Content Based Video Quality Estimation for H.264/AVC Video Streaming

Michal Ries, Catalina Crespi, Olivia Nemethova and Markus Rupp

Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology
Gusshausstasse. 25, A-1040 Vienna, Austria
(mries, ccrespi, onemeth, mrupp)@nt.tuwien.ac.at

Abstract— **The scope of this work is the estimation of video quality for low resolution video sequences typical in (mobile) video streaming applications. Since the video quality experienced by users depends considerably on the the spatial (edges, colors, ...) and temporal (movement speed, direction, ...) features of the video sequence, this paper presents a two-step approach to quality estimation. Firstly, shots between two scene changes are analyzed and their content class is found. Secondly, based on the content class, frame rate and bitrate, an estimation of quality is carried out. In this paper, the design of the content classifier as well as an appropriate choice of the content classes and their characteristics is discussed. Moreover, the design of quality metric is presented, based on the mean opinion score obtained by a survey. The performance of the proposed method is evaluated and compared to several common methods. The results show that the proposed approach, provides powerful means of estimating the video quality experienced by users for low resolution video streaming services.**

I. Introduction

For provisioning of streaming services it is essential to provide a required level of customer satisfaction, given by the perceived video stream quality. It is therefore important to choose the compression parameters as well as the network settings so that they maximize the end-user quality. Thanks to its significant video compression gain the newest video coding standard H.264/AVC allows for providing video streaming for low bit and frame rates while preserving the perceptual quality. This is especially suitable for video applications in 3G wireless networks.

Mobile video streaming is characterized by low resolutions, and low bit-rates. The commonly used resolutions are *Quarter Common Intermediate Format* (QCIF,176x144 pixels) for cell phones, *Common Intermediate Format* (CIF, 352x288 pixels) and *Standard Interchange Format* (SIF, 320x240 pixels) for data-cards and palmtops (PDA). The mandatory codec for UMTS streaming applications is H.263 but the 3GPP release 6 [1] already supports a baseline profile of the H.264/AVC codec [2]. The appropriate encoder settings for UMTS (Universal Mobile Telecommunications System) streaming services differ for different streaming content types [3], [4], [5], [6], [7] and streaming application (resolution, codec). In UMTS bearers with 64-384 kbit/s are used for multimedia (audio and video) streaming. Mobile terminals have limited complexity and power, so the decoding of higher rate videos becomes quite a challenging task. It can be assumed that the maximum supported video bit-rates for QCIF resolution are 105 kbit/s and for CIF and SIF resolutions are 200 kbit/s.

In the last years, several objective metrics for perceptual video quality estimation were proposed. The proposed metrics can be subdivided into two main groups: human vision model based video metrics [8], [9], [10], [11] and metrics based only on the objective video parameters [12], [13], [14], [15]. The complexity of these methods is quite high and significant computational power is necessary to calculate them. These metrics are designed for broadband broadcasting video services and do not consider mobile video streaming scenarios. Moreover, we are looking at the measures that do not need the original (non-compressed) sequence for the estimation of quality, because this reduces the complexity and at the same time broadens the possibilities of the quality prediction deployment. Hence, we are looking for an objective measure of the video quality simple enough to be calculated in real-time on transmitter side. In order to keep low complexity of video quality estimation it is necessary to estimate the video content character due to its content dependence of subjective video quality [3], [4], [6], [7]. Goals of our research are to recognize the most significant content types, estimate the video quality of mobile video streaming at the user-level (perceptual quality of service) and to find most suitable codec settings for these frequent content types.

The paper is organized as follows: In Section 2 we describe the classification and recognition of the most significant content types. In Section 3 the sequences selected for evaluation are described as well as the setup of survey, we performed to obtain the MOS values. The results are further processed in Section 4, where focus is given on the video quality estimation. Section 5 contains conclusions and some final remarks.

II. Content features extraction

The human visual perception of video content is determined by the character of the observed sequence. It is necessary to distinguish different content characters/classes because they strongly influence the subjective quality. The character of a sequence can by described by the amount of the edges (spatial information) in the individual frames and by the type and direction of movement (temporal information). The data rate of the video sequence is shared by the number of frames per second. Higher frame rates result in a lower amount of spatial information in individual frames and possibly in some compression artifacts. Thus, taking the data rate as an objective parameter, we can look either on the spatial information or on the temporal information. In the literature the focus is given mainly on the spatial information [14], [15].Such approaches come mainly from the quality estimation of still images [16], [17]. However, especially in small resolutions and after applying compression, not only speed of movement (influencing at most the compression rate) but also the type of the movement plays an important role in the user perception. Therefore, in this work we focus on the motion features of the video sequences that determine the perceived quality.

1) Scene change detector: A video stream can consist of more than one different scene with different content, spatial and temporal information (i.e. a typical video sequence of news consists of some shots of the moderator and diverse reportage shots of the described events). Since each shot of a sequence can have a different content character, splitting a video into its basic temporal units - shots - is the initial step in the process of video content classification due to content of shots variation within one sequence. A shot is a series of video frames taken by one camera (e.g. zooming in or out an object, panning along a landscape).Two consecutive shots are separated by a shot boundary, which can be abrupt or gradual. While an abrupt shot boundary (cut) is generated by simply attaching one shot to another without modifying them, a gradual shot boundary is the result of applying an editing effect to merge two shots.

The most suitable low-complexity method for our purpose is the scene change detection based on a dynamic threshold [18]. The method has been tuned up for our purpose: the coefficients of the thresholding function were modified and 10 upcoming frames were additionally taken into account. The scene change detector works with precision and recall [18] higher than 97%. Such accuracy is more than satisfying for the purpose of content classification.

2) Motion vectors: The block from the current frame for which a matching block is sought, is known as the *target block*. The relative difference in the locations between the matching block and the target block is known as the *motion vector (MV)*. If the matching block is found at the same location as the target block then the difference is zero, and the motion vector is known as *zero vector*.

The difference between target and matching block increases (approximately linearly) with the size of the blocks and smaller blocks better describe the actual motion in the frame. On the other hand an increase of the objective accuracy does not always imply a better performance. We have observed that, if the blocks are selected too small, the resulting MVs do not reflect anymore the motion as it is perceived by a viewer. Due to the unavoidable presence of noise in video sequences, and the characteristics of the human visual system, it happens that movement is detected although a human observer does not see it. Such behavior is not suitable for our purpose. After several trials with videos of different character, we found a block size of 8×8 pixels to be a good trade-off for QVGA resolution sequences. The 320×240 pixels are divided into 30×40 blocks, which gives a total number of 1200 MVs per frame.

The second part of the process, and the most time and resource consuming one, is block matching. Each block in the current frame is compared to a certain search region in the past frame in order to find a matching block. This operation is performed only on the luminance component of the frame. A matching criterion has to be used to quantify the similarity between the target block and the candidate blocks. Because of its simplicity and good performance, we decided to use the sum of absolute differences (SAD), computed as the pixel wise sum of the absolute differences between the two blocks being compared:

$$\mathrm{SAD}_{n,m} = \sum_{i=1}^{N} \sum_{j=1}^{M} |B_n(i,j) - B_m(i,j)| \qquad (1)$$

where B_n and B_m are the two blocks of size $N \times M$, and i and j denote pixel coordinates. If more than one SAD minimum is detected, priority is given to the matching block the position of which is most similar to that of the target block, or equivalently, to the MV of smallest size.

3) Extraction of sequence motion and color parameters: Once we obtained MVs, the information about the motion (motion features) in the sequence has to be extracted. The static or dynamic character of a sequence is one of the main causes for the differences in perceived quality. We intended to perform a classification not only in terms of ”static sequences” and ”dynamic sequences”, but also to investigate this aspect more in depth and determine typical levels of quantity of movement for every main content class. The overall amount of movement, or equivalently, the lack of movement in a frame, can be easily estimated from the proportion of blocks with zero vectors, that is, blocks that do not move from one frame to the other. Therefore, the average proportion of static blocks in a sequence of frames is very useful when it comes to distinguishing contents with typical different ”levels” of overall movement.

The length of the MV indicates how far the block has moved from one frame to the next, and its angle tells us in which direction this movement occurred. Therefore, the mean size of the MVs in a frame or sequence of frames is an indicator of how fast the overall movement happens. On the other hand, knowing exactly in which direction the movement is taking place seems useless (redundant) for our purpose. Moreover detecting a main direction of movement, that correspons to big proportion of MVs pointing in the same direction, is a valuable information. Thus, it can be assumed that the analysis of the distribution of sizes and angles of the MVs can give substantial information about the character of the motion in the sequence. A set of statistical calculations on the MV was implemented in order to study their level of significance and find out which features can be used to identify perceptual content types.

Finally the content classification is based on the following statistical and resolution independent features of MVs within one shot (over all the frames of the analyzed sequence):

- **Zero MV ratio N:**
  Percentage of zero MVs in a frame. It is the proportion of the frame that does not change at all (or changes very slightly) between two consecutive frames. It usually corresponds to the background if the camera is static within one shot.
- **Mean MV size n:**
  Proportion of mean size of the non-zero MVs within one frame normalized to the screen width, expressed in percentage. This parameter determines the amount of the global motion.

- **Uniformity of movement** d**:**
  Percentage of MVs pointing in the dominant direction (the most frequent direction of MVs) in the frame. For this purpose, the granularity of the direction is 10 degrees.
- **Horizontalness of movement** h**:**
  We define horizontalness as the percentage of MVs pointing in horizontal direction. Horizontal MVs are from intervals $\langle -10; 10 \rangle$ or $\langle 170; 190 \rangle$ degrees.

In order to increase the accuracy of the content classifier, color features were considered. Color histograms provide additional information about the spatial sequence character because in different types of contents, the density and magnitude of colors differ as well. Soccer sequences for example contain a lot of varying green colors while cartoon sequences exhibit discrete saturated colors. This characteristic has important consequences to the compression and transmission artifacts. Therefore, we also use the following parameter:

- **Greenness** g :
  We define greenness as percentage of green pixels in a frame. For this purpose the RGB color space was down sampled to two bits per color component resulting in 64 colors. Five colors out of the 64 colors cover all variation of the green color.

III. Content Classification

In this section five content classes are identified and presented based on the content features defined in Section II. Furthermore decision algorithms for automatic content classification are designed and evaluated.

A. Content classes

For the mobile video streaming content classification we define the five most frequent content classes with different impact on the user perception:

Fig. 1. Snapshots of typical content classes

1) Content class (news): The content class number one includes sequences with a small moving region of interest (face) on a static background. The movement in the region of interests (ROI) is mainly determined only by eyes, mouth and face movements. The ROI covers up to approximately 15% of the screen surface.

2) Content class (soccer): This content class contains wide angle camera sequences with uniform camera movement (panning). The camera is tracking a small rapid moving object (ball) on the uniformly colored (typically green) background.

3) Content class (cartoon): In this content class object motion is dominant, the background is usually static. The global motion is almost not present due to its artificial origin of the movies (no camera). The movement object has no natural character.

4) Content class (panorama): Global motion sequences taken with a wide angle panning camera. The camera movement is uniform and in a single direction.

5) Content class (rest): The content class contains a lot of global and local motion or fast scene changes. Scenes shorter than three seconds are also associated to this content class. The content class covers scenes which do not fit any of the previous four classes.

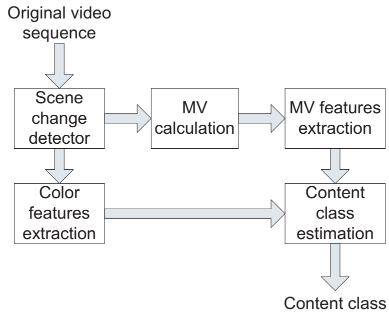

Fig. 2. Content classifier design

B. Hypothesis testing and content classification

The content classification is based on above defined parameters. Due to extensive set of objective parameters, the statistical method were used for data analyzes and content classification. This assumption excludes content classifying based on threshold which is a limited and not accurate method for evaluating larger data sets.

We use a statistical method based on hypotheses testing. Each of the described content classes is determined by unique statistical features of motion and color parameters (see Figure 3). Due to their unique statistical features of well defined content classes it is not necessary to perform M-ary hypothesis testing and it is sufficient to formulate a null hypothesis (H0) for each content class based on these statistical features separately. The obtained empirical cumulative distribution functions (ECDF) from the typical set of sequences for each content class show substantial mutual differences (see Figure 3). From the next investigation it results that it is very difficult to determine single parametric distribution model representation from obtained model ECDF. For this purpose we were looking for hypotheses testing method which allows for defining non parametric, distribution free H0 hypothesis.

For our hypothesis evaluation a method is needed capable of working with empirical (sample) distributions. For this purpose the most suitable is the non parametric and distribution free: the Kolmogorov-Smirnov (KS) test [19]. The KS test is used to determine whether two underlying probability distributions differ, or whether an underlying probability distribution differs from a hypothesized distribution, in either case based on finite samples. The two-sample KS test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples.

From the typical set of sequences for each content class the ECDFs are obtained. The model ECDFs were derived from a set of 142 typical sequences. Each content class is described with five model ECDFs (zero MV ratio, mean MV size, uniformity of movement, horizontalness of movement, greenness), which correspond to their H0 hypothesis, respectively. Furthermore, it is necessary to find the maximal deviation ($D_{cc\ max}$) within one content class for all parameters (for each model ECDF). If the $F_n(x)$ is the model ECDF and $F(x)$ is the ECDF of the investigated sequence. D_n; is the maximal difference between $F_n(x)$ and $F(x)$:

$$D_n = \max_x \|F_n(x) - F(x)\| \,. \tag{2}$$

The content class estimation is based on a binary hypothesis test within first four content classes. With the KS test the ECDFs of the investigated sequence and all model ECDFs of the first four content classes are compared. The KS test compares five ECDF (of defined MV or color parameters) of defined content classes specified by the H0 hypothesis with all five ECDFs of the investigated content. If the obtained D_n for all parameters and of (the first four) content classes is smaller than $D_{cc\ max}$ for each parameter than the investigated sequence matches this content class.

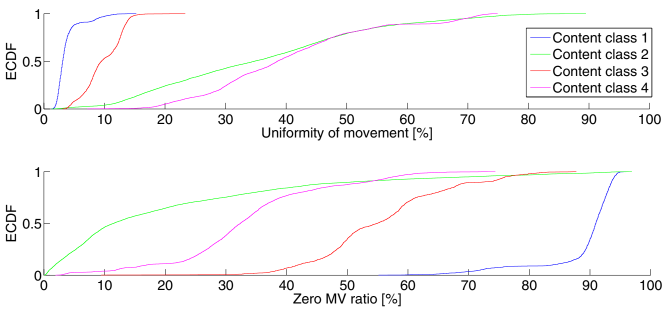

Fig. 3. Model ECDF of zero MV ratio and uniformity of movement

If the ECDFs of the investigated sequence have not a fit with any of first four content classes, the content classifier (Figure 2) decides for the rest content class number five. The classifier estimates the content on transmitter side from the original sequence.

The performance of the content classifier was evaluated with two parameters. **False detection** reflects the ratio of improper detection of a content class, in the case when investigated sequences belong to any **other** content class. **Good match** reflects the ratio of successful classification of investigated sequences, when investigated sequences belong to any of the first four classes. Note, in our sequences we had almost only cuts and no gradual changes. The scene change detector was sensitive on gradual shot boundaries (dissolve, fades or wipes). The achieved precision of the content classifier is shown in Table I, what is a satisfying result for further quality estimation.

Content class	False detection [%]	Good match [%]
1	0	97
2	0	100
3	5,6	92
4	0	100
Num. of sequenc.	786	98

TABLE I

THE EVALUATION RESULTS OF CONTENT CLASSIFIER

IV. THE TEST SETUP FOR VIDEO QUALITY EVALUATION

For the tests we selected two sets of five video sequences each having ten-second duration and SIF resolution. Screenshots of these sequences are depicted in Figures 1. All sequences were encoded with an H.264 baseline profile 1b. For subjective quality testing we used frame and bit rate combinations shown in Table II. In total they were 36 combinations.

FR [fps]/BR [kbit/s]	24	50	56
5	Ne, Ca	Vi	Ne, Ca
7.5	Ne, Ca		Ne, Ca
10	Ne, Ca		Ne, Ca
15	Ne		Ne

FR [fps]/BR [kbit/s]	60	70	80	105
5				Ne
7.5	Vi	Vi		Ne, So, Vi
10		Vi	Vi	Ne, So, Vi
15			Vi	Ne, So, Vi

TABLE II

TESTED COMBINATIONS OF FRAME RATES AND BIT RATES. ABREVIATION OF SEQUENCE TYPES: CA = CARTOON, NE = NEWS, SO = SOCCER, PA = PANORAMA, VI = VIDEOCLIP

To obtain a MOS (Mean Opinion Score), we worked with 36 test persons for two different sets of test sequences. The first set was used for metric design and the second for evaluation of the metric performance. The training test set was carried out with 26 test persons and the evaluation test set was carried out with 10 test persons. The training and evaluation tests were collected of different sets of five video sequences. The chosen group ranged different ages (between 20 and 30), gender, education and experience with image processing.

The tests were consistent with the ITU-T Recommendation [20], using the absolute category rating (ACR) method as it better imitates the real world streaming scenario. Thus, the subjects had not the original sequence as a reference, resulting in a higher variance. People evaluated the video quality using a five grade MOS scale (1-bad, 2-poor, 3-fair, 4-good, 5-excellent). According to our experiences with previous psycho-visual experiments [4], [5] the subjective results are slightly different if they are displayed on UMTS handsets or PC monitors. Due to this experience we did not follow in this only case ITU-T Recommendation [20] in this point and in order to emulate real conditions of the UMTS service, all the sequences were displayed on a PDA VPA IV UMTS/WLAN (see Figure IV).

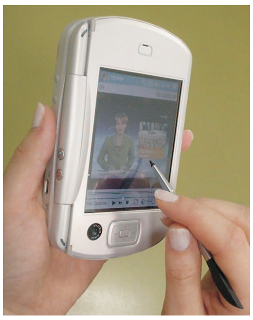

Fig. 4. Test equipment: VPA IV UMTS/WLAN

The viewing distance from the phone was not fixed, but selected by the test person. We have noticed that all subjects were comfortable to take the PDA at a distance of 20-30 cm. At the beginning of the test session, three training sequences were presented to the test persons. Test sequences were presented in an arbitrary order, with the additional condition that the same sequence (even differently degraded) did not appear in succession. Two runs of each test were taken. In order to avoid the learning effect we made a break of half an hour between the first and the second run. In the further processing of our results we have rejected the sequences which were evaluated with an individual standard deviation higher than one. Following this rule, we excluded 12.4% of the test results.

V. Video quality estimation

Our second step was to design a real time video quality estimator at user equipment. The estimation on the receiver side has to be based only on the compressed sequence without the original (uncompressed) sequence and the information about content class is in parallel signalized with the video streaming (see Figure 5), in order to reduce processing complexity as much as possible. Such measurement setup allows for continuous real time video streaming quality measurement on both sides: user and provider. The video quality is estimated after content classification within one cut (see subsection II-.1 and III-A.5).

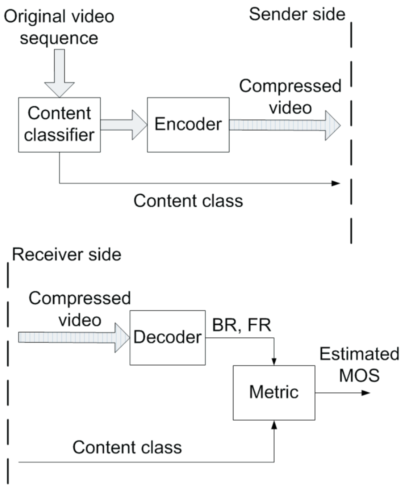

Fig. 5. Content based video quality estimator design

A. Objective data analysis

Due to limited processing power of the user equipments it was necessary to identify low complexity objective parameters. In order to keep the complexity as low as possible the most suitable parameters are already provided frame rate (FR) and bit rate (BR). These parameters are the codec compression settings and signalized during the initiation of the streaming session, requiring no computational complexity for estimation as they are known at both transceiver and receiver. Furthermore, it is necessary to describe the influence of these two parameters on a investigated dataset for each content class separately.
For this purpose, we used a well known multivariate statistical method, the Principal Component Analysis (PCA) [22]. The PCA was carried out to verify further applicability of the objective parameters BR and FR for the metric design. The PCA was performed for all content classes separately. In our case the first two components proved to be sufficient for an adequate modeling of the variance of the data (see Table III)

Sequence	Variab. of PC1 [%]	Variab.of PC2 [%]
Content class 1	61.7	23.1
Content class 2	51.8	32.9
Content class 3	54.8	30.4
Content class 4	53.1	42.7
Content class 5	63.5	28.2

TABLE III
The total variability of the first two components for all content classes.

The PCA results (see Figure 6) show sufficient influence of

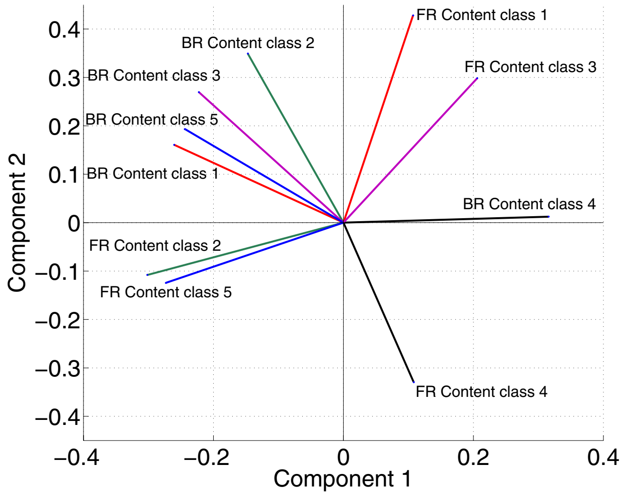

Fig. 6. Visualization of PCA results for all content classes.

B. Subjective quality metric design and evaluation

The proposed low complexity metric is based on two objective parameters (BR and FR) for each content class (see (3)).

$$\widehat{\mathrm{MOS}} = \mathrm{f}(\mathrm{BR}, \mathrm{FR}, \mathrm{Content\ class}). \qquad (3)$$

We proposed one common model (see (3)) for all content classes. Therefore, the model has linear and hyperbolic elements (see (4)) and the coefficients vary substantially for the content classes. They can even have zero values. On the other hand rather good correlation was achieved with one offset and two non-zero coefficients (see Table IV).

$$\widehat{\mathrm{MOS}} = A + B \cdot \mathrm{BR} + \frac{C}{\mathrm{BR}} + D \cdot \mathrm{FR} + \frac{E}{\mathrm{FR}}. \qquad (4)$$

The metric coefficients were obtained by a linear regression of the proposed model with our training set (MOS values averaged over two runs of all 26 subjective evaluations for particular test sequence). To evaluate the quality of the fit

Coeff.	CC 1	CC 2	CC 3	CC 4	CC 5
A	4.0317	1.3033	4.3118	1.8094	1.0292
B	0	0.0157	0	0.0337	0.0290
C	−44.9873	0	−31.7755	0	0
D	0	0.0828	0.0604	0.0044	0
E	−0.5752	0	0	0	−1.6115

TABLE IV
Coefficients of metric model for all content classes (CC)

Content type	CC 1	CC 2	CC 3	CC 4	CC 5
r	0.9277	0.9018	0.7559	0.9030	0.9307
r'	0.9964	0.8863	0.8409	0.9812	0.9695

TABLE V

METRIC PREDICTION PERFORMANCE BY CORRELATION ON EVALUATION SET

of our proposed metric, we used a Pearson (linear) [21] correlation factor:

$$r = \frac{\mathbf{x}^T\mathbf{y}}{\sqrt{(\mathbf{x}^T\mathbf{x})(\mathbf{y}^T\mathbf{y})}}, \quad (5)$$

and the Spearman rank correlation factor [21]:

$$r' = 1 - \frac{6(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})}{N(N^2-1)}. \quad (6)$$

Here, the vector $\mathbf{x}$ corresponds to the **average** MOS values of the evaluation set (averaged over two runs of all 10 subjective evaluations for particular test sequence) for all tested encoded sequences. Vector $\mathbf{y}$ corresponds to the prediction made by the proposed metric. The dimension of $\mathbf{x}$ and $\mathbf{y}$ refers to N.

The performance of the subjective video quality estimation compared to the subjective quality data is summarized in Table V and shown in Figure 7. Obtained correlations with the evaluation set show very good performance of the proposed metric for all content classes except for content class number three, containing two and three dimensional cartoon movies. This feature increases variability of the MOS results within this content class and reduces the metric fitting performance.

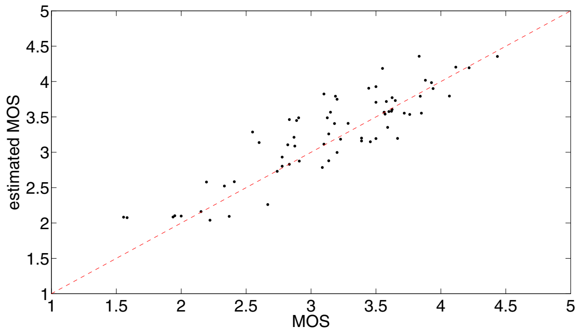

Fig. 7. Estimated vs. subjective MOS results

VI. CONCLUSION

In this paper we proposed content based perceptual quality metrics for the most frequent content types for mobile video streaming services and investigated their performance. Furthermore, our proposed method allows for a continuous quality measurement on both transceiver and receiver side, since it has a low processing complexity.

The automatic content classification enables video quality estimation within one content class. Our proposed automatic content classification recognizes with high accuracy the most frequent content types. Moreover classification based on hypothesis testing is a universal statistical method for content classification, which provides almost unlimited opportunities for the definition of new content classes.

Therefore, it is sufficient to design content dependent low complexity metrics for each defined content type. Our proposed metrics based on basic codec compression setting parameters have minimal complexity on the one hand and excellent prediction performance on the other hand.

Our approach to video quality estimation allows for a reliable method which can be easily extended.

VII. ACKNOWLEDGMENT

The authors would like to thank mobilkom austria AG for supporting their research. The views expressed in this paper are those of the authors and do not necessarily reflect the views within mobilkom austria AG.

REFERENCES

[1] 3GPP TS 26.234 V6.8.0:"End-to-end transparent streaming service; Protocols and codecs".
[2] ITU-T Recommendation H.264 (03/05): "Advanced video coding for generic audiovisual services" — ISO/IEC 14496-10:2005: "Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding".
[3] M. Ries, O. Nemethova, M. Rupp. "Reference-Free Video Quality Metric for Mobile Streaming Applications," Proc. of the DSPCS 05 & WITSP 05, Sunshine Coast, pp. 98-103, Australia, December, 2005.
[4] O. Nemethova, M. Ries, E. Siffel, M. Rupp, "Quality Assessment for H.264 Coded Low-Rate and low-Resolution Video Sequences," Proc. of Conf. on Internet and Inf. Technologies (CIIT), St. Thomas, US Virgin Islands, pp. 136-140, 2004.
[5] M. Ries, R. Puglia, T. Tebaldi, O. Nemethova, M. Rupp, "Audiovisual Estimation for Mobile Streaming Services," Proc. of International Symposium on Wireless Communication Systems, Sept, 2005.
[6] H. Koumaras, A. Kourtis, D. Martakos, "Evaluation of Video Quality Based on Objectively Estimated Metric," Journal of Communications and Networking, Korean Institute of Communications Sciences (KICS), vol. 7, no.3, Sep 2005,
[7] C. John, "Effect of content on perceived video quality," Univ. of Colorado, Interdisciplinary Telecommunications Program: TLEN 5380 Video Technology, 9 Aug. 2006
[8] A. W. Rix, A. Bourret, and M. P. Hollier, "Models of Human Perception," J. of BT Tech., vol. 17, no. 1, pp. 24-34, Jan. 1999.
[9] S. Winkler, F. Dufaux, "Video Quality Evaluation for Mobile Applications," Proc. of SPIE Conference on Visual Communications and Image Processing, Lugano, Switzerland, vol. 5150, pp. 593-603, July 2003.
[10] S. Winkler, Digital Video Quality, JohnWiley & Sons, Chichester, 2005.
[11] E.P. Ong, W. Lin, Z. Lu, S. Yao, X. Yang, F. Moschetti, "Low bit rate quality assessment based on perceptual characteristics," Proc. of Int. Conf. on Image Processing , Vol. 3, pp. 182-192, Sept. 2003.
[12] ANSI T1.801.03, "American National Standard for Telecommunications - Digital transport of one-way video signals. Parameters for objective performance assessment," American National Standars Institute, 2003.
[13] M.H. Pinson, S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Transactions on broadcasting, Vol. 50, Issue: 3, pp. 312-322, Sept, 2004.
[14] T. M. Kusuma, H. J. Zepernick, M. Caldera; "On the Development of a Reduced-Reference Perceptual Image Quality Metric," Proc. of the 2005 Systems Communications (ICW05), pp. 178-184, Montreal, Canada, August, 2005.
[15] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A No-Reference Perceptual Blur Metric," IEEE Int. Conf. on Image Processing, pp. 57-60, Sep. 2002.
[16] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-Reference Perceptual Quality Assessment of JPEG Compressed Images," IEEE Int. Conf. on Image Processing, pp. 477-480, Sep. 2002.
[17] S. Saha and R. Vemuri, "An Analysis on the Effect of Image Features on Lossy Coding Performance," IEEE Signal Processing Letter, vol. 7, no. 5, pp. 104-107, May 2000.
[18] A. Dimou, O. Nemethova, M. Rupp, "Scene Change Detection for H.264 Using Dynamic Threshold Techniques," in Proc. of the 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Service, Smolenice, Slovac Republic. July 2005.
[19] K. Bosch, "Statistik-Taschenbuch," Oldenbourg Wissensch. Vlg, Munich, 1998.
[20] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," September 1999.
[21] VQEG: "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment." 2000, available at http://www.vqeg.org/.
[22] W. J. Krzanowski, "Principles of Multivariate Analysis," Clarendon press, Oxford, 1988.