Modeling Sparse Spatio-Temporal Representation for No-Reference Video Quality Assessment

Muhammed Shabeer P
Indian Institute Of Technology Hyderabad
ee15mtech11008@iith.ac.in

Introduction

- Video content generation and consumption continues to grow exponentially
- Objective Video Quality Assessment (VQA) an indispensable tool for content management
- No-reference VQA (NRVQA) especially important when pristine source unavailable - a very common occurrence in reality
- We present a sparsity based NRVQA algorithm

Background

- NRVQA algorithms rely on finding distortion discriminative features - handcrafted and machine learnt [1, 2, 3]
- Supervised learning of functional relationships between features and Difference Mean Opinion Scores (DMOS)
- The Human Visual System (HVS) hypothesized to sparsely represent visual stimulus [4]
- Several sparsity based image QA algorithms proposed [5]
  - Hypothesis is that sparse representations are distortion discerning
- Proposed sparsity based NRVQA algorithm among the first of its kind

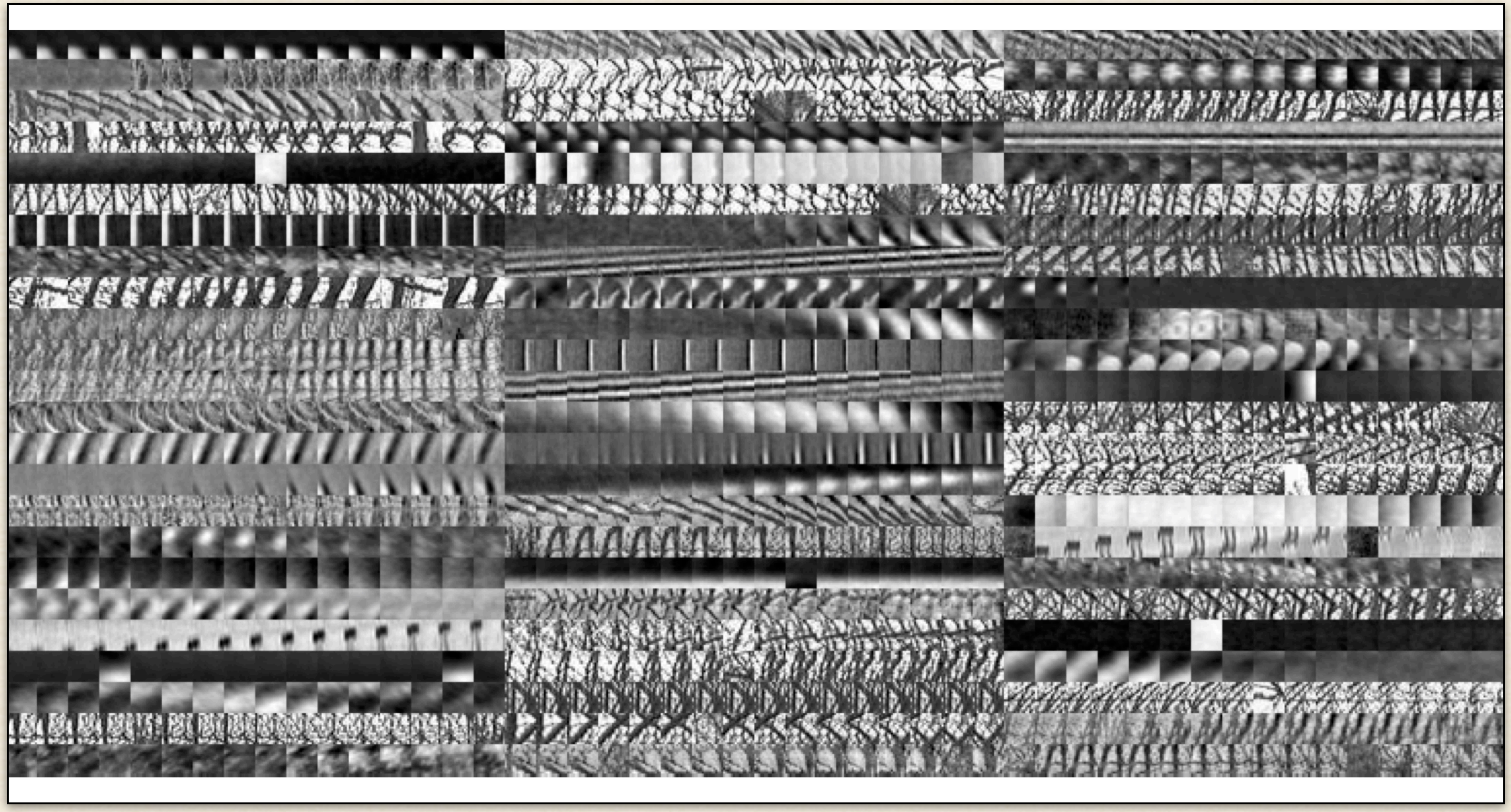
Sparse Representation of Spatio-Temporal Volumes

- A spatio-temporal volume is expressed in terms of a linear combination (using a_i) of atomic volumes (ϕ_i) from an overcomplete dictionary:

$$V(x,y,t) = \sum_i a_i \phi_i(x,y,t)$$

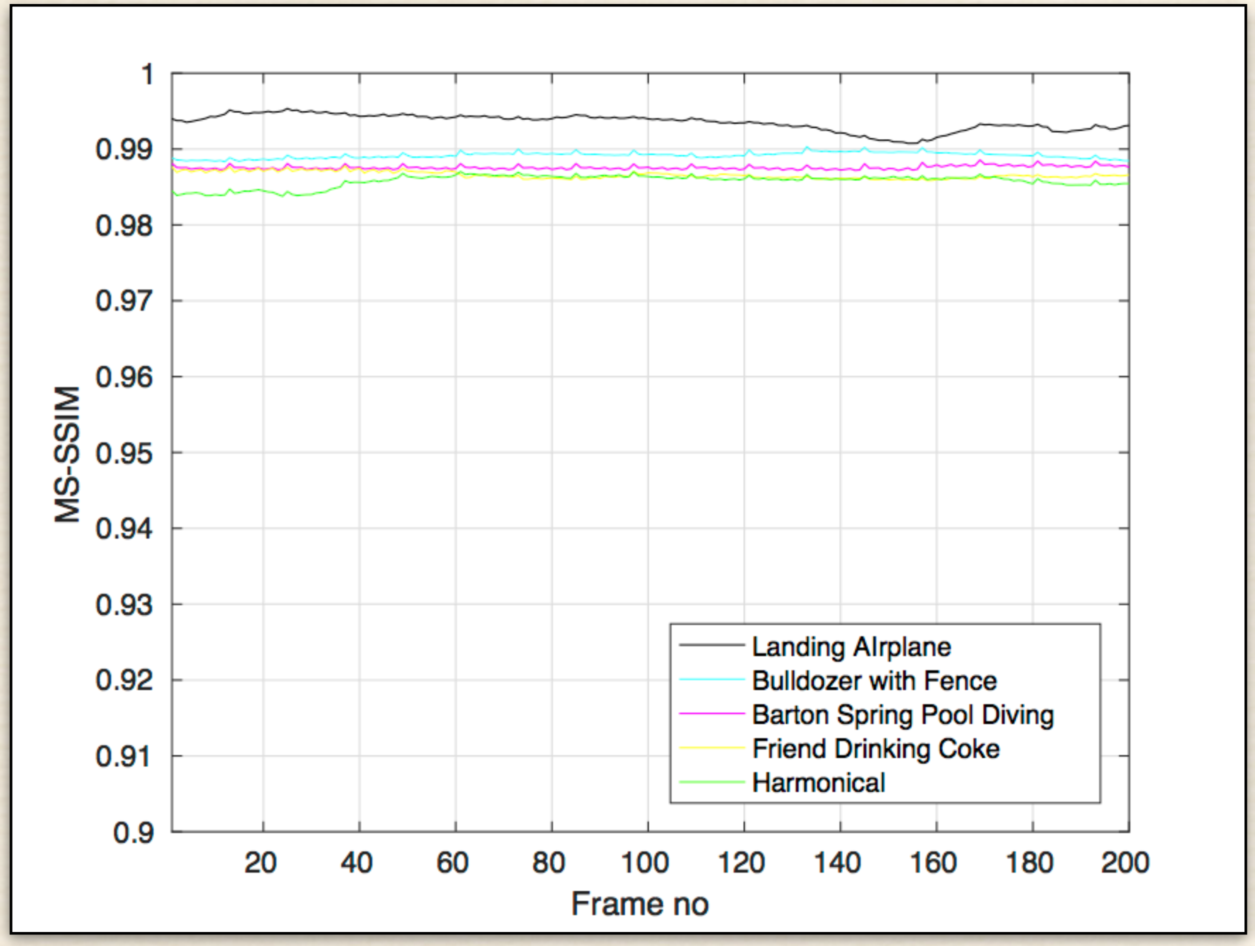
- The dictionary of video volumes is constructed using pristine video volumes
- The KSVD algorithm [6] used for t construction
  - Volume unwrapped into a vector and standard approach followed
  - Unwrapping in a particular order retains spatio-temporal correlation
  - Dictionary size $N \times 2N$
  - Various volume sizes (x, y, t) considered: **5 x 5 x 3** to **16 x 16 x 16**
  - Pristine videos from LIVE SD Video Database [8]

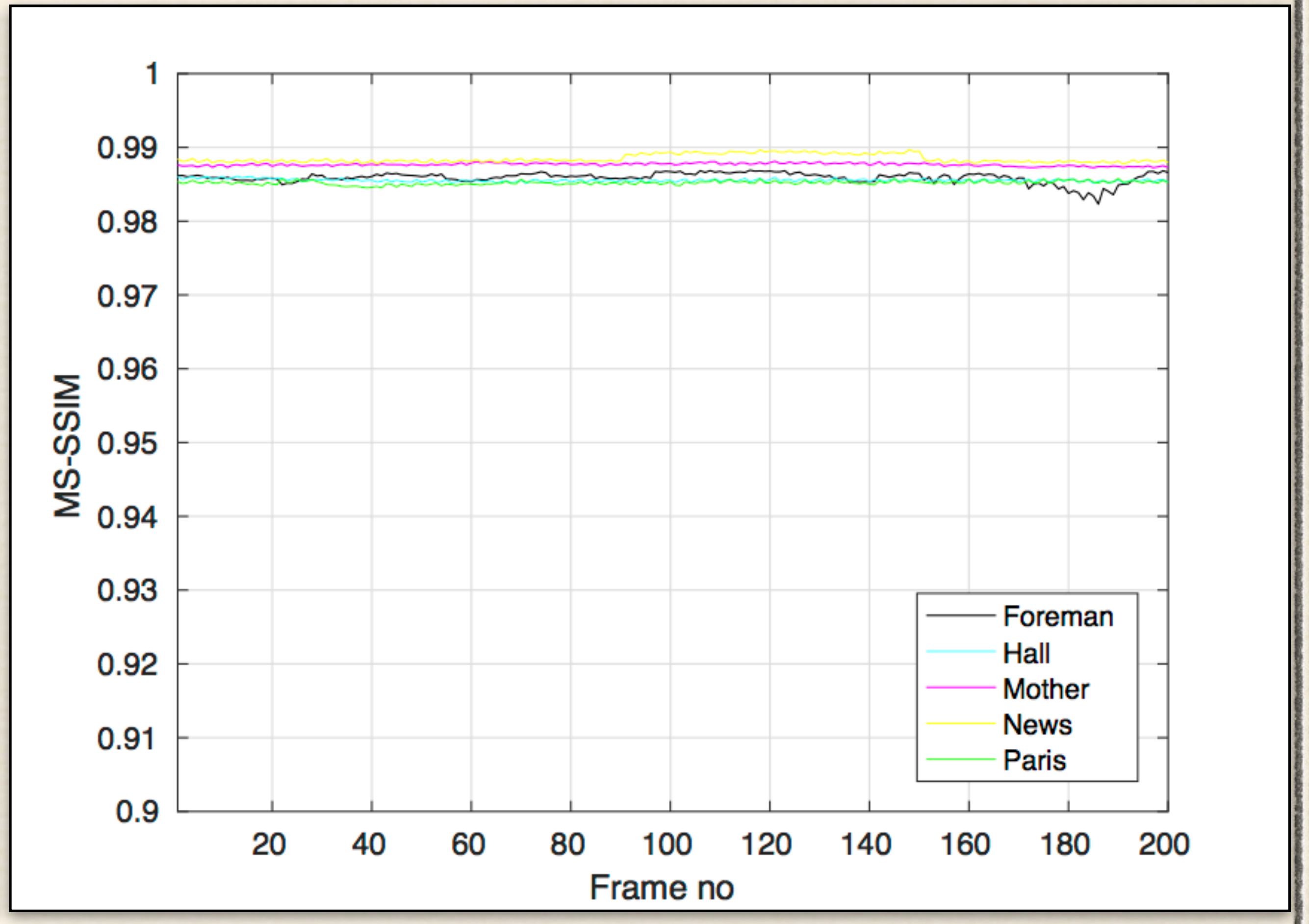
72 atoms from Dictionary of size16 x 16 x 16



Robustness and Reliability of Dictionary

- Volume size of **5 x 5 x 3** was found to give best reconstruction performance. Plots show frame-wise MS-SSIM index [7]





- LIVE Mobile HD Database [9]

EPFL Database [10]

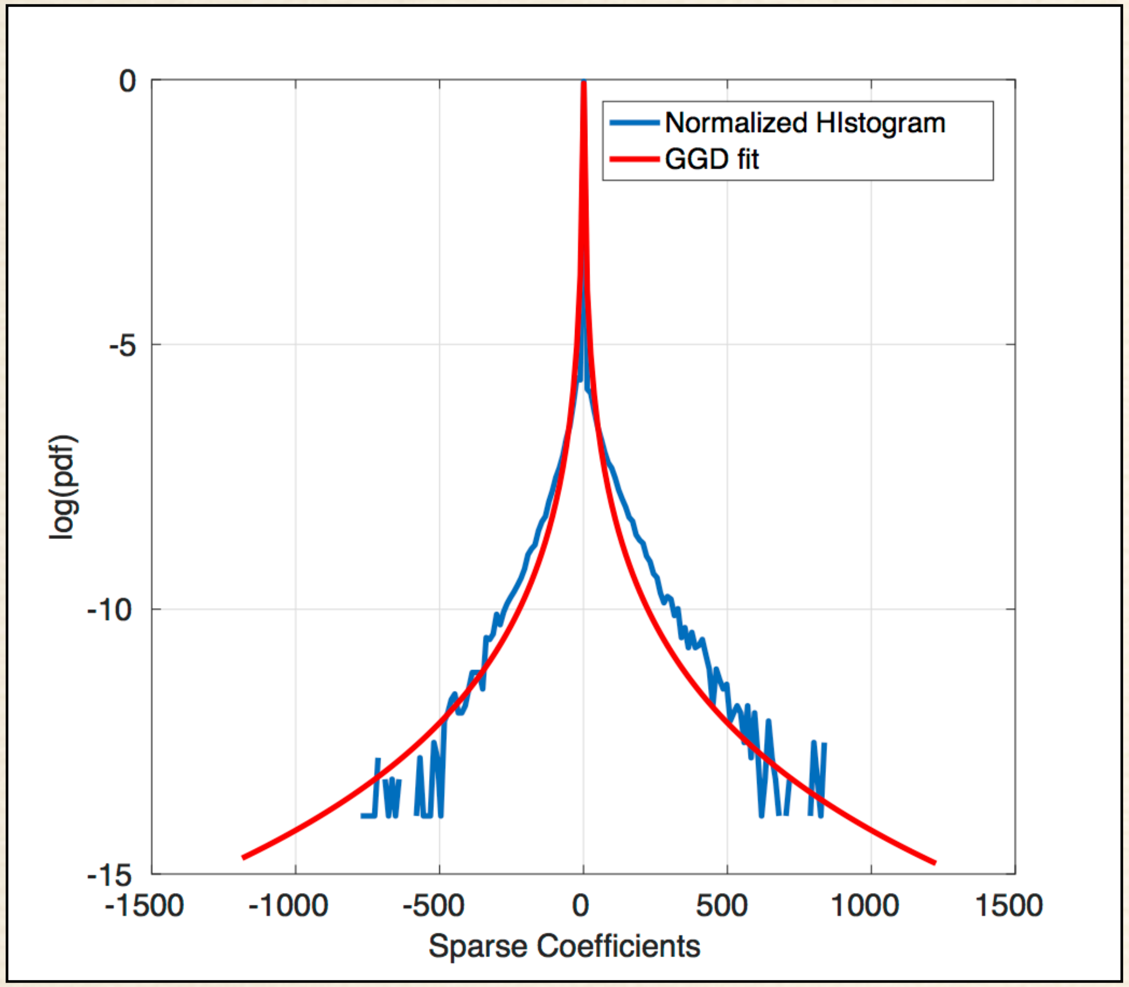
Proposed NRVQA Algorithm

- NRVQA algorithm uses dictionary with atom size **5 x 5 x 3**
- The histogram of sparse coefficients corresponding to each atom is modeled using a generalized Gaussian distribution:

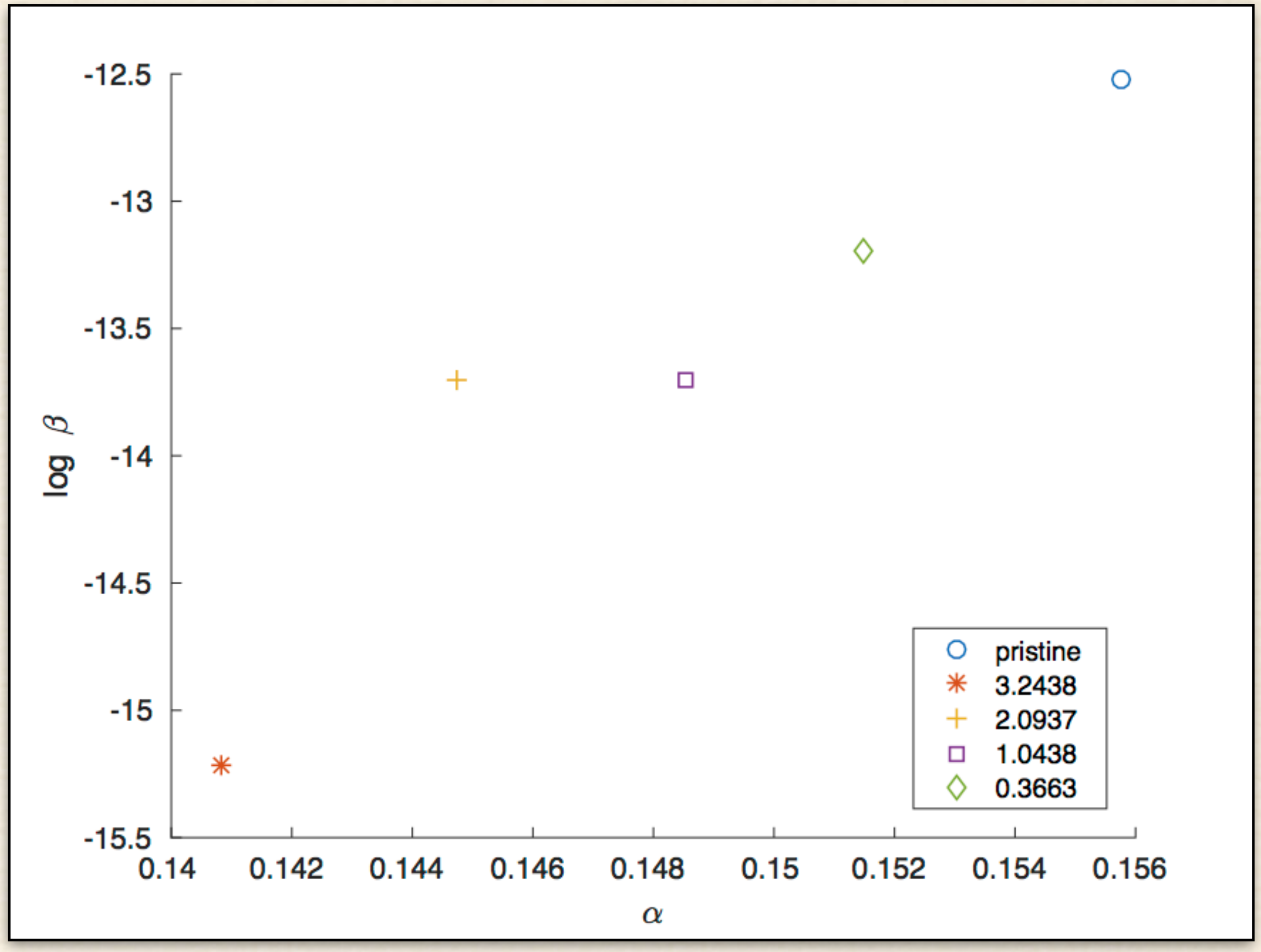
$$f(x;\alpha,\beta)=\frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)}exp\left(-\frac{|x|}{\beta}\right)^{\alpha}$$

- The model parameters(α,β) serve as excellent distortion discriminatory features
- These features are used for supervised learning of DMOS labels using support vector regression (SVR)

GGD Model and its Effectiveness on the LIVE Mobile Database [10]



Model fit for Atom 1



Distortion discrimination (legend DMOS)

Evaluation Databases

Database	# Videos (Distorted + Reference)	Frame Rate (FPS)	Resolution	Distortions
LIVE SD [8]	150 + 10	25/50	768 x 432	MPEG2, H264, IP, Wireless
EPFL SD [9]	144 + 12	30	352 x 288/ 704 x 576	Packet loss
LIVE Mobile HD [10]	160 + 10	30	1280 x 720	Compression, Rate adaptation, Temporal dynamics, wireless

Results on SD Databases

	LIVE SD [8]		EPFL SD [9]	
	LCC	**SROCC**	**LCC**	**SROCC**
NIQE [11]	0.2668	0.2250	0.5160	0.4998
VIIDEO [1]	0.6510	0.6240	0.1840	0.2025
Video BLIINDS [2]	0.8810	0.7590	0.7520	0.8070
FLOSIM-NR [3]	0.6076	0.5864	0.8915	**0.8961**
Lie et al. [12]	**0.8910**	**0.7820**	0.8050	0.7960
Proposed	*0.7082*	*0.6621*	***0.9107***	*0.8764*

Results on HD Database

	LIVE Mobile HD [10]		LIVE Tablet HD [10]	
	LCC	**SROCC**	**LCC**	**SROCC**
NIQE [11]	0.7560	0.7410	0.7569	0.7559
VIIDEO [1]	0.2451	0.2164	0.5430	0.5027
Video BLIINDS [2]	0.3734	0.4392	-	-
FLOSIM-NR [3]	0.8450	0.8352	0.9140	0.8647
Proposed	***0.9253***	***0.9007***	***0.9686***	***0.9382***

Results: Computational Cost

	LIVE SD [8]	
	Time/Video (secs)	**Improvement (%)**
Video BLIINDS [2]	311	90
VIIDEO [1]	132	77
FLOSIM-NR [3]	43	28
Proposed	***31***	-
Tested on: 3.1 GHz Intel Core i7, 16 GB RAM, Ubuntu 16.04		

Conclusions

- Developed a NR-VQA algorithm based on natural video statistical features
- A spatio-temporal dictionary designed for sparsely representing natural video volumes
- Modeled sparse coefficients using a GGD
- GGD model parameters able to discern spatial and temporal distortions jointly
- A simple and computationally efficient NR-VQA algorithm dubbed SParsity based Objective VIdeo Quality Evaluator (SPOVIQE)
    - Supervised learning (model parameter features, DMOS labels) using SVR
- SPOVIQE has very competitive performance on the LIVE SD, EPFL-SD and LIVE HD databases in addition to having low computational complexity

References

1. Anish Mittal, Michele A Saad, and Alan C Bovik, “A completely blind video integrity oracle,” IEEE Transactions on Image Processing, vol. 25, no. 1, pp. 289–300, 2016.

2. Michele A Saad, Alan C Bovik, and Christophe Charrier, “Blind prediction of natural video quality,” IEEE Transactions on Image Processing, vol. 23, no. 3, pp. 1352–1365, 2014.

3. K Manasa and Sumohana S Channappayya, “An optical low-based no-reference video quality assessment algorithm,” in Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016, pp. 2400–2404.

4. Bruno A, Olshausen et al., “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” Nature, vol. 381, no. 6583, pp. 607–609, 1996.

5. KVSNL Manasa Priya, Balasubramanyam Appina, and Sumohana Channappayya, “No-reference image qual- ity assessment using statistics of sparse representations,” in Signal Processing and Communications (SPCOM), 2016 International Conference on. IEEE, 2016, pp. 1–5.

6. Michal Aharon, Michael Elad, and Alfred Bruckstein, “k-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Transactions on signal processing, vol. 54, no. 11, pp. 4311– 4322, 2006.

7. Zhou Wang, Eero P Simoncelli, and Alan C Bovik, “Multiscale structural similarity for image quality assessment,” in Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on. IEEE, 2003, vol. 2, pp. 1398–1402

8. Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K Cormack, “Study of subjective and objective quality assessment of video,” IEEE transactions on image processing, vol. 19, no. 6, pp. 1427–1441, 2010.

9. Francesca De Simone, Matteo Naccari, Marco Tagliasacchi, Frederic Dufaux, Stefano Tubaro, and Touradj Ebrahimi, “Subjective assessment of h. 264/avc video sequences transmitted over a noisy channel,” in Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on. IEEE, 2009, pp. 204–209.

10. Anush Krishna Moorthy, Lark Kwon Choi, Alan Con- rad Bovik, and Gustavo De Veciana, “Video quality assessment on mobile devices: Subjective, behavioral and objective studies,” IEEE Journal of Selected Topics in Signal Processing, vol. 6, no. 6, pp. 652–671, 2012.

11. Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, “Making a completely blind image quality analyzer,” IEEE Signal Processing Letters, vol. 20, no. 3, pp. 209–212, 2013.

12. Xuelong Li, Qun Guo, and Xiaoqiang Lu, “Spatiotemporal statistics for video quality assessment,” IEEE Transactions on Image Processing, vol. 25, no. 7, pp. 3329–3342, 2016.