A Review of Low Complex Blind Video Quality Predictor based on Support Vector Machines

Amitesh Kumar Singam¹, Venkat Raj Reddy Pashike²

Abstract: Objective Video Quality Assessment plays a vital role in visual processing systems and especially in the mobile communication field; some of the video applications boosted the interest of robust methods for quality assessment. Out of all existing methods for Video Quality Analysis, No-Reference (NR) Video Quality Assessment is the one which is most needed in situations where the handiness of reference video is not available. Our challenge lies in formulating and melding effective features into one model based on human visualizing characteristics. Our research work explores the tradeoffs between quality prediction and complexity of a system. Therefore, we implemented support vector regression algorithm as NR-based Video Quality Metric (VQM) for quality estimation with simplified input features. The features are obtained from extraction of H.264 bitstream data at the decoder side of the network. Our metric predicted with Pearson correlation coefficient of 0.99 for SSIM, 0.98 for PEVQ, 0.96 for subjective score and 0.94 for PSNR metric. Therefore, in terms of prediction accuracy, the proposed model has good correlation with all deployed metrics and the obtained results demonstrate the robustness of our approach. In our research work, the proposed metric has a good correlation with subjective scores which concludes that proposed metric can be employed for real time use, since subjective scores are considered as true or standard values of video quality.

Keywords: VQM, NRVQM, SVM, PCA, LS SVM, MOS, MSE

1. Introduction

With the introduction of 3G technology, the usage of consumer video applications in mobile devices has increased to a large extent. Due to the extremely large competition between service providers and also between the application developers to provide better quality than each other, advanced methods to assess the video quality is in great demand now. Applications like video chatting and live video streaming in mobile or hand-held multimedia devices boosted the interest in no-reference objective video quality assessment where the availability of reference video is very unlikely. Due to the huge demand in this area, it is necessary to provide a required level of customer satisfaction given by the perceived video streaming quality. So, No-reference video quality assessment takes major role for these conditions.

Most of proposed VQA approaches are applied to luminance layer but few are applied to both luminance and chrominance layers in every respective frame and combined by means of gray scale quality metrics. [3] proposed a full reference VQA method based on CMSVD algorithm, that treats chrominance and luminance layers of frame as whole and considers video features and characteristics of HVS. The proposed method was tested on VQEG phase I full reference testing data and obtained perceived video quality with good correlation.

[2] proposed No Reference objective video quality assessment method based on MSE estimation in order to predict video quality degradation because of various pattern packet losses after the transmission of data in network. The proposed algorithm predicts initial error by motion information of data received. After transmission of data in the network initial error occurs due to packet loss in each frame. The proposed method shows high Pearson correlation to match actual MSE and obtain higher correlation than conventional method DMOS and PSNR.

2. Kernel Based Learning Concept

Generally, kernel-based learning methods are classified into supervised and unsupervised learning algorithms. Kernel method solves any problem by mapping the input data set into high dimensional feature space via linear or nonlinear mapping or kernel trick. These algorithms are used for regression, classification and other jobs

2.1 Support Vector Machines

Support Vector Machine was invented by Vladimir VaDnik [1] and it is a supervised learning method that analyzes and recognizes patterns, and it is used for classification and regression analysis. SVM is a powerful learning Based algorithm, its formulation is based on Structural Risk Minimization principle which has advantage over Empirical Risk Minimization principle-based learning algorithms like traditional Neural Networks (NN). Over fitting problem of NN has been solved in SVM. It works by mapping of nonlinear input data to high dimensional induce space employed by nonlinear mapping which leads to linear regression and classification which performs better than conventional artificial neural network.

2.2 Least Square Support Vector Machines

LS-SVM is variant of SVM which uses the least-square linear system instead of quadratic programming method for estimating the function and which leads to improvement of generalization performance then standard SVM. Least Square Support Vector Machine is implemented for improving the accuracy of water quality retrieval, it is suitable for the small-sample fitting. It was Proposed LSSVM was used to monitor concentration of suspended matter. To achieve it, the Radial Basic Function was chosen as the kernel function of the retrieval model. The grid search, k-cross validation method is used for selecting and optimizing the parameters. Simulation results shows that proposed method obtained good performance. At the same time, the complexity of sys-
tem was reduced and speed of modeling was rapidly increased into the final published version.

3. Leave One Out Cross Validation Technique

Cross-validation is a re-sampling strategy used to validate the performance of our proposed model by random subsampling of the available data. In other words, the original data is subdivided randomly into k folds. In each round of CV (k-1) folds are used for training and the remaining one-fold is used for validation, this procedure is repeated for k times with each of the k folds should use exactly once for the model validation. Performance estimation of our proposed model is determined by average of results obtained in k folds for k rounds [4].

Leave one out cross validation is a special case of K-fold cross validation. It uses a single instance from the original input data of video sequences as the validation data, and the remaining instances as the training data. Each instance or data point of input is used once as the validation data.

In LSSVM model, we selected Radial Basis Kernel (RBF) function for realization of implicit mapping of input data into higher dimensional feature kernel space. Since RBF provides better training and testing errors. We used an optimization algorithm for tuning the hyper parameters sigma and gamma with respect to good performance measure. A grid search method was employed, which performs an exhaustive search through a subset of parameter space in machine learning algorithm for solving the problem of selection of model by finding optimal parameters. This algorithm is guided by performance metric (MSE) which is measured by leave one out cross validation in training.

4. Curse of dimensionality

Generally, when we are dealing with high dimensional data, addition of more features will effect the system or model performance and increase complexity of system. Therefore in order to overcome curse of dimensionality factors such as efficiency, classification performance and ease of modeling are to be considered. Efficiency is further classified into measurement, storage and computational costs.

5. Proposed Idea

Our main idea is to reduce the dimension of inputs with minimal loss of information. Dimensionality reduction is suitable in visualizing data, noticing a compact representation, and minimizes computational load. In addition, reducing the number of dimensions can separate the features with significant data from less significant ones which provides further vision into the nature of the data which may not be discovered otherwise. There are various dimensionality reduction techniques like principle component analysis, singular value decomposition, kernel principle component analysis, factor analysis and hierarchical clustering etc. In our research work, we are using three methods for dimensionality reduction.

1) Principle component analysis (PCA)
2) Based on correlation matrix between parameters as
3) Based on correlation between individual parameter and quality metric

We implemented Principle Component Analysis in Singular Vector Decomposition form for dimensionality reduction by low rank approximation method. PCA alone cannot be used for dimension reduction for very larger dimensional data and it cannot find eigen values for non-square matrix.

All the input features are characterized in this plot by a vector, the contribution of each feature towards two principle components are indicated by the distance and direction of the vector. For instance, the first principal component is illustrated in the plot by the horizontal axis with positive coefficients in right half of the bi-plot. The second principal component is illustrated by the vertical axis in the plot.

Each of the data points were represented in the bi-plot by a point and their positions indicates the score of each data point for the two principal components. For instance, first principal component has lowest scores near the left edge of this bi-plot. All the data points were fitted in unit square by scaling them by using Z-score function, so that locations of all data points can be determined.

6. Subjective Video Quality Assessment

Based on ITU-R Rec.BT.500 [2], we conducted subjective tests, since the results obtained from MOS scores are considered as true values of video quality, we also compared the correlation between MOS scores and all the quality metrics used in our research work. Subjective scores are obtained by evaluation of video quality with involvement of human observers according to his/her perception level of quality. Subjects will evaluate the visual quality of video sequences by grading them in the form of Mean Opinion Score. Therefore average of overall scores results in obtaining subjective measure of video quality. Our subjective experiments were conducted under laboratory viewing environment specified by ITU-R BT.500-12 Standards We selected Single Stimulus Continuous Quality Evaluation (SSCQ) out of Single Stimulus and Stimulus Comparison Quality Evaluation.

Our proposed LSSVM Model is trained and tested with 120x17 input features out of 120 video sequences generated in our thesis, input data contains 17 selected features extracted from bit stream information at decoder side of channel as
mentioned in section 2.2. Objective scores of 120 video sequences for PEVQ, PSNR, SSIM metrics and mean opinion scores obtained from subjective experiment are used as target values for our metric [5]. LSSVM model is trained randomly with 80x17 features with corresponding 80x1 target values. Trained LSSVM model has been tested with remaining 40x17 features randomly. The performance of LSSVM model is measured between true and predicted values for selected quality metrics. The prediction accuracy of our proposed models for generated video sequences was evaluated by Pearson correlation coefficient between true and predicted scores after non-linear regression analysis and regression plots are generated for all the proposed models and for different quality metrics used. All regression plots illustrate predicted values or residuals after regressing with true values Y of various quality metrics. Consistency of proposed model is evaluated by outlier ratio and prediction of monotonicity of proposed model is assessed by Spearman correlation coefficient

Statistical analysis illustrates that our proposed models predicted with high correlation for SSIM which is expected. Since all the 120 video sequences which we generated for our research work are encoded in JM Reference 16.1 based on H.264/AVC software uses Rate distortion optimization algorithm for improving video quality while video compression and distortion measure in JM encoder is correlating with SSIM metric.

Our proposed models also obtained good correlation with Subjective and PEVQ scores. Since our metric is predicting with correlation coefficient of 0.96 for MOS which concludes that our metric has good correlation with Human Perception level because subjective scores are considered as true or standard scores of video quality and [6] PEVQ is only metric which has quite close correlation with subjective values. our metric also obtained better correlation with PSNR. Performance of proposed model was measured by performance metrics.

To obtain the 95 %confidence interval of Pearson correlation coefficient we need to find upper and lower bounds, but Pearson correlation coefficient's sampling distribution is not normal distribution, therefore we adapted Pearson values to Fisher’s Z transformation.

<table>
<thead>
<tr>
<th>Metrics/Stats</th>
<th>Mean square error</th>
<th>Standard deviation</th>
<th>Mean absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEVQ</td>
<td>0.0895</td>
<td>0.27</td>
<td>0.2360</td>
</tr>
<tr>
<td>PSNR</td>
<td>2.700</td>
<td>1.2223</td>
<td>1.3078</td>
</tr>
<tr>
<td>VSSIM</td>
<td>0.0025</td>
<td>0.0439</td>
<td>0.0363</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.0020</td>
<td>0.0424</td>
<td>0.0308</td>
</tr>
<tr>
<td>NTIA</td>
<td>0.0267</td>
<td>0.1612</td>
<td>0.1199</td>
</tr>
</tbody>
</table>

**Figure:** Video Quality Assessment Framework of Our Proposed Mode

7. Conclusion

We implemented support vector regression algorithm as NR-based VQM for quality prediction with simplified input features. The features are obtained from extraction of H.264 bit stream data at the decoder side of the network. After investigating, the strength of all extracted features we found out that all features have good contribution. The proposed method predicted with Pearson correlation coefficient of 0.99 for SSIM, 0.98 for PEVQ, 0.96 for Subjective score and better for PSNR metric. The SSIM metric results are expected, since JM Reference software uses rate distortion optimization algorithm for improving video quality and the video compression and distortion measure in JM encoder is correlating with SSIM metric. We conclude that prediction
accuracy of proposed model has good correlation for all deployed metrics including subjective scores and the obtained results demonstrates robustness of our approach.

References