World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:6, No:6, 2012

The content based Objective Metrics for Video Quality Evaluation

Michal Mardiak and Jaroslav Polec

***Abstract*—**In this paper we proposed comparison of four content based objective metrics with results of subjective tests from 80 video sequences. We also include two objective metrics VQM and SSIM to our comparison to serve as "reference" objective metrics because their pros and cons have already been published. Each of the video sequence was preprocessed by the region recognition algorithm and then the particular objective video quality metric were calculated i.e. mutual information, angular distance, moment of angle and normalized cross-correlation measure. The Pearson coefficient was calculated to express metrics relationship to accuracy of the model and the Spearman rank order correlation coefficient to represent the metrics relationship to monotonicity. The results show that model with the mutual information as objective metric provides best result and it is suitable for evaluating quality of video sequences.

***Keywords*—**Objective quality metrics, mutual information, region recognition, content based metrics.

I. Introduction

THE twentieth century brought many innovations and inventions. One of the most widespread and popular innovations is video in all its variations like cinema, television, videoconference etc. As the number of users asking for transfer of video increases, the quality of video becomes more important. The reliability in the terms of automatically measuring visual quality becomes important in the emerging infrastructure for digital video [1]. This can be essential for evaluation of codec, for ensuring the most efficient compression of sources or utilization of communication bandwidth. Thus the measuring of video quality plays an important role. The most reliable results provide subjective video quality metrics which anticipate more directly the viewer's reactions [2]. However the quality evaluation of the video by subjective methods is expensive and too slow to be used in real-time applications. Therefore the objective methods are starting to be used. The main goal in the objective quality assessment research is to design metric, which can provide sufficient quality evaluation in terms of correlation with the subjective results [3].

In this paper we propose comparison of four objective video quality metrics, i.e. mutual information, angular distance, moment of angle and normalized cross-correlation measure with two reference objective methods (Structural similarity index (SSIM) and Video Quality Metric (VQM) described in [4],[5]) and subjective results. We used set of ten video sequence presented in [6,7]. We preprocessed each of these objective methods (except for two reference methods) with region recognition algorithm, which can simulate human visual system [8] and weighted each region differently. These content-based metrics were correlates with the subjective methods. In the second section we described the objective metrics, which was used for testing. The region recognition and classification is described in section 3. Then the results of our metrics are presented in section 4 and discussion in section 5. The final section concludes our proposed work.

M. Mardiak is with the Institute of Telecommunication, Slovak University of Technology, Bratislava, Slovak Republic (e-mail: mardiak@ktl.elf.stuba.sk).

J Polec is with the Institute of Telecommunication, Slovak University of Technology, Bratislava, Slovak Republic (e-mail: polec@ktl.elf.stuba.sk).

Research described in the paper was financially supported by the Slovak Research Grant Agency (VEGA) under grant No. 1/0602/11.

II. Correlation Based Metrics

A. Normalized cross-correlation measure

The normalized cross-correlation measure is simple correlation based metrics. They quantify similarity between frame from original video sequence $x_k(i,j)$ and corresponding frame $\hat{x}_k(i,j)$ from test video sequence.

Normalized cross-correlation measure can be expressed as:

$$cc = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}x_k(i,j)\hat{x}_k(i,j)}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}x_k(i,j)^2} \quad (1)$$

where $K = 1,...,k$ represents the total number of components. In our case $K = 3$.

B. Moments of angle

Moments of angle metric belongs to the correlation based metrics. It considers the statistics of angles between the pixel vectors of frame from original video sequence and corresponding frame from test sequence [9]. Vectors pointing in the same direction will indicate 'same' color in the frame while vectors with different direction will represent significantly different colors [9]. It means that vector $x(i,j)$ for frame from original video sequence and $\hat{x}(i,j)$ for test sequence can have positive and negative values. Therefore we are constrained to one quadrant of Cartesian space. The maximum difference which can be reach is equal to $\pi/2$ so the normalization coefficient will be $2/\pi$. The moments of

angle metric is defined as [9]:

$$moa = 1 - \left[\frac{2}{\pi} cos^{-1} \frac{\langle x(i,j), \hat{x}(i,j) \rangle}{\|x(i,j)\| \|\hat{x}(i,j)\|} \right] \left[1 - \frac{\|x(i,j) - \hat{x}(i,j)\|}{\sqrt{3 \cdot 255^2}} \right] \quad (2)$$

C. Angular distance

Angular distance metric between vectors $x(i,j)$ for frame from original video sequence and $\hat{x}(i,j)$ for test sequence can express similarity between colors. It quantifies the orientation difference between two color signals so it measures their similarity. Angular distance metric can be expressed as [10]:

$$ad = 1 - \frac{2}{\pi} cos^{-1} \frac{\langle x(i,j), \hat{x}(i,j) \rangle}{\|x(i,j)\| \|\hat{x}(i,j)\|} \quad (3)$$

D. Mutual information

Mutual information has been calculated between two frames; frame from the original sequence and the corresponding frame from the test sequence. Mutual information was calculated separately for each RGB component. Let us assume that the pixel at position (i,j) of the k-th component has value $x_k(i,j) \in \langle 0, G \rangle$. The values of intensity are l and l'. Then the $P^k_{x,\hat{x}}(l'/l)$ represents the number of changes from level of intensity l in the frame from original sequence to level of intensity l' in the frame from test sequence for the k-th according to total number of pixels in the frame. $P^k_x(l)$ stands for the number of level intensity l in the frame from original sequence and $P^k_{\hat{x}}(l')$ represents the number of level intensity l' form test sequence according to total number of pixels in the frame [11].

The mutual information for the k-th component is defined as:

$$I^k_{x,\hat{x}} = \sum_{l=0}^{G} \sum_{l'=0}^{G} P^k_{x,\hat{x}}(l'/l) log_2 \frac{P^k_{x,\hat{x}}(l'/l)}{P^k_x(l) \cdot P^k_{\hat{x}}(l')} \quad (4)$$

And the total mutual information is given by:

$$I = \sum_{k=1}^{K} \sum_{l=0}^{G} \sum_{l'=0}^{G} P^k_{x,\hat{x}}(l'/l) log_2 \frac{P^k_{x,\hat{x}}(l'/l)}{P^k_x(l) \cdot P^k_{\hat{x}}(l')} \quad (5)$$

Where $K = 1,...,k$ represents the total number of components. In our case $K = 3$.

III. Recognition of Region

It is well known fact that human visual system has ability to perceive different part of video sequence frame in different way based on local spatio-temporal content of the sequence [8]. For example intensity of pixels which belong to the edge certainly contains relevant information in terms of perceiving video quality [12].

In our approach we divided each frame into following three different regions: edge region, smooth region and texture region. To determine the correct region we computed gradient magnitude. The algorithm consists of the following steps [13]:

1. First, calculate the gradient magnitudes for frame from original video sequence and frame from test video sequence using Sobel operator separately to obtain two gradient fields with the same size as frame from original video sequence.
2. Compute two thresholds T_1 and T_2 defined as follows:

$$T_1 = 0.12 \cdot g_{max} \quad (6)$$

$$T_2 = 0.06 \cdot g_{max} \quad (7)$$

where g_{max} is the maximum value of gradient magnitude for frame from original video sequence calculated in first step.

3. The value of gradient magnitude of pixel at the position (i,j) in the frame from original video sequence is $g(i,j)$ and for frame from test sequence $\hat{g}(i,j)$. Assign each particular pixel from frame into one of the three regions according to the following rule:
   - Pixel belongs to the edge region if $g(i,j) > T_1$ or $\hat{g}(i,j) > T_2$.
   - If $g(i,j) < T_2$ or $\hat{g}(i,j) \le T_1$ than the pixel is considered as a part of smooth region.
   - Pixel, which does not match any of the above criteria is part of texture region.
4. Apply particular weight for every pixel according to the region where it belongs. The chosen values of weigh for every region are described below.

This recognition of region in video sequence was applied for all of the objective methods (except two reference metrics SSIM and VQM), i.e. mutual information, angular distance, moment of angle and normalized cross-correlation measure.

IV. Results

To prove relevance of results from proposed method, we compared two reference objective metrics and four other objective methods, which are not often used for evaluation of video quality with results from subjective method.

All subjective tests were performed on the set of video sequence from LIVE Video Quality Database [6, 7]. Set of test video sequences consists of following ten video sequences [6, 7]:

- bs – blue sky with some trees in circular movement of camera
- mc – video shows toy train moving horizontally in foreground. In the same time the calendar is moving vertically in the background.
- pa – shows people in the street. They are moving in different direction but the camera does not move.
- pr – man with the umbrella running along the river. Camera is moving in the same direction as the running man.

- rb – still camera shows river floor with some rocks and moving water.
- rh – video shows traffic on the street. Camera is still and cars are moving towards and away from the camera.
- sf – detail of sunflower with a bee.
- sh – guide in a museum who is walking along the wall with coat-of-arms.
- st – shows a railway station, railroads and some persons walking through railroad. Camera is zooming out during the whole video sequence .
- tr – shows tractor plowing on farmland and camera moves horizontally as the tractor moves.

Eight different modifications of video sequence were created from each of the ten test video sequences. To create these modified video sequences, different types of distortions were chosen to represent different visual appearance. Four of these distortions were caused by MPEG-2 compression with the compression rates varied from 700 Kbps to 4 Mbps, depending on the reference sequence [6, 7]. Another four distorted video sequences were created by the H.264 compression with rates varied from 200 Kbps to 5 Mbps [6, 7]. Fps parameter for test sequences was 25 or 50 fps. Eighty video sequences were used for testing in total.

The subjective results were obtained from single stimulus procedure and the rated on continuous scale. The quality of every video was evaluated by 38 viewers but only 28 viewers were considered as valid according to ITU-R BT 500.11 [2]. The value of subjective evaluation of quality use in this paper is mean DMOS values (averaged across all viewers scores).

We chose six different objective methods to compare them with results of subjective evaluation of quality. Four of them belong to those metrics, which are not often used for objective quality measurement for video, i.e. mutual information, angular distance, moment of angle and normalized cross-correlation measure. The two metrics (VQM and SSIM) are commonly used for quality evaluation and VQM was also standardized in 2003 by ANSI organization [14] and in 2004 by ITU-T [15]. SSIM and VQM metrics serve as "reference" objective metrics because their pros and cons have already been published.

To gain ability to assign different importance to particular region we used different weights for different regions. To obtain ideal combination of weights we decided to fix one weight. We fixed value of weight for texture region due to two following facts. First fact is that there is no specific condition for assigning the pixel to texture region (i.e. pixel, which does not belong to edge or smooth region is considered

TABLE I
PEARSON CORRELATION COEFFICIENT FOR EVERY OBJECTIVE METRICS AND VIDEO SEQUENCE

objective metric	*Video sequence*									
	bs	**mc**	**pa**	**pr**	**rb**	**rh**	**sf**	**sh**	**st**	**tr**
SSI M	-0.7520	-0.9505	-0.8467	-0.9749	-0.9724	-0.9717	-0.6796	-0.8932	-0.6951	-0.9792
VQM	0.7511	0.9731	0.8130	0.8844	0.9637	0.9617	0.5959	0.9376	0.7343	0.9393
I	-0.4623	-0.9053	-0.9901	-0.8344	-0.9781	-0.8367	-0.7683	-0.8698	-0.3041	-0.9811
angle	-0.7641	-0.1857	-0.3507	-0.3187	-0.2082	-0.2866	-0.3033	-0.3876	-0.0821	-0.2496
moa	0.7669	0.4619	0.7929	0.4243	0.5056	0.4898	0.5086	0.4705	0.1898	0.4486
cc	0.6921	-0.3290	-0.1775	-0.4992	-0.4918	-0.0854	-0.3854	-0.2091	-0.0593	-0.3225

TABLE II
SPEARMAN CORRELATION COEFFICIENT FOR EVERY OBJECTIVE METRICS AND VIDEO SEQUENCE

objective metric	*Video sequence*									
	bs	**mc**	**pa**	**pr**	**rb**	**rh**	**sf**	**sh**	**st**	**tr**
SSI M	-0.6946	-0.9524	-0.8810	0.4124	-0.8503	-0.9524	-0.7619	-0.8810	-0.8095	-0.9762
VQM	0.7619	0.9524	0.8095	0.4124	0.8503	0.9762	0.6667	0.9048	0.8095	0.9286
I	-0.5509	-0.9286	-0.9762	-0.8571	-0.9701	-0.7619	-0.7619	-0.9048	-0.3810	-0.9762
angle	-0.6946	-0.1429	-0.3810	-0.4524	-0.4192	-0.2857	-0.3571	-0.5714	-0.0238	-0.3333
moa	0.6946	0.4524	0.7381	0.4524	0.5629	0.4524	0.3571	0.5714	0.1190	0.3810
cc	0.5509	-0.2381	-0.1667	-0.1905	-0.2275	-0.1905	-0.5476	-0.3810	-0.0238	-0.5714

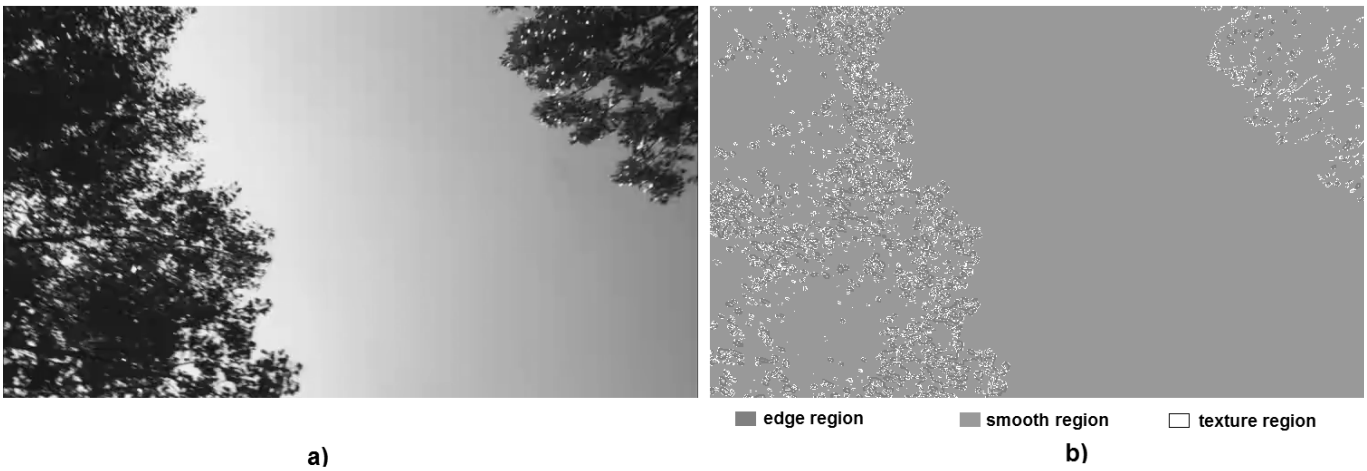

Fig. 3 Ability of Sobel filter to recognize three different region. Figure a) shows the original frame and b) shows weighted mask for bs video. Finding of texture region is not very good in the part of frame where the leafs are.

as texture). The second reason for fixing texture weight is that Sobel filter is not very good for finding the region of texture. Fixing the weight means that we set it equal to one so no additional importance is added to the metrics values belonging to texture region. Following figure shows dependency of weight values for edge and smooth region when the texture weight is set to one:

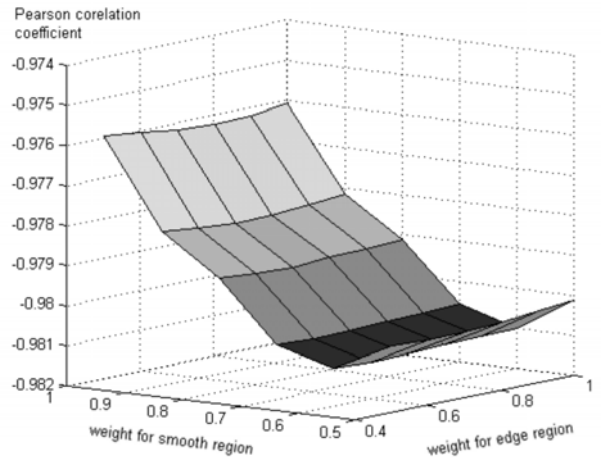

Fig. 1 Correlation between weights for smooth and edge region in case of fixed texture weight (equal to one)

The highest Pearson correlation equal to 0.9811 is achieved when the edge weight is equal to 0.7 and weight of smooth region is equal to 0.6. Therefore, the final combination for texture, smooth and edge region is 1, 0.6 and 0.7.

Results from objective methods were processed according to [15]. The score of particular objective metric was calculated for every frame in every video sequence. Then the mean value for each sequence was used for calculation correlation coefficients. The Pearson linear correlation coefficient was calculated to express metrics relationship to accuracy of the model and the Spearman rank order correlation coefficient to represent the metrics relationship to monotonicity. The final values of both correlation coefficients can be found in Table I and Table II. The relationship between the tested subjective method and DMOS does not need to be linear, as subjective testing can have nonlinear quality rating compression at extremes of the test range [16]. Due to this fact the nonlinear regression was used to data set with the following logistic function [16]:

$$f_{log} = \frac{a}{1 + e^{-b(OM - c)}} \tag{8}$$

where OM is value of objective metric.

Fig. 2 shows objective method, which has the best correlation coefficient with DMOS value for particular video sequence and logistic function.

V. Discussion

The final correlation coefficients were calculated for each of the test sequence and all its distorted modification. As the results show the mutual information with weighted region predicts quality better than other objective methods (including SSIM and VQM) in four out of ten videos (video sequences pa, rb, sf and tr). The highest Pearson correlation -0.99012 achieves mutual information with weighted region in pa video sequence. On the other hand, the significantly worst result comparing to SSIM and VQM are obtain by mutual information on two video sequences (bs and st). This might be caused by utilization of Sobel filter in process of determining the regions. Bs video sequence contains many tree leafs. Pixels on the edge of each leaf are detected as edge region pixels. However, these detected edges can be considered as 'false' edges in terms of quality and should be rather evaluated as part of the texture region as shows on Fig. 3. Same behavior can be observed in st video. It contains many railroads and trolleys with details, which cause higher number of detected edges (with many 'false' edges). The texture region should fit better there.

Moment of angle metric weighted with the same combination as mutual information predict quality of one video sequence better than any other metric (bs video

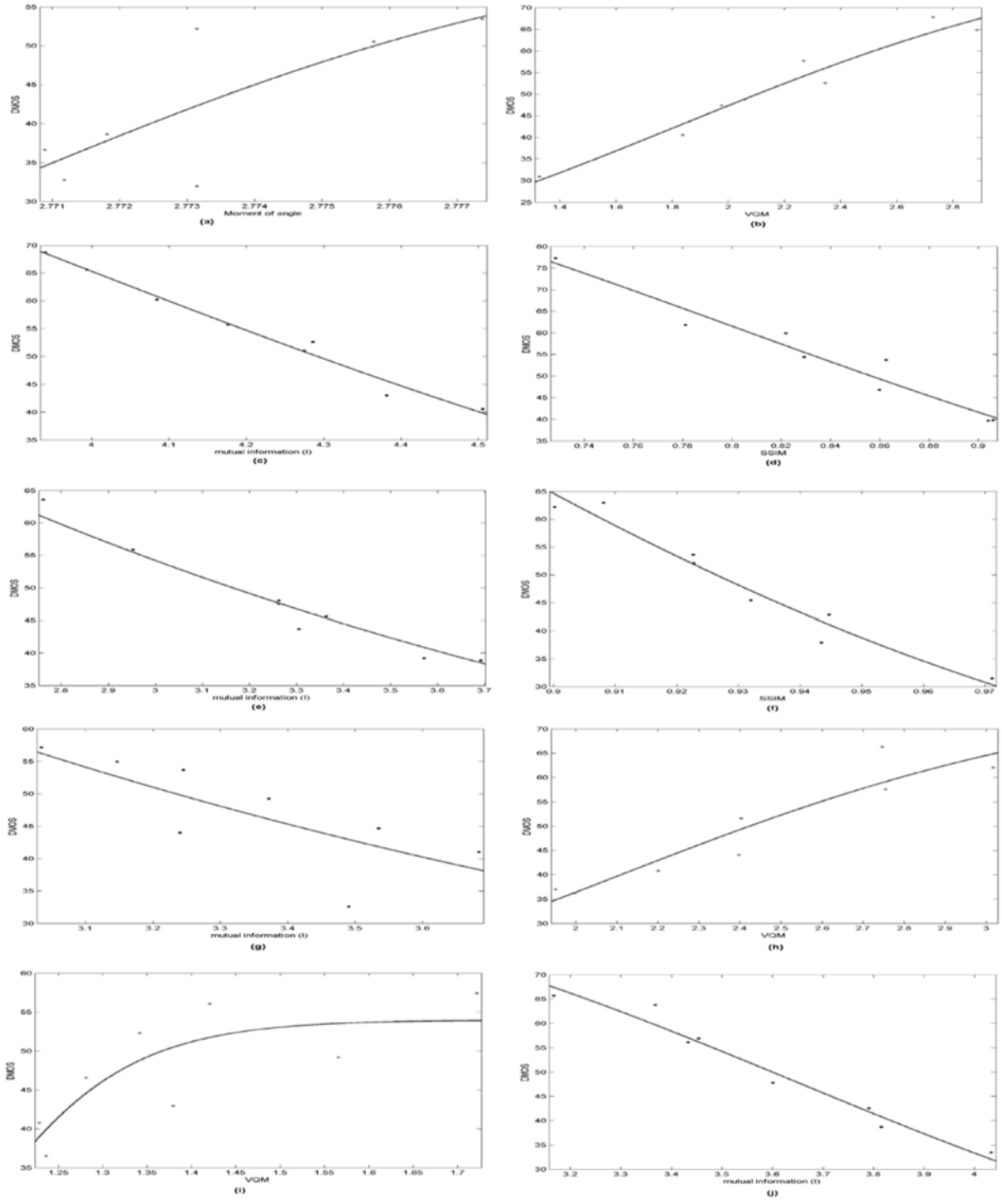

Fig. 2 objective methods which have the best correlation coefficient with DMOS value together with logistic function space: (a)moment of angle for bs video sequence (b)VQM metric for mc video sequence (c)mutual information for pa video sequence (d)SSIM for pr video sequence (e)mutual information for rb video sequence (f)SSIM for rh video sequence (g)mutual information for sf video sequence (h)VQM metric for sh video sequence (i)VQM metric for st video sequence (j)mutual information for tr video sequence

sequence). Moment of angle provides best result despite of the disability of Sobel filter to determine correct texture region in this video sequence. In two other video sequences (pa and sf) moment of angle metric does not provide best results. However the correlation value is comparable to objective metrics, but objective metrics provide better result in comparison with subjective results. In the rest seven video sequences the ability to predict subjective evaluation of

quality is minimal. Overall Pearson correlation coefficient is very small with maximum value around -0.5.

The rest two objective methods, i.e. angular distance and normalized cross-correlation measure do not correlate well with the subjective evaluation of quality. Only in the bs video sequence are values of correlation coefficient for both metrics comparable with reference metrics VQM and SSIM. In the rest of the video sequences is the Pearson correlation coefficient is markedly lower than in the case where quality is evaluated with mutual information, VQM, SSIM or with moment of angle metric.

VI. Conclusion

In this paper we presented content based weighted objective methods and their comparison with two well know objective methods and result from subjective testing. The overall results show that content based weighted mutual information provides best results in term of correlation. Pearson correlation coefficient was highest in four out of all ten video sequences. The VQM and SSIM are considered as well known objective methods in this paper. VQM provides best Pearson correlation coefficient in three out of ten video sequences and SSIM only in video sequences. Pearson correlation coefficient is highest only in one case, where we use moment of angle metric as objective method.

The method proposed in this paper provides best result in case when content based weighted mutual information is used. Therefore, it can be used as objective criterion of quality even if it is pixel based metric.

References

[1] A. B. Watson, J. Hu, and J. F. McGowan, "Dvq: A digital video quality metric based on human vision," *Electronic Imaging*, vol.10, pp.20-29, 2001.

[2] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union – Radiocommunication Sector, Tech. Rep. BT.500-11, 2002.

[3] J. L. Martinez, P. Cuenca, F. Delicado and F. Quiles, "Objective video quality metrics: A performance analysis," in *Proc EUSIPCO* Proc., Florence, 2006.

[4] Z. Wang., A. C Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process*., vol. 13, pp. 1-14, 2004.

[5] F. Xiao, "DCT-based Video Quality Evaluation," MSU Graphics and Media Lab (Video Group), 2000.

[6] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video", *IEEE Trans. on Image Processing*, vol.19, pp.1427-1441, June 2010.

[7] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "A Subjective Study to Evaluate Video Quality Assessment Algorithms", in *SPIE Proceedings Human Vision and Electronic Imaging*, Jan. 2010.

[8] S. Pechard, D. Barba, and P. Le Callet, "Video quality model based on a spatio-temporal features extraction for H.264-coded HDTV sequences," in *Proc. of the Picture Coding Symposium (PCS '07)*, Lisboa, 2007.

[9] I. Avcibaş, B. Sankur and K. Sayood, "Statistical evaluation of image quality measures," *J. of Electronic Imaging*, vol.11, pp. 206-223, 2002.

[10] D. Androutsos, K. N. Plataniotis and A. N. Venetsanopoulos, "Distance Measures for Color Image Retrieval," in *Proc ICIP '98*, Chicago, 1998, pp. 770-774.

[11] Z. Cernekova, "Temporal video segmentation and video summarization," Ph.D. dissertation, Comenius Univ., Bratislava, 2009.

[12] C. Li and A. C. Bovik, "Content-weighted video quality assessment using a three-component image model," *J. Electronic Imaging*, vol. 29, pp. 011003-1-9, 2010.

[13] J. L. Li, G. Chen, and Z. R. Chi, "Image coding quality assessment using fuzzy integrals with a three-component image model," *IEEE Trans. Fuzzy Syst.*, vol. 12, pp. 99–106, 2004.

[14] ANSI T1.801.03, "American National Standard for Telecommunications – Digital transport of one-way video signals – Parameters for objective performance assessment," 2003.

[15] ITU-T, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," Recommendation J.144, 2004.

[16] The Video Quality Experts Group, "Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment, phase II," 2003.