Paper ID: 1506

ICIP 2021

A Foveated Video Quality Assessment Model Using Space-Variant Natural Scene Statistics

Yize Jin[1], Todd Goodall[2], Anjul Patney[2], Richard Webb[2], Alan Bovik[1]
[1]The University of Texas at Austin, [2]Facebook Reality Labs

The University of Texas at Austin
LIVE
Laboratory for Image & Video Engineering

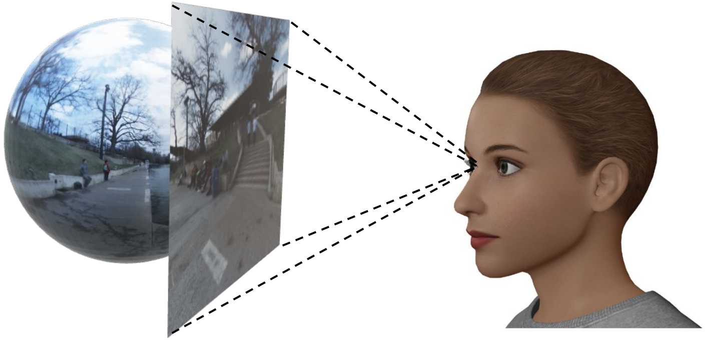
FACEBOOK REALITY LABS

Contents

- Introduction
  - Background
  - Challenges
  - Foveated Video Compression
- Proposed SVBRISQUE Model
  - Limitations of Traditional Models
  - Space-Variant Natural Scene Statistics
  - Summary of the Proposed Method
- Evaluation & Performance
  - Evaluation Framework
  - Results & Conclusion

Background

- In VR, immersive videos are usually rendered on a 3D geometric shape
- Only a certain field of view (FOV) is captured by virtual cameras and displayed in the head mounted displays.




Oculus Quest 2, ~20 ppd


Valve Index, ~15 ppd


Vive Pro, ~15 ppd


Oculus rift s, ~25 ppd


Pimax Vision 8K, ~40 ppd

Challenges in Immersive Videos

- The human vision system can resolve 120 ppd at the fovea
- High dynamic range (HDR), high framerate (HFR)
- Low latency requirement for human interactions (head movements etc.)

In summary, huge bandwidth consumption under low latency constraints.

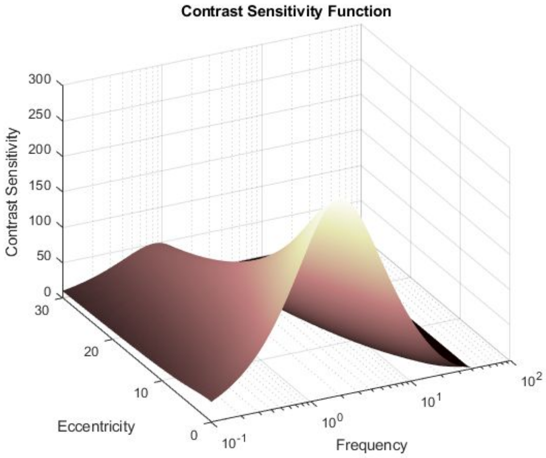
An Estimate of Video Bandwidth Consumption Based on YouTube Recommendation[1]. (Encoder: H264)

Spatial Resolution (ppd)	Temporal Resolution (fps)	Monitor Spatial Resolution (Non-VR)	Estimated Bitrate (Non-VR)	HMD Spatial Resolution (VR)	Estimated Bitrate (VR)
15	30	~ 1024x768	~ 1Mbps	~ 5400x2700	~ 40Mbps
30	30	~ 1920x1080	~ 4Mbps	~ 10800x5400	~ 160Mbps
	60		~ 6Mbps		~ 240Mbps
60	30	~ 3840x1920	~ 20Mbps	~ 20000x10000	~ 600Mbps
	60		~ 30Mbps		~900Mbps
	90		~ 40Mbps		~1200Mbps

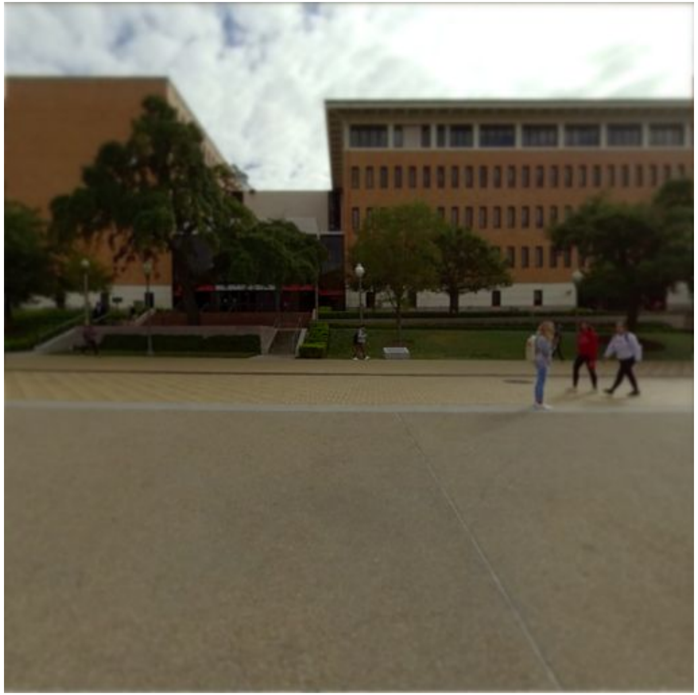
[1] https://support.google.com/youtube/answer/2853702

Foveated Video Compression

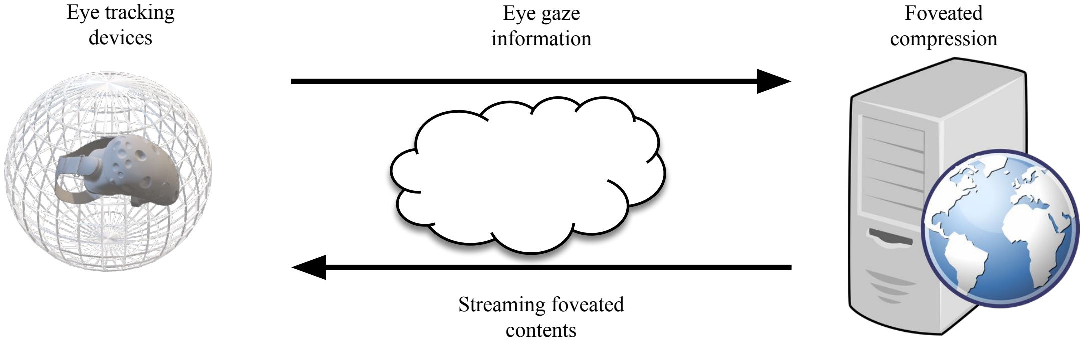
- The human vision system has decreasing acuity away from the foveal center
- Foveated video compression is regaining attention



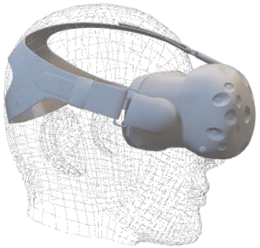
A Foveated Frame



Foveated Video Compression



- Earlier systems such as CSF-embedded multiresolution-based encoding systems: FMP[1], EFIC[2]…
- Foveated encoding systems based on modern codecs (AVC/HEVC): C-C. Ho et al.[3], J. Ryoo et al.[4], Romero-Rondon et al.[5], H. Kim et al.[6] …



[1] W. S. Geisler and J. S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," SPIE Conference on Human Vision and Electronic Imaging, 1998.
[2] Z. Wang and A. C. Bovik, "Embedded foveation image coding," IEEE Transactions on Image Processing, vol. 10, no. 10, pp. 1397-1410, 2001.
[3] Chia-Chiang Ho, Ja-Ling Wu, and Wen-Huang Cheng, "A practical foveation-based rate-shaping mechanism for MPEG videos," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 11, pp. 1365-1372, 2005.
[4] J. Ryoo, K. Yun, D. Samaras, S. R. Das, and G. Zelinsky, "Design and evaluation of a foveated video streaming service for commodity client devices," ACM International Conference on Multimedia Systems, New York, 2016
[5] M. F. Romero-Rondon, L. Sassatelli, F. Precioso, and R. Aparicio-Pardo, "Foveated Streaming of Virtual Reality Videos," ACM Multimedia Systems Conference, New York, 2018.
[6] H. Kim, J. Yang, M. Choi, J. Lee, S. Yoon, Y. Kim, and W. Park. "Eye tracking based foveated rendering for 360 VR tiled video," ACM Multimedia Systems Conference, New York, 2018.

Contents

- Introduction
  - Background
  - Challenges
  - Foveated Video Compression
- Proposed SVBRISQUE Model
  - Limitations of Traditional Models
  - Space-Variant Natural Scene Statistics
  - Summary of the Proposed Method
- Evaluation & Performance
  - Evaluation Framework
  - Results & Conclusion

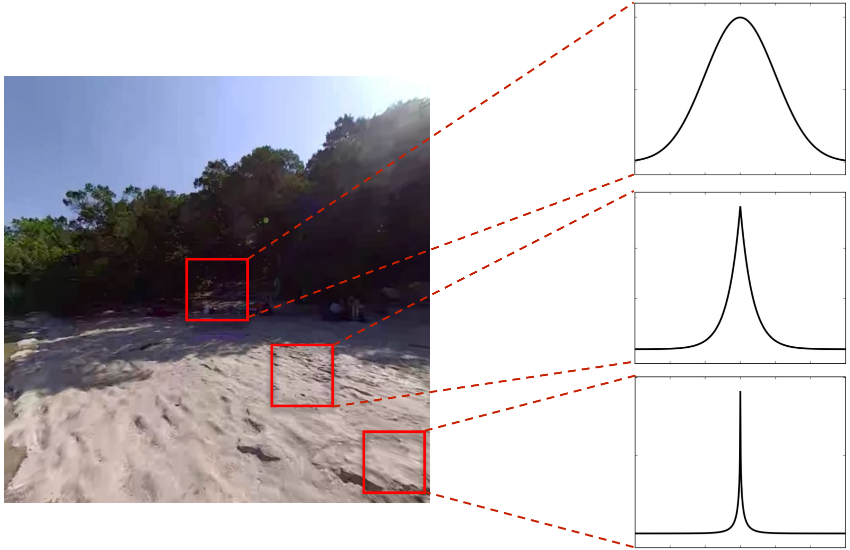
Limitations of Traditional Methods

- Natural Scene Statistics (NSS) have been successfully deployed in NR IQA / VQA models.
- Mean subtracted contrast normalized coefficients of natural images follow a standard Gaussian (normal) distribution, while distortions destroy this regularity.

MSCN Coefficients: $\hat{I} = \dfrac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}$

- Underlying assumption in traditional IQA/VQA: distortions and the NSS features are invariant across spatial domain.

No longer true for foveated videos!



A Foveated Image

MSCN Distribution

Space-Variant Natural Scene Statistics

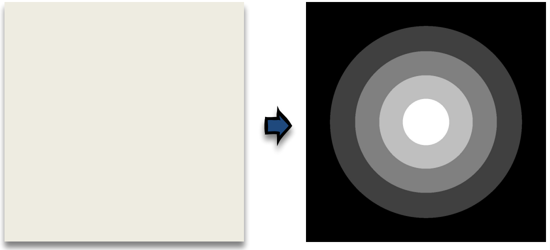
GGD and AGGD Models

$$\hat{I} = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}$$

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} I_{k,l}(i,j)$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} (I_{k,l}(i,j) - \mu(i,j))^2}$$

SV-GGD and SV-AGGD Models



GGD to SV-GGD

$$f(x;\alpha,\sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right) \longrightarrow \begin{aligned}&\hat{I}(\mathbf{r}) \sim f(x;\alpha(\mathbf{r}),\sigma(\mathbf{r})^2)\\ &= \frac{\alpha(\mathbf{r})}{2\beta(\mathbf{r})\Gamma(\frac{1}{\alpha(\mathbf{r})})} \exp\left(-\left(\frac{|x|}{\beta(\mathbf{r})}\right)^{\alpha(\mathbf{r})}\right)\end{aligned}$$

$$P(i,j) = \hat{I}(i,j)\hat{I}(i+d_1, j+d_2)$$

AGGD to SV-AGGD

$$f(x;\nu,\sigma_l^2,\sigma_r^2) = \begin{cases} \frac{\nu}{(\beta_l+\beta_r)\Gamma(\frac{1}{\nu})} \exp\left(-\left(\frac{-x}{\beta_l}\right)^{\nu}\right) & x < 0 \\ \frac{\nu}{(\beta_l+\beta_r)\Gamma(\frac{1}{\nu})} \exp\left(-\left(\frac{-x}{\beta_r}\right)^{\nu}\right) & x \geq 0 \end{cases} \longrightarrow$$

$$f(x;\nu(\mathbf{r}),\sigma_l^2(\mathbf{r}),\sigma_r^2(\mathbf{r})) = \begin{cases} \frac{\nu(\mathbf{r})}{(\beta_l(\mathbf{r})+\beta_r(\mathbf{r}))\Gamma(\frac{1}{\nu(\mathbf{r})})} \exp\left(-\left(\frac{-x}{\beta_l(\mathbf{r})}\right)^{\nu(\mathbf{r})}\right) & x < 0 \\ \frac{\nu(\mathbf{r})}{(\beta_l(\mathbf{r})+\beta_r(\mathbf{r}))\Gamma(\frac{1}{\nu(\mathbf{r})})} \exp\left(-\left(\frac{-x}{\beta_r(\mathbf{r})}\right)^{\nu(\mathbf{r})}\right) & x \geq 0 \end{cases}$$

Space-Variant Natural Scene Statistics

$$\hat{I}(\mathbf{r}) \sim f(x; \alpha(\mathbf{r}), \sigma(\mathbf{r})^2)$$

$$= \frac{\alpha(\mathbf{r})}{2\beta(\mathbf{r})\Gamma(\frac{1}{\alpha(\mathbf{r})})} \exp\left(-\left(\frac{|x|}{\beta(\mathbf{r})}\right)^{\alpha(\mathbf{r})}\right)$$

$$f(x; \nu(\mathbf{r}), \sigma_l^2(\mathbf{r}), \sigma_r^2(\mathbf{r})) =$$

$$\begin{cases} \frac{\nu(\mathbf{r})}{(\beta_l(\mathbf{r})+\beta_r(\mathbf{r}))\Gamma(\frac{1}{\nu(\mathbf{r})})} \exp\left(-\left(\frac{-x}{\beta_l(\mathbf{r})}\right)^{\nu(\mathbf{r})}\right) & x < 0 \\ \frac{\nu(\mathbf{r})}{(\beta_l(\mathbf{r})+\beta_r(\mathbf{r}))\Gamma(\frac{1}{\nu(\mathbf{r})})} \exp\left(-\left(\frac{-x}{\beta_r(\mathbf{r})}\right)^{\nu(\mathbf{r})}\right) & x \geq 0 \end{cases}$$

Smoothness assumption in concentric regions

Estimating the parameters of SV-GGD and SV-AGGD models.

- Maximum likelihood is a functional of these space-variant parameters.
- Locally stationary assumption. Can be extended to ring-shaped concentric regions.

Proposed SVBRISQUE Model

A Spatial Neural Noise Model

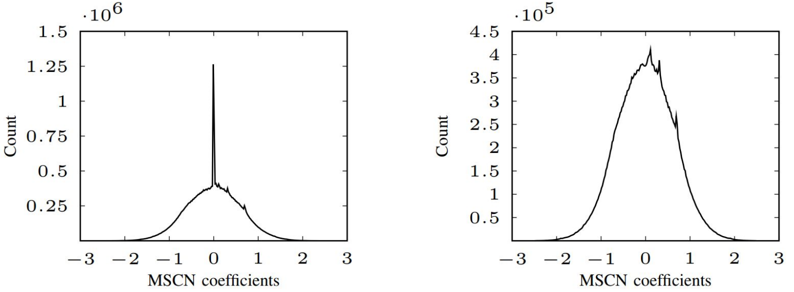
- Introduce a neural noise model.
- Account for uncertainty of visual perception.
- Add a small amount of variation on saturated, over- or under-exposed regions.

The Neural Noise Model

$$\hat{I} = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \quad \Rightarrow \quad \tilde{I}(i,j) = I(i,j) + \mathcal{W}_s \qquad \hat{I} = \frac{\tilde{I}(i,j) - \tilde{\mu}(i,j)}{\tilde{\sigma}(i,j) + C_s}$$

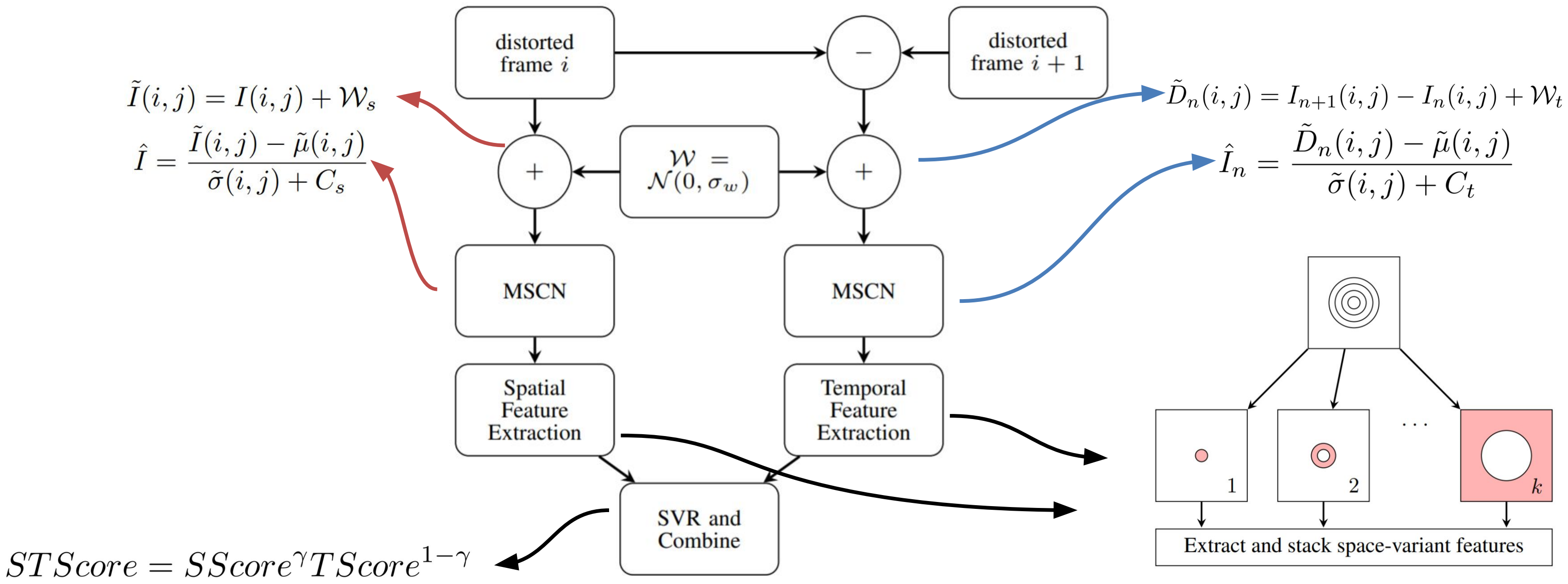
Saturated Regions in indoor and outdoor scenes

MSCN Distribution Before (left) and After (right) Applying Neural Noise Model



Summary of the Proposed Method

Overview of SVBRISQUE



$$\tilde{I}(i,j) = I(i,j) + \mathcal{W}_s$$

$$\hat{I} = \frac{\tilde{I}(i,j) - \tilde{\mu}(i,j)}{\tilde{\sigma}(i,j) + C_s}$$

$$\tilde{D}_n(i,j) = I_{n+1}(i,j) - I_n(i,j) + \mathcal{W}_t$$

$$\hat{I}_n = \frac{\tilde{D}_n(i,j) - \tilde{\mu}(i,j)}{\tilde{\sigma}(i,j) + C_t}$$

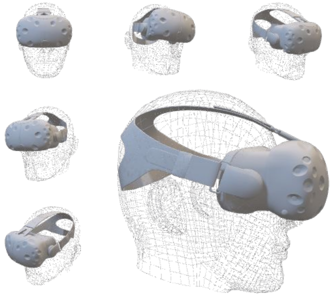
$$STScore = SScore^{\gamma} TScore^{1-\gamma}$$

Contents

- Introduction
  - Background
  - Challenges
  - Foveated Video Compression
- Proposed SVBRISQUE Model
  - Limitations of Traditional Models
  - Space-Variant Natural Scene Statistics
  - Summary of the Proposed Method
- Evaluation & Performance
  - Evaluation Framework
  - Results & Conclusion

Evaluation Framework

- The newly created LIVE-FBT-FCVR databases[1] were used to compare SVBRISQUE with existing foveated IQA/VQA models.
- To recover foveated viewing experience, we adopted a viewport-based assessment framework.
- We sampled 18 viewing directions, and created 18 foveated viewport videos for each distortion, whose field of view is 90° and resolution is 1024×1024.
- SVR with RBF kernel was used to train a model on the obtained features. Median performance was reported after 1000 train-test splits.



[1] Y. Jin, M. Chen, T. Goodall, A. Patney and A. C. Bovik, "Subjective and Objective Quality Assessment of 2D and 3D Foveated Video Compression in Virtual Reality," in *IEEE Transactions on Image Processing*, vol. 30, pp. 5905-5919, 2021

Results & Conclusion

	Methods	SROCC↑	KROCC↑	PLCC↑	RMSE↓
2D	BRISQUE	0.797±0.22	0.639±0.18	0.708±0.18	9.60±3.29
	SVBRISQUE	**0.900±0.11**	**0.736±0.12**	**0.884±0.10**	6.91±2.53
	NIQE	0.605±0.32	0.457±0.24	0.675±0.31	**6.47±2.27**
	V-BLIINDS	0.440±0.25	0.327±0.20	0.431±0.25	11.11±2.07
	TLVQM	0.509±0.36	0.381±0.26	0.470±0.36	10.38±3.09
	FWQI	0.791	0.785	0.591	–
	FASSIM	0.757	0.742	0.553	–
3D	BRISQUE	0.751±0.19	0.587±0.15	0.699±0.17	9.11±2.97
	SVBRISQUE	**0.875±0.12**	**0.695±0.12**	**0.877±0.12**	**5.99±1.79**
	NIQE	0.732±0.19	0.570±0.15	0.781±0.17	6.59±2.00
	V-BLIINDS	0.391±0.24	0.283±0.18	0.300±0.22	9.33±1.50
	TLVQM	0.696±0.21	0.517±0.17	0.699±0.21	7.82±2.43
	FWQI	0.804	0.784	0.592	–
	FASSIM	0.755	0.740	0.543	–

Conclusion

We have designed a blind foveated VQA model called SVBRISQUE which relies on space-variant NSS. The model achieved SOTA performance on LIVE-FBT-FCVR databases.

Limitation

- The proposed method rely on foveation radii information inherited from the databases.

Future Directions

- Remove the dependence on foveation radii information.
- Further improve the performance.

Thank you!