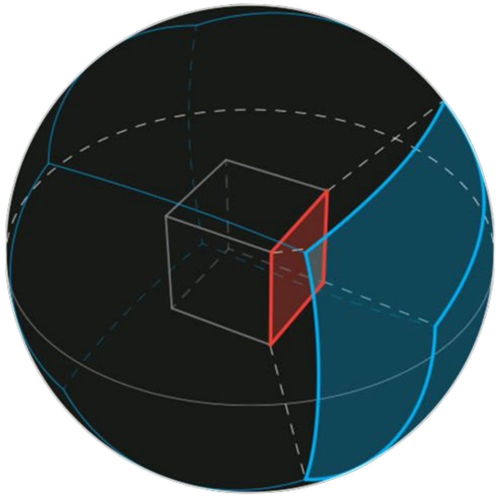
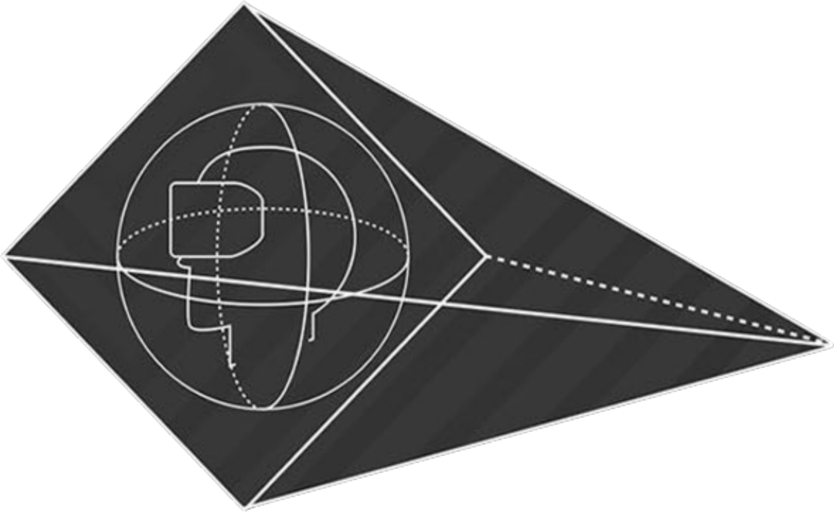
A Framework to Evaluate Omnidirectional Video Coding Schemes

M. Yu, H. Lakshman and B. Girod, "A Framework to Evaluate Omnidirectional Video Coding Schemes," 2015 IEEE International Symposium on Mixed and Augmented Reality, Fukuoka, 2015, pp. 31-36.

Introduction

- Omnidirectional videos are mapped onto one or more planes before encoding to interface with modern video coding standards.
- Different mappings and different test criteria have been employed in many literatures to report coding efficiency.





Contributions

- They propose a method to compare the original and the coded omnidirectional videos by generating viewports corresponding to head motion data to compute PSNR between viewports

- They also propose sphere based PSNR (S-PSNR) computations to approximate the average quality over all possible viewing directions

Review of Panoramic Projections

- **Equirectangular projection:**
  - constant spacing:
    - latitude $\phi \in [-\pi/2, \pi/2]$
    - longitude $\theta \in [-\pi, \pi]$
  - Address the vertical and horizontal positions in a panorama using ϕ and θ
  - Due to the constant spacing of latitude, this projection has a constant vertical sampling density on the sphere
  - Horizontally, each latitude ϕ (whose circumference is given by $\cos\phi$) is stretched to a unit length to fit in a rectangle. Therefore, the horizontal sampling density at latitude ϕ is given by $1/\cos\phi$, which tends to infinity near the poles.

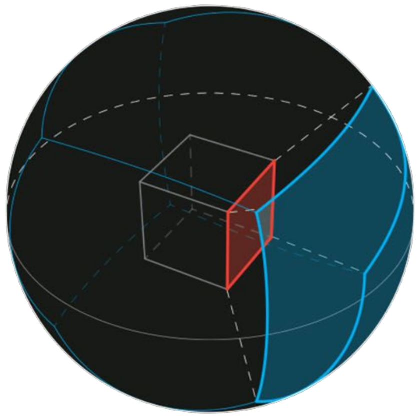
Review of Panoramic Projections

- **Lambert Cylindrical Equal-area:**
  - Compensate for the increasing horizontal sampling density as we go near the poles by decreasing the corresponding vertical sampling density.
  - The vertical sampling density is set to $\cos\phi$ so that the combined sampling density is constant throughout the sphere

Review of Panoramic Projections

- **Dyadic:**
  - Directly decrease the horizontal oversampling of equirectangular projection.
  - Halve the horizontal resolution of the panorama for $|\phi| \geq \frac{\pi}{3}$.

Review of Panoramic Projections



- **Cubic projection:**
  - Place the sphere of unit diameter at the center of a cube with unit length sides
  - Each face of the cube is generated by rectilinear projection with a 90-degree field of view in horizontal and vertical directions
  - The sampling density is lowest at the center of the cube faces and highest where the cube faces meet

Viewport-Based Quality Evaluation

- If we uniformly span the spherical coordinates in the visible region of the sphere and pass rays from O to the points on the sphere, they will intersect the viewport plane ABDC with non-uniform spacing between the pixels

- In order to compute a uniform grid of pixels in the viewports, they start with the desired locations in the viewport and **reverse** the mapping to compute corresponding location on the sphere

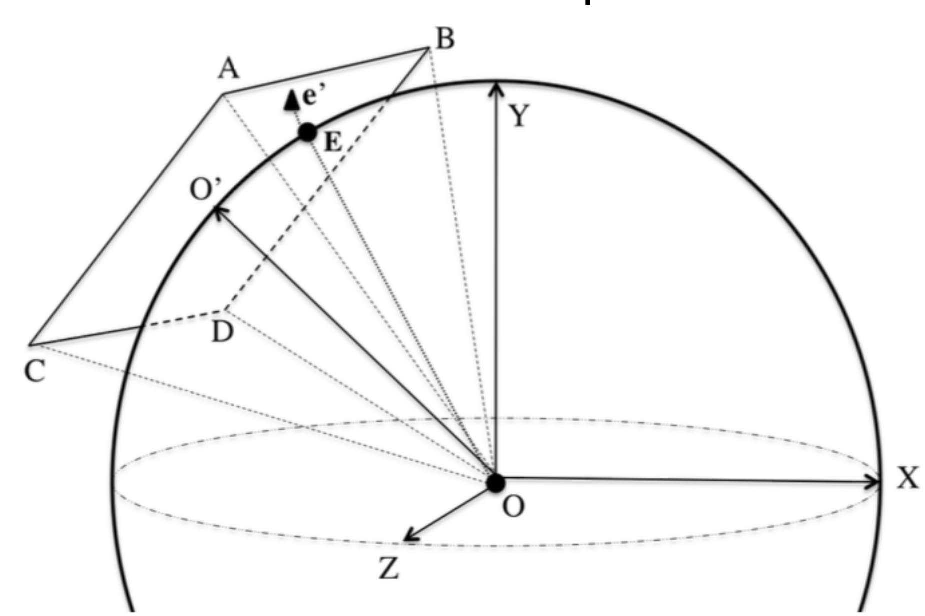


Figure 3: Example of a viewport.

Viewport-Based Quality Evaluation

$$w \cdot \mathbf{e}' = \mathbf{K} \cdot \mathbf{R}^T \cdot \mathbf{E} \qquad \mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{E} = [x, y, z]^T$$

- f_x and f_y are the focal length expressed in pixels
- c_x and c_y are the texture coordinates of principal point O' in the viewport
- **R** represents the rotation of the user's head relative to the canonical position which is that the user is looking down the negative z-axis
- **E** denotes a point on the sphere in the currently visible region

$$\mathbf{E} = \mathbf{R} \cdot \frac{\mathbf{K}^{-1}\mathbf{e}'}{||\mathbf{K}^{-1}\mathbf{e}'||_2}$$

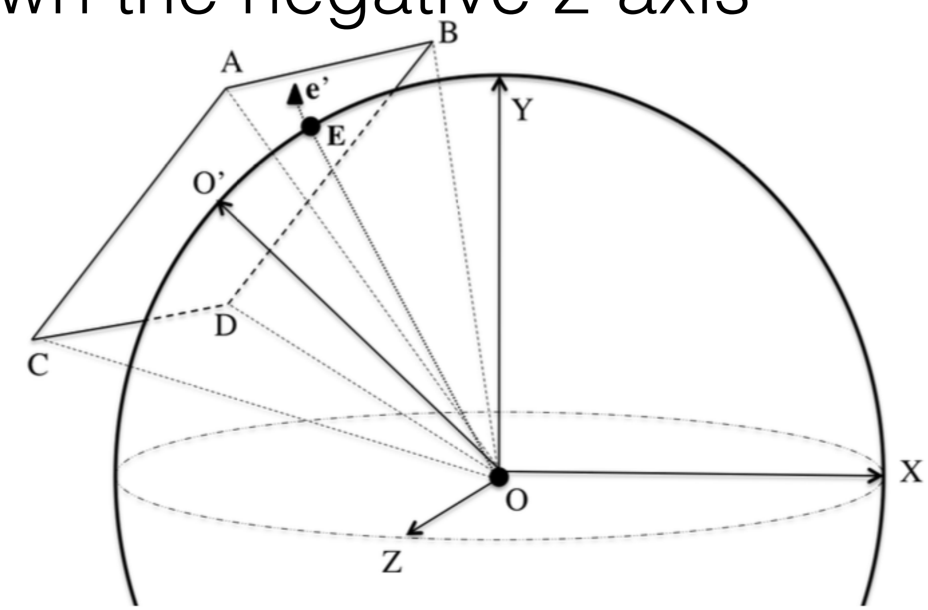
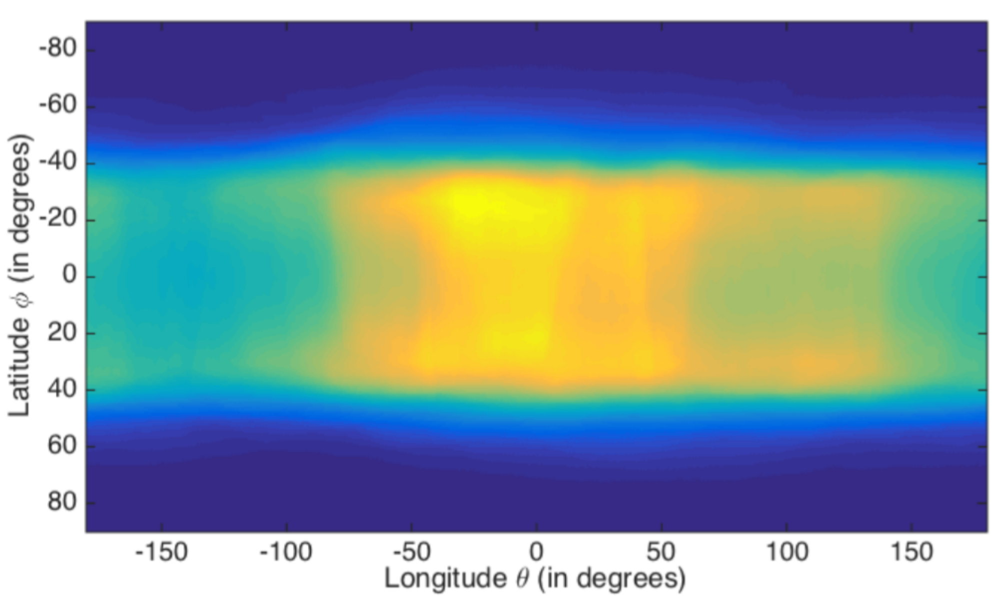


Figure 3: Example of a viewport.

Spherical Domain Comparison

- Based on the previous evaluation we can determine the corresponding set of points on the sphere
- They develop spherical PSNR (S-PSNR) to summarize the average quality over all possible viewports
- Next, they also observe that not all viewing directions are equally likely. They use relative frequencies to weight the coding errors



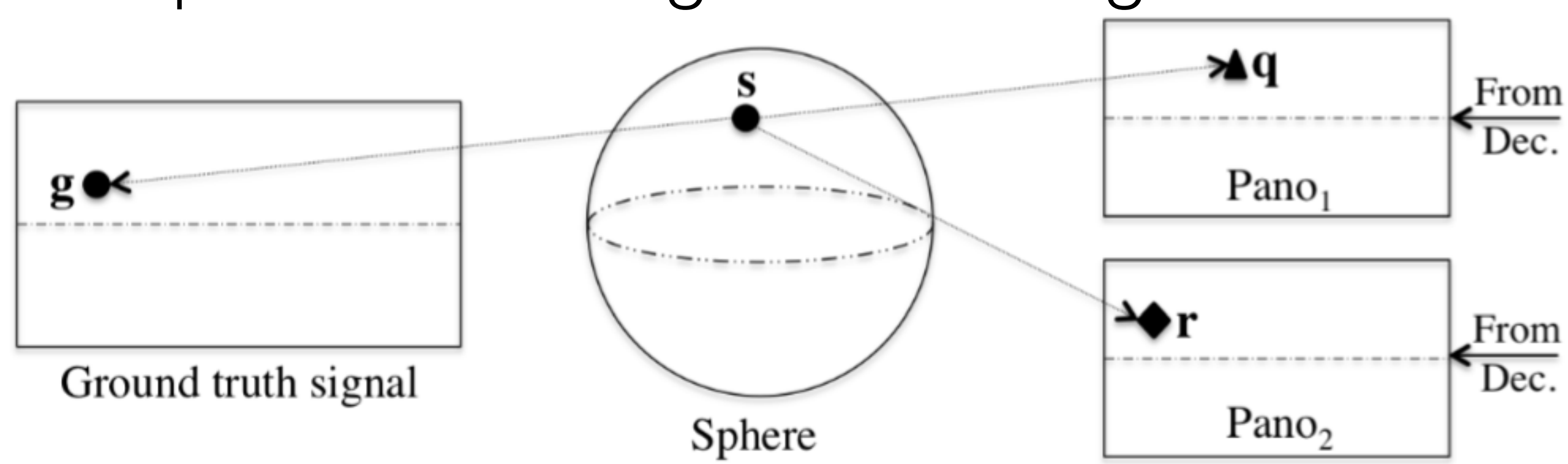


Figure 4: Comparison of the ground truth signal with coded panoramas on a set of uniformly sampled points on a sphere.

Experimental Results

- They use the previous methods to study how the quality varies with respect to the bitrate using an H.264/AVC codec

- In the first part they evaluate various mapping schemes and their impact on the coding efficiency

- Then, they consider the case of testing a coding system without the explicit knowledge of head motion trajectories

Experimental Results
- Mapping Comparisons

- The average bitrate savings of the Equal-area projection over the Equirectangular projection is approximately 8.3%

- The Dyadic projection also shows similar improvement over the Equirectangular projection

- The average performance of **cubic** projection is **lower** than the Equal-area and Dyadic projection

Sequence	Equal-area	Cube	Dyadic
BMX	9.4%	11.4%	3.3%
Cannes	-0.2%	7.0%	-0.8%
China1	-7.3%	-4.0%	-6.1%
China2	-8.3%	7.7%	-7.1%
Kauai1	-9.4%	-10.4%	-9.0%
Kauai2	-20.1%	-16.7%	-16.4%
Kauai3	-11.3%	-8.1%	-7.9%
London	5.6%	10.7%	2.1%
Monument	-36.4%	-29.7%	-27.7%
Waterfall	-5.4%	1.3%	-3.3%
Avg	-8.33%	-3.09%	-7.29%

Table 1: BD-rate comparison of various projections relative to the Equirectangular projection using the viewport evaluation method described in Sec. 4.

Experimental Results
- Mapping Comparisons

- The viewport PSNR at a given bitrate is the lowest when using the Equirectangular projection, while it is the highest when using the Equal-area projection.

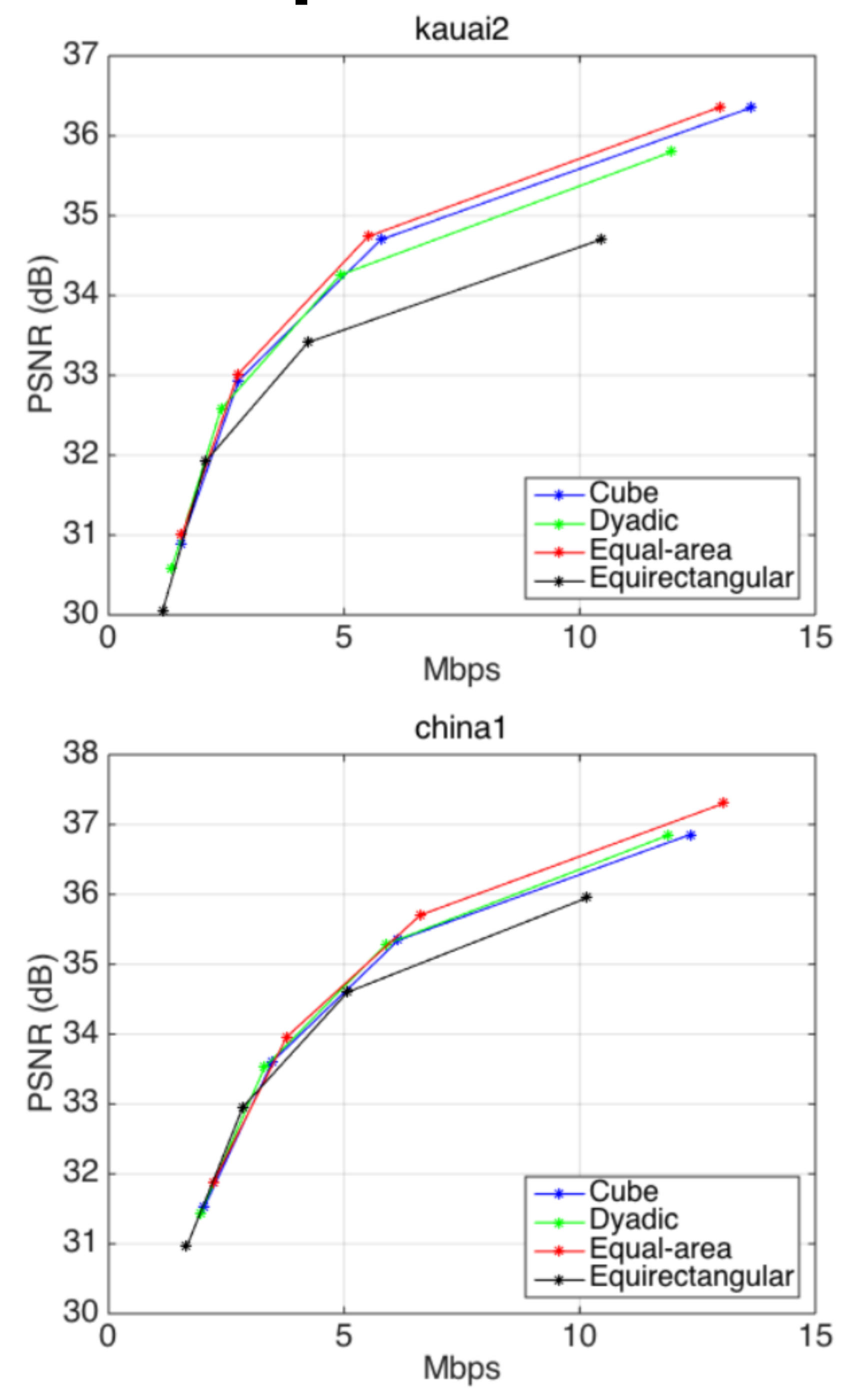


Figure 6: RD curves of two sequences for different panoramic projections using the viewport quality evaluation method.

Experimental Results
- Spherical PSNR vs. Viewport PSNR

- WeightSph: S-PSNR with sphere points weighted by point access frequency.
- LatSph: S-PSNR with sphere points weighted by the corresponding latitude access frequency.
- Sph: S-PSNR where all points are weighted equally.
- Quad: PSNR calculated by mapping both the ground truth and the coded videos to the same 6Kx3K Equirectangular projection.
- The WeightSph and LatSph methods differ from the reference by less than 7% on average without explicit head motion data

Projection	WeightSph	LatSph	Sph	Quad
Equirectangular	6.85%	7.18%	16.46%	23.36%
Equal-area	5.42%	6.03%	13.10%	26.28%
Cube	6.48%	6.66%	13.55%	19.81%
Dyadic	6.08%	6.31%	13.90%	20.06%
Avg	6.21%	6.55%	14.25%	21.38%

Table 2: BD-rate comparison of various metrics described in Sec. 6.2 relative to the viewport evaluation method when using different projections.

Experimental Results
- Spherical PSNR vs. Viewport PSNR

- WeightSph and LatSph methods are able closely approximate the viewport method with only general head motion statistics

- The Sph and Quad methods yield significantly different approximations

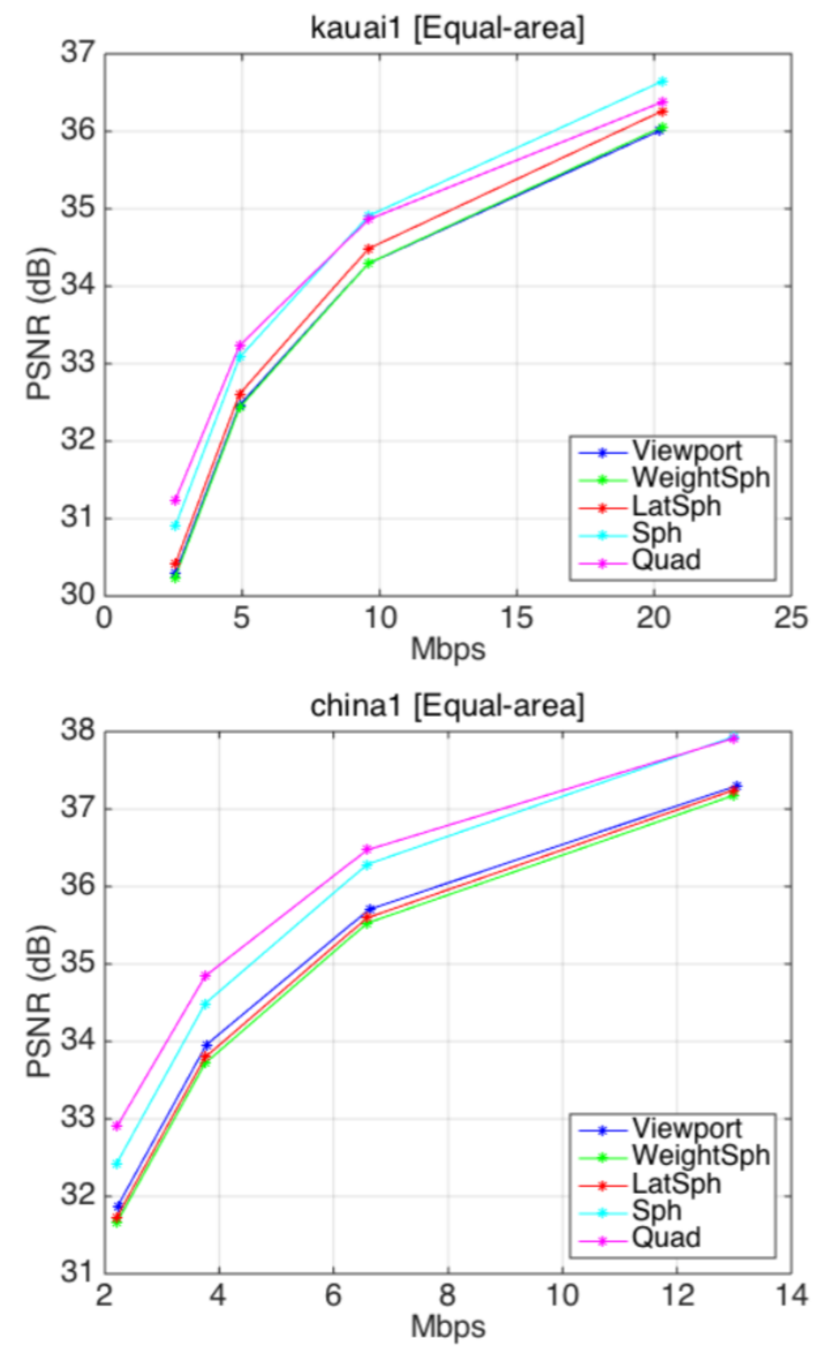


Figure 7: RD curves of two sequences coded using the Equal-area projection where the distortion is measured using various viewport quality approximation methods.

Conclusions

- They propose a framework which allows us to compare various sphere-to-plane mappings without bias toward any specific mapping or resolution.
- In the experimental results, it is possible to approximate the average viewport quality by exploiting general head motion statistics.