PERFORMANCE EVALUATION OF MOBILE VIDEO QUALITY ESTIMATORS

Michal Ries
Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, Austria.

USC 15th of January 2008
Overview

• Subjective quality tests

• Results of subjective quality tests

• Video quality estimation

• Audiovisual quality estimation
Scope of the Work

• GOAL: Design low complexity video quality metric for mobile video streaming.

• Define motion and color features suitable for content classification and quality estimation.

• Quality and content estimation within one cut.
  – Temporal video segmentation

• Introduce new statistical method for content classification.
Subjective Quality Tests

• Test methods: (rec. ITU-T P.910)
  – Absolute Category Rating (ACR)
  – Scaling: 5 grade MOS scale. 1: Bad
    5: Excellent

• Test conditions:
  – Terminal: VPA IV UMTS/WLAN
  – Codec: H.264 baseline profile
  – Resolution: SIF (240 x 320)
  – Two rounds with identical conditions
  – 26 persons in training set
  – 10 persons in evaluation set – different sequences
Content Classes

We have defined five most frequent content classes:
Design of Video Quality Estimator

- Video quality estimation is based on character of the sequence.
- The video quality is estimated within two cuts:
  - Scene change detector allows for temporal segmentation of video stream.
- Scene change detection is based on variable threshold algorithm.
Scene Change Detector

- Statistical features are computed over sliding window \([n - 10, n + 10]\), 10 previous and 10 following frames.
- Scene change is detected using a local sequence statistical features of sum of absolute differences (SAD):
  - Empirical mean
    \[
    m_n = \frac{1}{N+1} \sum_{i=n-N}^{n-1} X_i
    \]
  - Standard deviation:
    \[
    \sigma_n = \sqrt{\frac{1}{N} \sum_{i=n-N-1}^{n-1} (X_i - m_n)^2}
    \]
- The thresholding function:
  \[
  T_n = a \cdot m_n + b \cdot \sigma_n
  \]
Scene Change Detector

- Dynamic video without cuts
- Video with rapid scene changes
Content Based Video Quality Estimation - Highlights

• Content classification
  – Motion and color parameters
  – Hypothesis testing

• Design of video quality estimator
  – Metric design
  – Prediction performance

• Conclusions
Design of Video Quality Estimator

- Two step approach:
  - content class estimation between two cuts on sender side,
  - subjective video quality estimation on receiver side.
Content Classification

- Content classification character of the movement and color features.
Motion and Color Parameters

• Color and motion parameters are extracted within one cut and calculated for each frame.

• Content classification is based on following features:

  • Color features:
    – Percentage of green pixels in the frame,
    – Green color is determined by 5 bins from 64 bin color space.
Motion and Color Parameters

- Type and character of movement is defined:
  - Percentage of zero motion vectors (MV) in one frame.
  - Mean size of non-zero MVs in one frame normalized by width of sequence resolution.
  - Percentage of MVs pointing in the dominant direction (the most frequent direction of MV) in the frame.
  - Dominant horizontal direction of non-zero MVs. The directions of MVs are in intervals (−10; 10) or (170; 190).
Hypothesis Testing

• We set up hypotheses for all content classes based on color and motion features.

• Find out the most suitable test for our empirical data sets:
  – Seek for difference between two datasets
  – Non parametric and distribution free

• Kolmogorov-Smirnov (KS) test is the most suitable
  – Max difference between two CDFs or ECDFs

\[ D_n = \max \left[ \text{abs} \left( F_n(x) - F_m(x) \right) \right] \]
Hypothesis Testing - Data Processing

- ECDF of percentage of MVs pointing in the dominant direction.

![Graph showing Model ECDF, Tested resampled ECDF, and Tested ECDF]

- Model ECDF
- Tested resampled ECDF
- Tested ECDF
Metric Design

- Low complexity reference free video quality metric.
- The content is signaled parallel with video stream.
- Metric is based on a zero complexity streaming/codec parameters.
Metric Design

• Reference-free video quality metric does not require knowledge about the original (non-compressed) sequence only content class is signaled.

• One video quality metric model with different coefficients for each content class:

\[
MOS_V = K + A \cdot BR + \frac{B}{BR} + C \cdot FR + \frac{D}{FR}
\]

• Model coefficients vary substantially for the content classes – each class has two zero coefficients.
Full Reference Free Video Quality Estimation

Highlights

• Design motion based video quality metric for most frequently streamed content types

• Define the most relevant motion features based on multivariate statistical analysis.

• Single model estimation approach

• Ensemble based estimation approach
Design of Video Quality Estimator

- One step approach:
  - subjective video quality estimation on receiver side.
  - Quality is estimated with single or with ensemble of models
Motion Parameters

• MV features, BR and FR were investigated:
  – mean size of all MV
  – standard deviation of MV sizes
  – histograms of MV directions
  – variance of MV directions
  – proportion of horizontal movement
  – proportion of dominant MV direction

• Principal Component Analysis (PCA) was carried out to verify further applicability of the investigated parameters.
Motion Parameters

- Type and character of movement is defined:
  
  Z - Percentage of zero motion vectors (MV) in one **cut**.
  
  N - Mean size of non-zero MV in one **cut** normalized over width of sequence resolution (normalized to max length of MV).
  
  U - Percentage of MVs pointing in the dominant direction (the most frequent direction of MV) in one **cut**.
  
  S - Ratio of MV deviation within one shot refers to proportion of standard MV deviation within one shot.
Design of Video Quality Estimator

• Reference-Free video quality metric does not require knowledge about the original.

• Video quality metric model for all content classes:

\[ MOS_v = k + a \cdot BR + c \cdot Z + d \cdot S^e + f \cdot N^2 + g \cdot \ln(U) + h \cdot S \cdot N \]
I’d like to Ask the Audience, …

• Ensemble based systems adopted following process for decision making:
  – Seeking additional opinions before making a decision.
  – Goal in doing so, is to improve our confidence that we are making the right decision, by weighing various opinions.

• Suitable application for ensemble based system is mapping objective data on subjective MOS results.

• Design of ensemble based perceptual quality metric for mobile video streaming services.
Ensemble of Models

• The ensemble generalization error is always smaller than or equal to the expected error of the individual models.

• An ensemble should consist of well trained but diverse models in order to increase the ensemble ambiguity:
  – estimators with significantly different decision boundaries from the rest of the ensemble set.
Ensemble of Models

- Cross validation improves generalization property of our ensemble set.
- Cross validation scheme:
  - Our data set is divided in two subsets and the models are trained on the first set.
  - The models are evaluated on the second set, the model with the best performance becomes ensemble member.
  - The data set is divided with light overlapping with previous subsets into two new subsets and the models are trained on the first set.
Ensemble of Models

• We use six estimation models in our ensemble.

• Estimation models:
  – *k-nearest neighbor rule* (kNN) decision rule assigns to an unclassified sample point the classification of the nearest sample point of a set of previous classified points,
  – *artificial neural network* (ANN).
    • 90 neurons in hidden layer
    • learning method is improved resilient propagation with back propagation
Estimator’s Performance

- Prediction performance (with evaluation set) of content based video quality metric:

<table>
<thead>
<tr>
<th>Content class</th>
<th>Content based</th>
<th>Motion Based</th>
<th>Ensemble based</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>0.9277</td>
<td>0.768043</td>
<td>0.9292</td>
</tr>
<tr>
<td>Soccer</td>
<td>0.9018</td>
<td>0.971771</td>
<td>0.9685</td>
</tr>
<tr>
<td>Cartoon</td>
<td>0.7559</td>
<td>0.861316</td>
<td>0.7697</td>
</tr>
<tr>
<td>Panorama</td>
<td>0.9030</td>
<td>0.799703</td>
<td>0.9068</td>
</tr>
<tr>
<td>“Rest”</td>
<td>0.9307</td>
<td>0.93667</td>
<td>0.9733</td>
</tr>
</tbody>
</table>

Pearson correlation factor:

- Ensemble based: 0.8387
- Motion based: 0.8148
- Content based: 0.8554
Conclusions

• Content estimation based on KS-test is robust and an easily extendable method to new content classes and parameters.

• Content estimation on sender side allows for low complexity metrics on user/receiver side.

• Motion based video quality estimation allows for full reference free estimation.

• All proposed methods are less complex than the ANSI metric.
Thank you for your attention