VIDEO QUALITY ASSESSMENT ON CELL

Maja Pokrić, Dragan Kukolj, Ivitan Pap, Nemanja Lukić, Nikola Teslić, Miodrag Temerinac, Zoran Marčeta, Vladimir Zlokolica

Faculty of Technical Sciences, Novi Sad, Serbia

ABSTRACT

This paper presents a novel system for a real-time no-reference video quality monitoring on a commercially available multi microprocessor platform. System design stage incorporates, firstly, automated statistically based selection of appropriate objective measures from a large predefined set for characteristic artifacts. Secondly, this reduced set of objective measures is used in conjunction with subjective assessments for derivation of a non-linear model which can then be used for future predictions. Different sets of relevant objective measures are used for video material with various contents, and hence for derivation of content based non-linear video assessment models. The system is capable of capturing live video, selecting appropriate assessment models and extracting video quality indicators in real-time.

1. INTRODUCTION

There is an increased need to measure and assess quality of the video sequences, which greatly depends on the video codec, bit-rates used for video coding and context of the video material. User oriented video quality assessment (VQA) provides means for monitoring of perceptual service quality. There are many papers that discuss different measures for artifacts which appear in coded video sequences, such as ringing, blurring and blocking [1]-[10].

In the paper we describe how the objective measures can be used in conjunction with the subjective VQAs, for the selection of the ‘optimal’ set of measures and generation of a content based non-linear VQA models. The selection of smaller subset of objective measures is performed through statistical analysis and it was concluded that the appropriate set of measures is content dependent. Hence, the framework models are computed only once in the training phase and are based on a video content and set of video artifacts considered. The presented system is aimed to capture and measure quality of standard and high definition video in a real-time [11] making it possible to estimate the video quality.

2. METHODOLOGY

In the proposed no-reference video quality assessment scheme we consider blurring, ringing and blockness artifacts present in coded video sequences. The general block-scheme of the proposed framework for a video quality assessment system training phase is shown in Fig. 1. Overall we are aiming to derive non-linear model which will derive quality of a video material based on previously assessed data. The following subsections describe each stage of a system training phase.

---

1 Cell Broadband Engine (CBE), IBM
2.1 Objective Assessment

There is an ever present need to derive a set of objective measures for image degradation, reflected by the appearance of image artifacts, due to encoder/decoder implementations as part of transport stream at various bit rates. Blocking appears in all block-based compression techniques due to coarse quantization of frequency components [1, 4, 8]. It can be observed as surface discontinuity (edge) at block boundaries. These edges are perceived as abnormal high frequency components in the spectrum. Ringing is observed as periodic pseudo edges around original edges [9]. It is due to improper truncation of high frequency components. This artifact is also known as the Gibbs phenomenon or Gibbs effect. In the worst case, the edges can be shifted far away from the original edge locations. This effect is observed as false edge. Blurring appears as edge smoothness or texture blur, is due to the loss of high frequency components when compared with the original image [10]. Blurring means that the received image is smoother than the original one.

In our approach we consider 21 different objective measures as a starting set of feature-based quality descriptors which describe dominant artifacts in video sequences. Some of these descriptors, to name just few, are: blockiness measure [1,4,8], edge activity [2], and homogeneity measure based on co-occurrence matrix [3]. In order to be able to capture motion characteristics of video material we included the motion intensity, which is a measure of the average magnitude of motion in a frame. This measure is separated into two categories, namely: (i) the global motion intensity, calculated from the global motion field, and (ii) the object motion intensity, calculated by first subtracting the global motion from the MPEG motion vectors [5].

2.2 Subjective Assessment

The most representative approach of assessing the quality of a video is subjective evaluation by human observers as ultimately their assessment will truly reflect viewer perception. The mean opinion score (MOS), which is a subjective quality measure obtained by averaging scores from a number of human observers, is derived from tests created according to ITU-R BT.500-10 [12] recommendations. For the purpose of initial system training, a number of SD sequences made available by Video Quality Experts Group (VQEG) [13] and HD sequences where encoded and decoded by MPEG2 video codec at variable bit rates. Basically, Double Stimulus Continuous Quality Scale (DSCQS) method was used where the pairs of sequences were presented to the viewer, where first one is an original sequence and the other is a processed impaired sequence. Then the final test video was formed by randomly pairing original and degraded video sequence and the observers were asked to evaluate the quality of overall impaired sequences using a five-point grading scale, from 1 to 5, according to perceived quality. Number of viewers had to be at least 20 for each test run to be able to obtain statistically meaningful results, and the test run was kept to maximum of 30 minutes in order to maintain viewer attention.

2.3 Feature Space Reduction

The feature space of objective measures is reduced through the selection of the feature descriptors with dominant influence on accurate prediction of the subjective assessments. By reducing the number of features two important effects are achieved: (i) reduction in computational complexity, which aids real-time assessment procedure and (ii) reduction of number of examples necessary for model training. The selection of dominant features for perceptual video quality assessment is based on feature vectors statistical analysis. Simultaneously two approaches were used: (i) selection of important features using statistical analysis of non-linear mean opinion scores (MOS) prediction models formed as NNs with one measure at an input and (ii) principal component analysis (PCA) and independent component analysis (ICA) [14] models for feature space reduction.

Generally, effectiveness of objective measures for video quality assessment is reflected through their correlation with subjective assessments, i.e. mean opinion scores (MOS). Identification of potential non-linear dependency between objective measures and subjective assessments is done by forming Multilayer Perceptrons (MLPs) neural network [15] for each feature. MLP has a task to do non-linear mapping of each feature at an input to predicted MOS at an output using a number of statistical measures such as: root mean square prediction error (RMSE), spearman rank-order correlation coefficient (SCC), Pearson linear correlation coefficient (PCC), number of outliers, Euclidean distance between outputs of each NN-based predictor. The final selection of features which will be used to form non-linear prediction model is obtained through the analysis of initial set of features obtained by combining results from PCA, ICA approach and statistical analysis single feature NN-based predictor.

2.4 Feature Space Clustering

Once we selected the features which are the best descriptors of ringing, blocking and blurring artifacts, the measurements were performed on the most representative test sequences with various contents and dynamics. The clusters were formed by partitioning the feature space...
which was defined by selected features over a set of video frames from available set of test video sequences. The clusters were formed in a feature space using k-means and fuzzy c-means classification techniques which roughly divide space into content dependant regions. The frames from test sequences can belong to different clusters to certain degree. Although, in system training phase each frame can appear only in a single cluster, and that will be the one to which it belongs to the highest degree.

2.5 Non-linear Mapping

The final step in VQA system design is to derive a non-linear model for the selected objective measures by mapping them to the subjective assessments. Non-linear mapping between selected features and averaged MOSs of video sequences under consideration reflects unknown perceptual sense of a viewer. For this purpose the multilayer perceptron NN variable number of nodes in a hidden layer is used.

3 REAL-TIME IMPLEMENTATION

Calculation of the objective measures is computationally intensive tasks and can not be executed in a real-time on a common x86 platforms. A multi-core microprocessor technology is capable of executing VQA methods in real-time, even for HD content [11].

The non-linear expert model is selected based on where measured features of input unknown video lie in clustered feature space (see Fig. 3). For example, for a current frame under consideration its location in a feature space is determined and also to which degree it belongs to a certain cluster, based on reciprocal value of Euclidean distance between position of the current frame and the cluster centroid. To which degree the frame belongs to the cluster determines to what extent the output of the associated neural network will contribute to the final output. Hence, the final output which represents the overall video quality represents the weighted sum of all NNs nodes.

The quality scores are displayed as in Fig. 4 with semaphores indicating the state of each artifact for a current frame. For every artifact type system outputs one normalized value in range from 0 to 1, where 0 means that no artifacts are present in input sequence and value 1 means that great amount of appropriate artifacts are detected. The system also communicates degree of artifacts presents through semaphores for each artifact. Measuring is field/frame based, meaning that no temporal processing is performed outside of the current field/frame.
4. RESULTS

For every input frame three normalized values arrive for each representative artifact. Fig. 5 shows time graph for all three measures received one sample input. The measured quality is highly correlated with the observed subjective quality.

5. CONCLUSIONS AND FUTURE WORK

This paper describes the concept of a no-reference video quality assessment system and its real-time implementation on the multi-processor platform. The developed system provides real-time quality monitoring of live high-definition video content at a remote site. It is able to process the large amount of data required for live SD and HD video thanks to the parallel features of the multi-core architecture and optimization techniques applied. The developed system can be easily integrated with existing broadcast environments as it supports several video interfaces like HDMI, SDI and CVBS. Presented method provides a novel framework for design and implementation of artifact adaptive VQA system. This approach will be used as a framework to test more measures for different artifacts and extended to include more content specific video material.

6. REFERENCES