

Technical Disclosure Commons

Defensive Publications Series

May 01, 2017

A No-Reference Video Quality Predictor For Compressed Videos

Sasi Inguva

Chao Chen

Anil Kokaram

Follow this and additional works at: http://www.tdcommons.org/dpubs_series

Recommended Citation

Inguva, Sasi; Chen, Chao; and Kokaram, Anil, "A No-Reference Video Quality Predictor For Compressed Videos", Technical Disclosure Commons, (May 01, 2017)
http://www.tdcommons.org/dpubs_series/487



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

A NO-REFERENCE VIDEO QUALITY PREDICTOR FOR COMPRESSED VIDEOS

ABSTRACT

A system and method to predict the perceptual quality of a compressed video by deploying a self-reference technique is disclosed. The method includes the steps of computing a frame difference image from the luminance component of at least one other alternate frame of an input test image. A blurred frame and a blurred frame difference image are then obtained by low-pass filtering of the input and frame difference images. A divisive normalization operator is applied on the four types of images independently and a generalized Gaussian distribution GGD fitted. Spatial features and temporal features are then extracted from the GGD. Absolute differences between the spatial and temporal features are computed and weighted based on motion in a given frame in the video. These features are pooled over different patches across the frames to obtain a final video quality score Q . The method shows superior results when compared to existing methods, while being computationally simple.

BACKGROUND

With the pervasiveness of visual media, measuring and improving the quality of images and videos are receiving increased attention. The quality of images is measured using objective quality assessment algorithms. Since the final receivers of visual media are human observers, the goal of objective quality assessment algorithms is to accurately predict the quality of images and videos as perceived by human observers. Depending on the amount of information available, objective video quality algorithms are divided into full-reference, reduced-reference, and no-reference algorithms. In the case of no-reference algorithms, except the input video whose quality needs to be predicted, the algorithm does not have any

additional information. Quality predictors such as SSIM, VQM, and PSNR are full-reference quality algorithms where, in addition to the distorted video, the original undistorted video is also assumed to be available. Though these algorithms are very accurate in predicting perceived quality, the dependency on the original undistorted video greatly limits the applicability of these models in numerous applications.

DESCRIPTION

A system and method are disclosed that predict the perceptual quality of a compressed video without reference to the original or undistorted version of the video. A self-reference technique as shown in FIG. 1 is deployed to extract video quality features from an input test video.

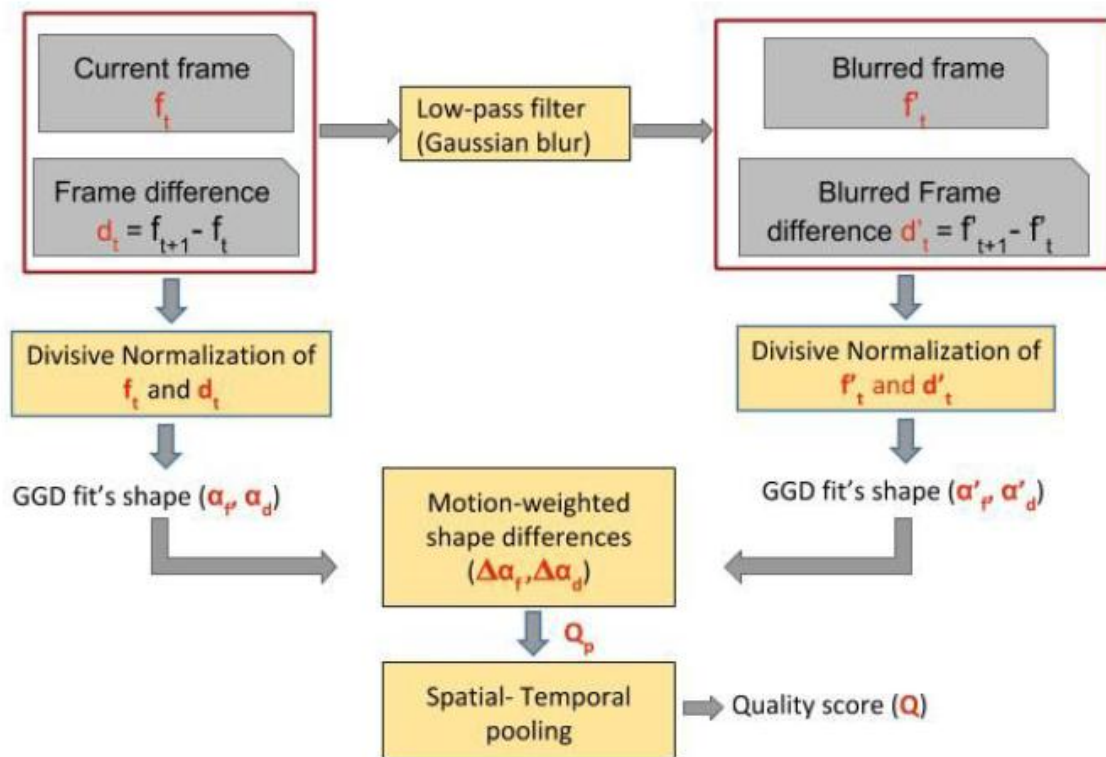


FIG. 1: Self-reference technique to extract quality features from an input video

The method includes the steps as shown in FIG. 2. At least one other alternate frame,

in this case, every alternate frame f_t from an input test video is considered. The luminance component of every alternate frame f_t of the video is measured. A frame difference image d_t is computed as the difference between the luminance components of the current frame and the next frame and is given by $I_{2t} = I_{2t+1} - I_{2t}$. A blurred frame f'_t and a blurred frame difference image d'_t is obtained by low-pass filtering of f_t and d_t respectively. A divisive normalization operator is applied on the images f_t , f'_t , d_t , and d'_t independently to obtain the histograms of normalized coefficients. A generalized Gaussian distribution GGD is fit to the resulting histograms. A total of four shape features α_s , α'_s , α_d and α'_d are extracted from the GGD. α_s and α'_s are spatial features extracted from the GGD of f_t and f'_t , respectively and α_d and α'_d are the temporal features extracted from d_t and d'_t respectively. An absolute difference between the spatial features $\Delta\alpha_f$ and temporal features $\Delta\alpha_d$ is computed and weighted based on motion in a given frame in the input test video. A final quality score Q is then obtained by pooling these weighted spatial and temporal features over different patches and across the frames in the video.

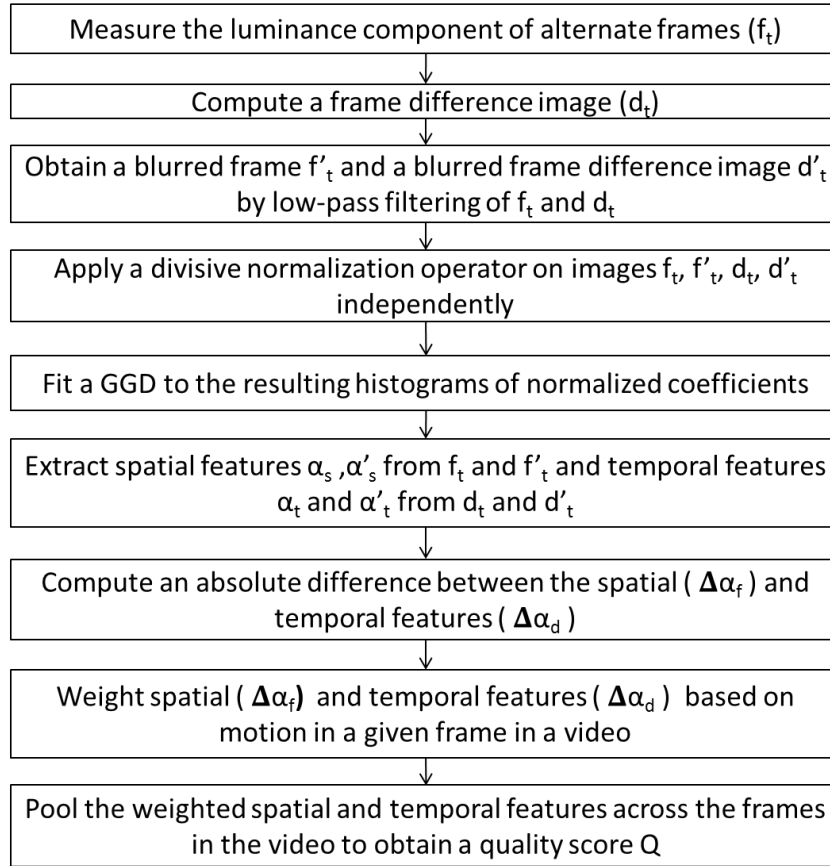


FIG. 2: Method of self-reference to extract video quality features from an input video

In order to compute the spatial and temporal feature differences $\Delta\alpha_f$ and $\Delta\alpha_d$, the frames f_t and the difference image d_t are divided into patches of predetermined size $B \times B$. In parts of the video where there is little motion, the frame differences do not contain sufficient information to capture structural regularity or irregularity and in such scenarios, the spatial quality takes precedence. However, in parts of the video where there is significant motion and in the presence of visual distortion the perceptual quality of a video could further be degraded. Hence in these parts temporal features $\Delta\alpha_d$ take precedence. The spatial and temporal features are weighted as given by:

$$Q_p = (1-m_p) * \Delta\alpha_f + m_p * \Delta\alpha_d \dots \dots \dots (1)$$

where m_p is the normalized average frame difference in a given patch p , and $\Delta\alpha_f$ and $\Delta\alpha_d$ are

the average spatial and temporal features computed in the patch p .

Furthermore, when the patches lack structure i.e. edge information, an average standard deviation σ of every patch in the frame f_t and the blurred frame f'_t are computed. An average difference of these σ s are computed, thus generating a difference of σ map (DoS), for every frame that is considered. Patches whose difference of σ value is below a p -th percentile of all the DoS values of the entire video are excluded. This technique may eliminate patches which are relatively smooth with respect to all the patches of any given video. The average of the motion-weighted features Q_p from the filtered patches of the frame and across all the frames in a given video are computed to obtain a single quality feature Q .

The system may be used to analyse distortions in user generated video contents for massive video processing such as cloud-based transcoding. . Advantages of the method are that the low pass filter, divisive normalization, GGD parameters, are computationally very simple. Additionally, these operations are independent of each other a patch level and also at frame level. This offers us a great scope for parallelization of these operations to further speed up the execution time and deploy it in real-world applications.