No-reference Video Quality Assessment for Noise, Blur, and MPEG2 Natural Videos

Department of Scientific Computing

Ji Shen & Qin Li   Advisor: Gordon Erlebacher

Email: js04j@fsu.edu; qli@math.fsu.edu

Florida State University

Abstract

In this paper, we propose a new no-reference VQA metric, called Video Hybrid No-reference (VHNR) method. It is based on natural video statistics built from the coefficients of 3D curvelet and cosine transforms. VHNR can blindly predict the quality of noisy, blurry, or MPEG2 compressed videos and requires no original reference video. The 3D curvelet transform is known to be sensitive to surface singularities, generated by noise or blur artifacts in the videos. On the other hand, the cosine transform is well-suited to detect MPEG2 compression artifacts. No-reference works because we studied tens of thousands of distorted videos and obtained a statistical relation between the video quality, the specific video characteristics in the transformed spaces, and the video motion speed. Intensive computations are required to analyze tens of thousands of simulated high resolution videos. Since 3D curvelet transform of each video requires 6GB memory and large amounts of computation time, the algorithm is implemented on the FSU High-Performance Computing (HPC) using MPI. The parallelism reduces the computational time of the whole experiment from 118 days to 9 days (a speedup of 13).

Introduction of Video Quality Assessment

When taking videos by a digital camcorder, there are always video distortions, for example, noise, blur, compression, or some combination thereof. How does one quantify such distortions, or the video quality? A video quality assessment (VQA) metric analyzes the video and assigns a numerical score to the video quality. There are three categories of VQA metrics: full-reference, reduced-reference, and no-reference. Full-reference methods, such as PSNR, compare the full set or subset of original high quality video and the distorted video. However, the original video often does not exist, such as YouTube videos or video taken by low quality camcorders. In such cases, no-reference VQA is the only method to give the quality score. Most no-reference VQA analyze the artifact patterns in the video; therefore the methods usually only work for one particular type of filter. The proposed algorithm has been applied successfully to images filtered by noise, blur, or MPEG2.

Introduction of 3D Curvelet Transform

We transform the videos into curvelet space to retrieve the features for each video. The curvelet transform is a recent transform which has been proved to have the best convergence properties when representing the surface-like singularities. In a 3D Cartesian grid, the 3D curvelet transform [1, 2] is defined as

$$\theta(j,\,l,\,k) := \sum_{0<n_1,n_2,n_3<n} f(n_1,n_2,n_3)\overline{\phi_{j,l,k}(n_1,n_2,n_3)}$$

where $j, l \in Z$ and $k = (k_1, k_2, k_3)$. We can think of the 3D curvelet transform as a convolution of the curvelet function and the video function. A curvelet decays along two axes and is oscillatory along the third orthogonal axis. And it has multi-scale, multi-angle, and multi-location. If a curvelet overlaps a surface singularity whose shape, angle and location correlates well with the curvelet, the corresponding curvelet coefficient is large; otherwise, the coefficient is nearly zero. This property is used to advantage when analyzing filtered videos.

VHNR Model

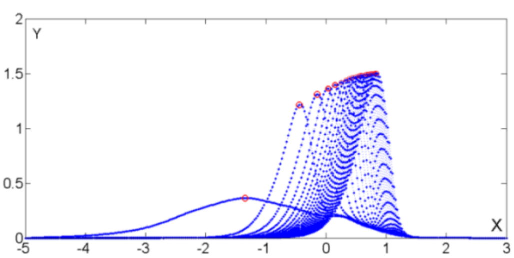

Fig. 1. LPMCCs of 25 noisy videos and their common original video. From bottom left to top right, the LPMCCs correspond to videos with increasing noise level. The top point (red circled) of LPMCC serves as the feature of each video.

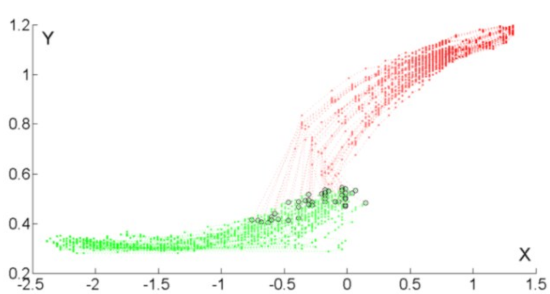

Fig. 2. Feature distribution of 10400 noisy (red) and blur (green) videos. The dots followed by the dotted track correspond to a series of videos filtered from the same original (black) video.

In [3], we proposed the Hybrid No-reference (HNR) model for natural images. In this paper, we extend the framework to video quality assessment. The feature of each video is a 4D vector. Once a 3D curvelet transform is applied onto the video, we take the peak of the log-Pdf of the magnitude of the curvelet coefficients (LPMCC) from each curvelet scale to form the feature vector (Figure 1). After getting the features for tens of thousands of videos, in the feature space of each scale, we obtain the relation between video quality and feature values $F(l, c) = 0$ (Figure 2) where l is video quality and c is video feature. As a result, after a training process to build the relation of F, for any new arbitrary video c' with feature c', we can approximate l' using F. We use curvelet coefficients to predict the noise or blur level but DCT coefficients to predict the MPEG2 compression level. A similar approach is applied to get the feature of MPEG 2 videos when using DCT coefficients.

HPC Parallel Implementation

HPC stands for High Performance Computing. The videos in VIVID have a size of $512 \times 512 \times 256$. The curvelet transform (CurveLab out-core version) of one video requires at least 6GB memory and takes about 600 seconds using a 2GHz level CPU. In order to build up a reliable statistical feature of various types of videos, our testing requires transforming tens of thousands of videos. We parallelized our algorithms to run on HPC, for efficiency.

Our software has three components (Figure 3): (1) P1: C++ MPI version of 3D curvelet transform or Matlab 3D DCT transform, (2) P2: Matlab video simulation program instances running simultaneously, (3) P3: the Matlab VHNR main program.

P2 builds a video pool that is always ready to feed the next request videos. P1 applies the 3D curvelet (using CurveLab MPI version) or DCT transform to each video and outputs the pdf of the coefficients. P3 does the training, prediction, and error estimation based on the pdfs. Without HPC, the total computational time is about 118 days. Running in parallel using 21 cores, the time is reduced to nine days.

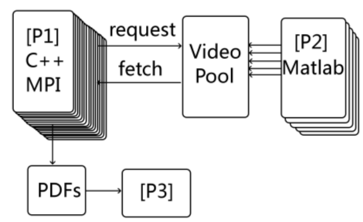

Fig. 3. Implementation flowchart on HPC

Results

Errors between the exact FLs of simulated videos and the predicted FLs are compared. Following the Video Quality Experts Group (VQEG) Phase II tests, we use Pearson's Linear Correlation Coefficient (CC) to analyze the prediction accuracy.

The CC of VHNR is very high for noise and fairly good for blur and MPEG2 videos. Compared to PSNR, our proposed method offers substantial improvements.

Model (FR/RR/NR)	Noise	Blur	MPEG2
Proposed VHNR	0.982	0.931	0.921
PSNR (FR)	0.836	0.560	0.838

Table. 1. Correlation Coefficient between exact and predicted video quality. (5200 videos per filter).

Conclusion

In this poster, we presented VHNR, a new no-reference video quality metric. The main contributions of this poster are threefold. First, we studied the natural video statistics (NVS) of the log-Pdf of the transformed coefficients in the transformed space and obtained the relation between video quality and the Pdf of curvelet coefficients, which makes no-reference assessment possible. Compared to other full-reference methods, VHNR does not require a reference video, yet it has the prediction correlation coefficients over 0.92 for all the three filters when tested on the VIVID video library. Second, the VIVID library can simulate 15600 various quality videos with different foregrounds, backgrounds, motion speeds, filters, and filter levels. It is very helpful for developing and testing video quality metrics, especially for training-based methods. Third, a parallel video analysis program is developed on HPC with interaction between Matlab and C++ programs. It makes possible the intensive computation of tens of thousands of videos to process reliable training data necessary for accurate quality assessment of videos.

References

[1] L. D. L. Ying and E. Candes. "***3d discrete curvelet transform***". In Proc. SPIE Wavelets XI, 2005.

[2] E.J. Candes and D. L. Donoho, "***New tight frames of curvelets and optimal representations of objects with piecewise-C2 singularities***". Comm. on Pure and Appl. Math. 57 (2004), 219–266.

[3] Ji. Shen, Qin Li, and Gordon Erlebacher. "***Hybrid no-reference natural image quality assessment of noise, blur, jpeg2000, and jpeg images***".