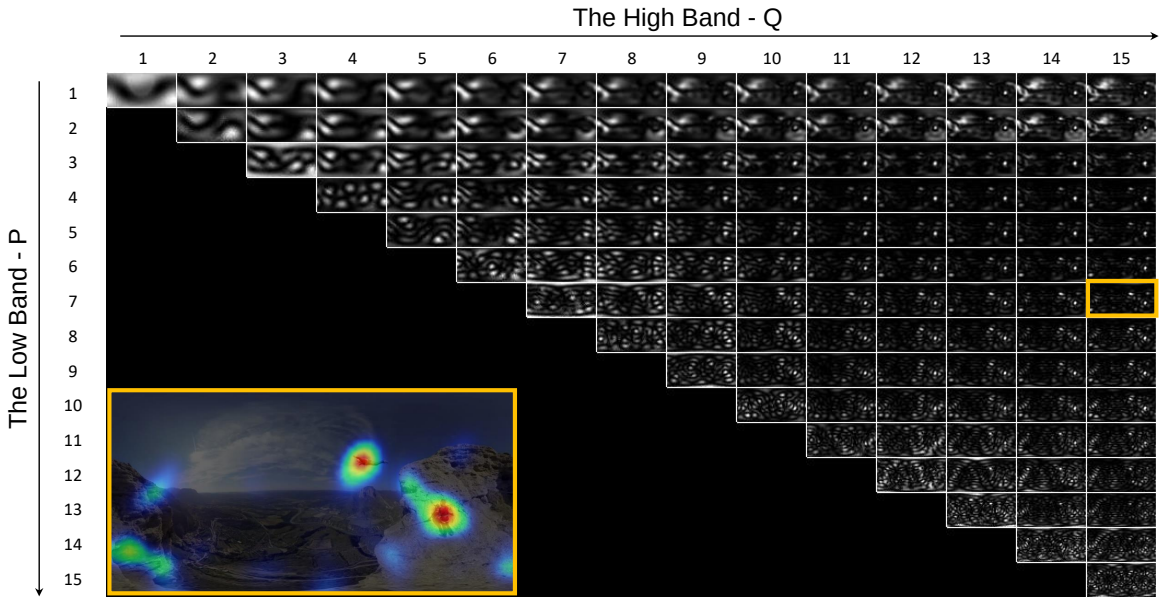



**Figure 2: The spectral residual maps between different bands of spherical harmonics. The number along the horizontal axis indicates the high band $Q$, while the vertical axis indicates the low band $P$. Note that the saliency maps within or close to the orange bounding box successfully detect the two people in the frame.**

## 2 SPHERICAL SPECTRAL RESIDUAL MODEL

We present a new approach to compute saliency for spherical 360° videos using the idea of spectral residual with spherical harmonics. The SH coefficients consist of $L^2$ values for $L$ bands. For the $m^{\text{th}}$ element of a specific band $l$, we evaluate the SH coefficients of the feature map $f$ as:

$$c_l^m(\theta,\phi) = \int_{(\theta,\phi)\in S} f(\theta,\phi)\cdot Y_l^m(\theta,\phi)\ \mathrm{d}\theta\ \mathrm{d}\phi \tag{1}$$

, where $0 \le l \le L$ is the band index, $m$ is the order of the band, and $-l \le m \le l$. $f(\theta,\phi)$ is the value of the feature map at $\theta,\phi$. $P_l^m$ are the associated Legendre polynomials. The lower bands of spherical harmonics contain low-frequency background information such as the sky and mountains, which are not essential for saliency. Therefore, the SSR model is aimed at discarding the low-frequency information and evaluating the saliency across multiple scales in the spectral domain.

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

$$\begin{aligned}\Re(\phi,\theta) &= \sum_{l=0}^{Q}\sum_{m=-l}^{l} c_l^m\cdot Y_l^m(\phi,\theta) - \sum_{l=0}^{P}\sum_{m=-l}^{l} c_l^m\cdot Y_l^m(\phi,\theta)\\ &= \sum_{l=P+1}^{Q}\sum_{m=-l}^{l} c_l^m\cdot Y_l^m(\phi,\theta)\end{aligned} \tag{2}$$

in which $Y_l^m(\phi,\theta)$ are pre-computed associated Legendre polynomials in the preprocessing stage. The SSR represents the salient part of the scene in the spectral domain and serves as a compressed representation using spherical harmonics.

We square the spectral residual to reduce the estimation errors. For better visual effects, we smooth the spherical saliency maps using the following post-processing method:

$$\mathcal{S}(\phi,\theta) = \mathfrak{G}(\sigma) \circledast [\Re(\phi,\theta)]^2 \tag{3}$$

, where $\mathfrak{G}(\sigma)$ is a Gaussian filter with standard deviation $\sigma$ ($\sigma = 5$ for the results presented in this paper).

We show the SSR results of the intensity channel with all different pairs of the lower band $P$ and the higher band $Q$ in Figure 2. As $P$ increases, the low-frequency information such as the sky and mountains are filtered out. The spectral residual results within and close to the orange bounding box reveal the salient objects, such as the two people, clearly.

We apply this model to intensity, color, and temporal motion channels separately. After a non-linear normalization, we combine all conspicuity maps into the final saliency map. Empirically, we choose $Q = 15, P = 7$. The final composed result is shown at the bottom left corner in Figure 2, as well as in the accompanying video.

In a pilot study, our method remains consistent for challenging cases like horizontal clipping, spherical rotations, and equator biases, and is 5 to 13 times faster than the classic Itti *et al.*'s model.

We envision our techniques will be widely be used for live streaming of events, video surveillance of public areas, as well as templates for directing the camera path for immersive storytelling. Future research may explore how to naturally place 3D objects with spherical harmonics irradiance in 360° videos, how to employ spherical harmonics for foveated rendering in 360° videos, and the potential of compressing and streaming 360° videos with spherical harmonics.