

Objective Video Quality Method Based on Mutual Information and Human Visual System

Michal Mardiak and Jaroslav Polec

Abstract—In this paper we present the objective video quality metric based on mutual information and Human Visual System. The calculation of proposed metric consists of two stages. In the first stage of quality evaluation whole original and test sequence are pre-processed by the Human Visual System. In the second stage we calculate mutual information which has been utilized as the quality evaluation criteria. The mutual information was calculated between the frame from original sequence and the corresponding frame from test sequence. For this testing purpose we choose Foreman video at CIF resolution. To prove reliability of our metric were compared it with some commonly used objective methods for measuring the video quality. The results show that presented objective video quality metric based on mutual information and Human Visual System provides relevant results in comparison with results of other objective methods so it is suitable candidate for measuring the video quality.

Keywords—objective video quality, HVS, mutual information, VQM, SSIM, PSNR.

I. INTRODUCTION

THE recent century became a golden age in the area of technical innovations. One of the most widespread innovations is video in all its variations like cinema, television, videoconference etc. As the popularity of the video grows the requirements for providing video grows too. The reliability in the terms of automatic measuring visual quality becomes important in the emerging infrastructure for digital video [1]. This can be essential for evaluation of codecs, for ensuring the most efficient compression of sources or utilization of communication bandwidth. The most reliable results provide subjective video quality metrics which anticipate more directly the viewers 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 start to be used. The main goal in the objective quality assessment research is to design metric which can provide sufficient quality evaluation regarding to the subjective results [3]. For better approximation of viewers visual perception in the terms of video quality the Human Visual System (HVS) models has been implemented. Various types of HVS have been used in the objective video quality evaluation.

We present objective video quality metric based on mutual information and Human Visual System in this paper. We compare proposed method with several objective methods which have been used for the quality assessment of video sequence

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

M. Mardiak and J. Polec are with the Institute of Telecommunications, Slovak University of Technology, Ilkovicova 3, Bratislava, 812 19, Slovak Republic (e-mails: mardiak@ktl.elf.stuba.sk; polec@ktl.elf.stuba.sk).

i.e. Structural similarity index [4], [5] (SSIM), Peak signal-to-noise ratio [6] (PSNR), Video Quality metric [7] (VQM) and Minkowski-form distance [8] with parameter $r = 3$. In the next section different models of HVS are described and the Section 3 contains the calculation of mutual information. Then the results of our metric are presented and concluded at the end of the paper.

II. HUMAN VISUAL SYSTEM

The purpose of Human Visual System is to simulate human visual perception of the video and consider all its distortions in evaluation of the quality. Image quality measures which utilize HVS should lead in general to a better quality of the reconstructed image [9]. However the HVS is too complex, even using the simplified HVS in the objective measure should lead to a better correlation with the results of subjective methods [10]. There are many models or methods how to simulate human perception of quality. We present some of them here:

A. Low-Pass Gaussian Filter

The HVS is more sensitive in dark areas than in light so the spatial frequency sensitivity of the HVS decreases for high frequencies. The frequency sensitivity should be simulated by low-pass filter [9]. In our paper we choose the following simple low-pass Gaussian filter:

$$HVS1 = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \quad (1)$$

B. Band-Pass Filter

Another way how to simulate HVS can be using the band-pass filters [9], [10]. One of these filters can be expressed by the transfer function in the polar coordinates [10]:

$$H(\rho) = \begin{cases} 0.05e^{\rho^{0.554}} & \text{for } \rho < 7 \\ e^{-9(|\log_{10} \rho - \log_{10} 9|)^{2.3}} & \end{cases} \quad (2)$$

where $\rho = (u^2 + v^2)^{1/2}$. Operator $U\{\bullet\}$ can be used for image processed by the transfer function $H(\rho)$ and afterwards transformed by the inverse discrete cosine transformation (DCT) as follows [10]:

$$U\{x(i, j)\} = DCT^{-1} \left\{ H \left(\sqrt{u^2 + v^2} \right) X_{DCT}(u, v) \right\} \quad (3)$$

where $x(i, j)$ is multispectral pixel vector of image, $X_{DCT}(u, v)$ represents the 2D DCT of the image and DCT^{-1} stands for 2D inverse DCT [10].

C. Laplacian of Gaussian Filters

HVS processing of the image can be simulated also by the Laplacian of Gaussian (LoG) filters [11]. LoG filters can emulate the fact that HVS is more sensitive to the angular resolution and not to the image resolution [12]. We used two LoG filters with the size 7×7 and parameter $\sigma = 1$ and $\sigma = 1.2$.

D. Temporal Filters

The last HVS model presented in this paper is based on the two temporal filters which are also used in the JND metric. These filters are defined by the following impulse response functions [13]:

$$h_1(t) = ae^{-at}u(t) \quad (4)$$

$$h_2(t) = b \left\{ \frac{(bt)^3}{3!} - \frac{(bt)^5}{5!} \right\} e^{-bt}u(t) \quad (5)$$

III. MUTUAL INFORMATION

The mutual information is part of the information theory. The purpose of the metric is to find out interconnection between visual quality of the video and amount of information which is shared between the test video sequence and the reference video sequence. The mutual information is statistical measure of the image fidelity [14].

Average value of mutual information for two random variables X and Y is defined as [15]:

$$I(X; Y) = S(X) + S(Y) - S(X, Y) \quad (6)$$

$$I(X; Y) = S(X) - S(X|Y) = S(Y) - S(Y|X) \quad (7)$$

where $S(X)$ stands for entropy as follows [10]:

$$S(X) = \sum_{X \in X^5} p(X) \log_2 p(X) \quad (8)$$

In the presented method we calculate mutual information for the original and test sequence in RGB color space. Let us assume that pixel of k -th component $x_k(i, j)$ has value $x_k(i, j) \in \langle 0, G \rangle$. The values of intensity level are l and l' . The $P_{x, \hat{x}}^k(l'/l)$ represents the count of changes from the intensity level l in the frame x from original sequence to the intensity level of l' in the corresponding frame \hat{x} from the test sequence for the k -th component regarding to the total amount of the pixel in the frame. Parameters $P_x^k(l)$ and $P_{\hat{x}}^k(l')$ stand for count of the intensity level l in the frame from original sequence and the intensity level of l' in the corresponding frame from the test sequence regarding to the total amount of the pixel in the image [16].

$$I_{x, \hat{x}}^k = \sum_{l=0}^G \sum_{l'=0}^G P_{x, \hat{x}}^k(l, l') \log_2 \frac{P_{x, \hat{x}}^k(l'/l)}{P_x^k(l) \cdot P_{\hat{x}}^k(l')} \quad (9)$$

Total mutual information is defined as:

$$I_{x, \hat{x}} = \sum_{k=1}^K \sum_{l=0}^G \sum_{l'=0}^G P_{x, \hat{x}}^k(l, l') \log_2 \frac{P_{x, \hat{x}}^k(l'/l)}{P_x^k(l) \cdot P_{\hat{x}}^k(l')} \quad (10)$$

IV. RESULTS

It is necessary to choose correct video sequence to ensure that results from presented method cannot be influenced by improper video sequence. In this paper we choose the Foreman video sequencer for the test purposes of our metric. The test sequence has been at CIF resolution (352 x 288 pixels) coded by H.264 codec using the CABAC entropy coding method. Also seven different sizes of the motion compensation blocks were used (16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4), hadamard transformation was performed on the DC coefficient of the video sequence. Inter pixel prediction was also enabled.

The calculation of the presented objective video quality metric based on mutual information and Human Visual System consists of two stages. In the first stage HVS is applied to the test and original video sequence to simulate human perception of quality. We use four different HVS, Gaussian filters, transfer function of band-pass filter in the polar coordinates, two LoG filters and two temporal filters described by their impulse response functions. All of these HVS are described in the section above. In the second stage we calculate mutual information between frame from the original sequence and corresponding frame from the test sequence.

In Figs. 1, 2, 4, 3 the gray curve and the left vertical axis correspond with the reference objective metric (i.e. SSIM, VQM, PSNR, Minkowski-form distance). The right vertical axis and the black curve correspond with the mutual information after applying the particular HVS.

Figure 1 shows comparison between SSIM metric and the mutual information in the case when each of the mentioned HVS is used to simulate property of human visual perception. From the beginning of sequence the mutual information evaluation of quality is slightly increasing and decreasing. Quality oscillation are caused by the movement of the foreman's head in the video sequence. However by applying the Gaussian filter the differences between peaks and bottoms become smaller so the run of mutual information is smoother. During mentioned movement the Gibbs phenomenon appears on the edges and the structural component of the SSIM metric change. That is why also the SSIM quality assessment in this part of the sequence varies. Applying the LoG filters before computation of mutual information caused that changes in the quality of overall run at the start of the sequence correspond more to the SSIM. However not all of the peaks or bottoms occur at the same place.

The first bigger notable improvement of quality SSIM indicates when the fast move of the hand blurs the part of the frame in the original sequence. The Gaussian filter implementation reacts on this fact also by improving the quality. However on the following decrease of the SSIM the mutual information with Gaussian filter reacts after 20 frames later. On the other hand the applying the LoG filters causes that the mutual information evaluates this blurring in the frame as a degradation of the quality.

The major ascent of the quality is indicated by SSIM when the camera moves and blurs the major part of the frame which contains the background with the few colors. This blurriness appears also in the original sequence. However the run of

mutual information with the Gaussian filter falling down when this happens. Implementation where LoG filters are used reacts also by the decreasing the quality even if there is the peak when the quality is falling down. However this peak indicates worse quality as before the decreasing and occurs a few frames later. On the other side the mutual information in case where the second and fourth HVS model is used, indicates some quality improvement at this part of the sequence. The biggest one indicates the mutual information with the impulse response function $h_2(t)$.

After major SSIM quality ascents in the frame 192, the quality has decreasing trend with some peaks and the lowest value has at the frame 231 due to mixture of colors on the wall during the encoding. This causes the change of the contrast and structural component of SSIM. The run of the mutual information with Gaussian filter rise and slightly vary in quality with no bigger peaks or bottoms. The mutual information together with LoG filters also indicates improvement of quality but the peaks and the bottom are more noticeable. The overall run of mutual information with second HVS is rising and falling during the whole sequence. The final correlation is higher even if there is no noticeable improvement or degradation of quality. The mutual information with last HVS model indicates quality very different, lots of peaks and bottoms despite of the SSIM run even if the implementation with $h_2(t)$ has smoother run in comparison with $h_1(t)$.

Comparison between VQM and mutual information pre-processed with the HVS is shown in Fig. 2. At the start of the sequence the run of VQM has some oscillation in terms of the quality. Moving of the head causes changes in the local contrast due to fact that the face is darker area and the helmet is brighter. This appears as oscillations in the VQM quality. Mutual information with each HVS also contains some peaks and bottom in this part of the sequence. The Gaussian filter reduces differences between quality oscillations which do not correspond to the VQM run. In the implementation with LoG filters, the mutual information indicates little improvement of the quality at the beginning but then the peaks better corresponds with VQM. The second HVS model causes that run of the mutual information is rising and falling down in the same frames but in the reverse order. It means that if the VQM indicates improvement of the quality, mutual information indicates its degradation. That is the reason why the final correlation coefficient in the Tab. I is negative. When movement of hand blurs just the part of the image in the VQM run does not show any evident change, only a very slight improvement regarding the previous oscillations.

In the frame 192 VQM has rapid quality grow. This is caused by moving the camera and thus blurring the frame in the original sequence. Mutual information together with filters for simulate HVS (Gaussian and LoG filters) indicate degradation of the quality at this point. LoG filters have one exception from decreasing trend but this peak shows just slightly quality grow and does not affect overall descending character of the run. The second HVS model has two peaks at this part of sequence and one of them is corresponding with the reducing VQM quality. After major improvement of quality in the frame 192, the overall quality is falling down. Mutual

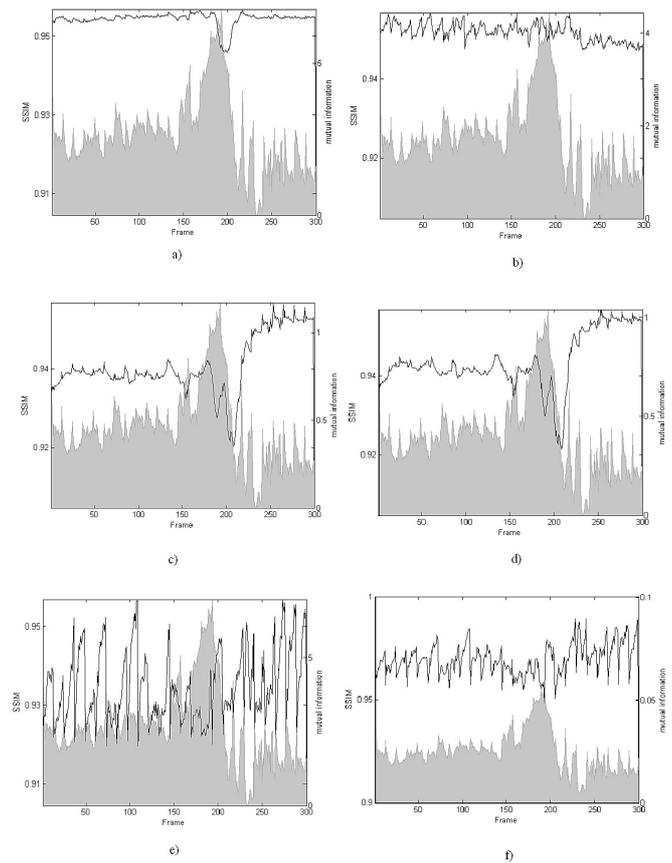


Fig. 1. Figures show the comparison between SSIM and the mutual information with different types of HVS. The left vertical axis and gray curve correspond to the SSIM. The right vertical axis and black curve correspond to the mutual information after applying the particular HVS: (a) Gaussian filter, (b) second HVS model, (c) LoG filter with $\sigma = 1$, (d) LoG filter with $\sigma = 1.2$, (e) HVS simulated by $h_1(t)$, (f) HVS simulated by $h_2(t)$.

information with first three HVS models start to raise from the frame 192. First and second HVS model reach approximately the same quality as before. However the third HVS model indicates better quality in comparison with the beginning of the sequence. Mutual information with the last HVS model contains many peaks where some of them occur at same place as the VQM peaks and some of them occur where VQM has bottoms.

Figure 3 and Fig. 4 show the comparison between the Peak signal-to-noise ratio, the Minkowski-form distance and mutual information with implemented every HVS model. The Minkowski-form distance and PSNR have very similar run. The quality peaks and bottoms alternate from the beginning of the sequence. As mentioned before mutual information has some quality oscillation for every HVS. By applying the Gaussian filter mutual information run becomes smoother so the correlation between mutual information and both of these metrics is not very high. In the case where third HVS model is used the changes in the quality are smaller than those indicated by Minkowski-form distance but corresponds well with the run

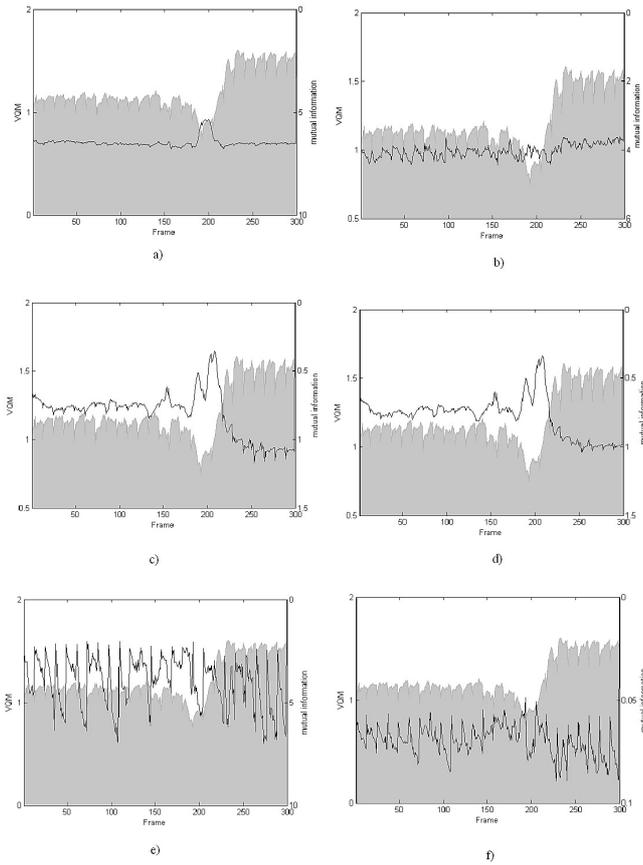


Fig. 2. Figures show the comparison between VQM and the mutual information with different types of HVS. The left vertical axis and gray curve correspond to the VQM. The right vertical axis and the black curve correspond to the mutual information after applying the particular HVS: (a) Gaussian filter, (b) second HVS model, (c) LoG filter with $\sigma = 1$, (d) LoG filter with $\sigma = 1.2$, (e) HVS simulated by $h_1(t)$, (f) HVS simulated by $h_2(t)$. Note that the axes are in the reverse order what corresponds with the fact that more similar pictures have mutual information higher but VQM value lower.

of PSNR. The second and fourth HVS model cause mutual information to vary more in the quality. Some of the peaks in the second HVS model correspond with peaks of both metrics.

A slight improvement of quality in the place where part of the frame is blurred indicates only PSNR metric. This change corresponds with the run of mutual information pre-processed by the second and the third HVS model. However in case of LoG filters implementation this change is inverse to the improvement indicated by the PSNR. The Minkowski-form distance does not indicate any noticeable improvement that corresponds with the mutual information combined by the Gaussian filter. Different behavior appears when the whole frame is blurred because of the camera movement. The Minkowski-form distance indicates degradation of quality in this part of sequence while PSNR quality grows in the same way as SSIM and VQM quality. All mutual information implementations indicate degradation of quality so the correlation between them and PSNR is not well enough except in the case when $h_2(t)$ is used. From this point up the end of the

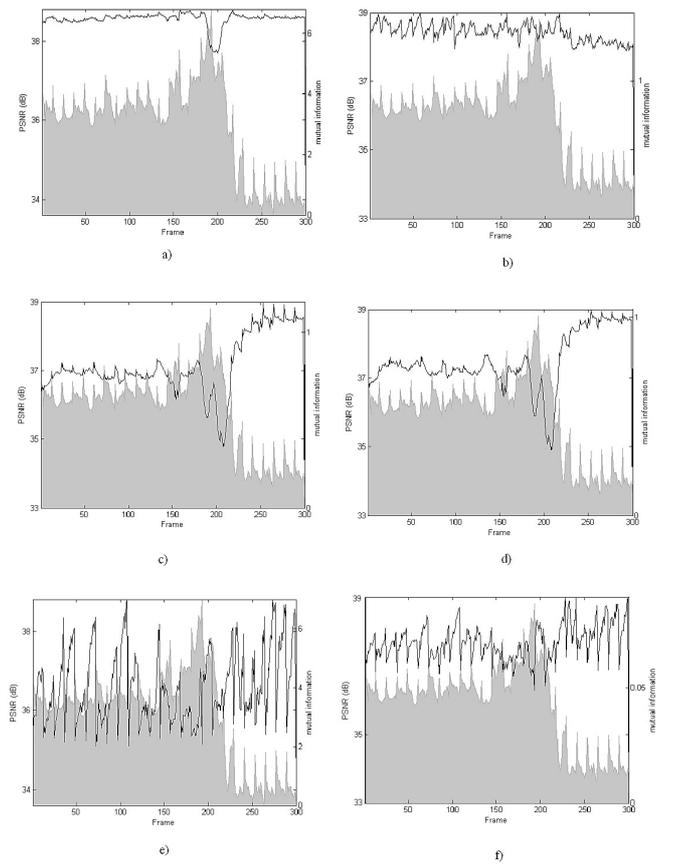


Fig. 3. Figures show the comparison between PSNR and the mutual information with different types of HVS. The left vertical axis and gray curve correspond to the PSNR. The right vertical axis and the black curve correspond to the mutual information after applying the particular HVS: (a) Gaussian filter, (b) second HVS model, (c) LoG filter with $\sigma = 1$, (d) LoG filter with $\sigma = 1.2$, (e) HVS simulated by $h_1(t)$, (f) HVS simulated by $h_2(t)$.

sequence the quality oscillates for both metrics. PSNR has the peaks and bottoms more visible due to changes of intensity of pixels belonging to wall. On the other side, Minkowski-form distance has run smoother with only few peaks and bottoms. These oscillations correspond with the runs of mutual information. Peaks and bottoms of LoG filters implementation best correlate with the PSNR changes. The smoother run of the Gaussian filter does not contain any essential peaks, while the fourth HVS model contains a lot of peaks especially implementation with $h_2(t)$.

V. CONCLUSION

In this paper we present objective video quality metric based on mutual information and Human Visual System. The evaluation of the quality in our metric is done in two steps. First of all, the original and the test sequence are pre-processed by the HVS. We choose four different types of HVS to find out which one will provide the best results. In the second step the mutual information is calculated between the frame from the original sequence and the corresponding frame from the test

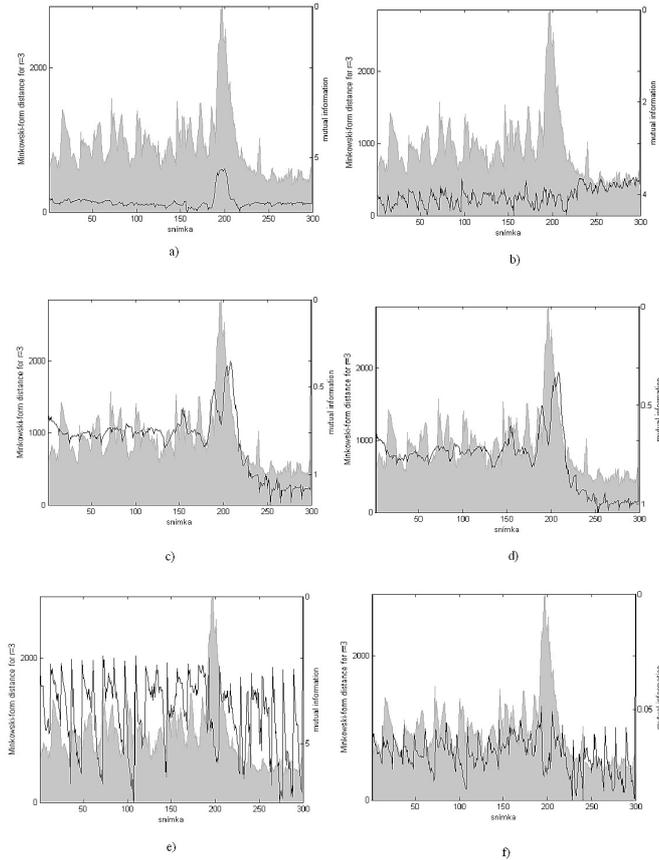


Fig. 4. Figures show the comparison between Minkowski-form distance and the mutual information with different types of HVS. The left vertical axis and gray curve correspond to the Minkowski-form distance. The right vertical axis and the black curve correspond to the mutual information after applying the particular HVS: (a) Gaussian filter, (b) second HVS model, (c) LoG filter with $\sigma = 1$, (d) LoG filter with $\sigma = 1.2$, (e) HVS simulated by $h_1(t)$, (f) HVS simulated by $h_2(t)$. Note that the axes are in the reverse order what corresponds with the fact that more similar pictures have mutual information higher but Minkowski-form distance value lower.

sequence. To verify the relevance of obtained results from our metric we compare it with standardized and some commonly used objective methods like SSIM, VQM, PSNR and with one representative of distance metrics Minkowski-form distance with parameter $r = 3$. Results shows that by implementing the simple low-pass Gaussian filter as simulation of human visual perception the run of the mutual information become smoother. In this case the overall results show only little correlation between mutual information and objective methods used for comparison. The second HVS based on band-pass filter and DCT provides better results as the Gaussian filter. However the correlation of the second HVS model is still small. The third set of results was obtained when the HVS model in the first step of calculation in our metric is simulated by LoG filters. We choose two types of these filters. Run of filter with $\sigma = 1.2$ is characterized by less peaks and smoother rising unlike the filter with $\sigma = 1$ where the crossing between bottoms and peaks is more rapid. In the last case we simulate HVS using

TABLE I
THE NORMALIZED CORRELATION COEFFICIENT BETWEEN THE MUTUAL INFORMATION WITH PARTICULAR HVS AND REFERENCE OBJECTIVE METRICS SSIM, VQM, PSNR AND MINKOWSKI-FORM DISTANCE

HVS model	metric			
	SSIM	VQM	PSNR	Mink
No HVS	0.1424	-0.2196	0.2593	-0.0667
Gaussian filter	-0.1125	0.0978	-0.0656	-0.2506
Second HVS	0.262	-0.4408	0.4354	0.3263
LoG 1.0	-0.4799	0.8085	-0.7843	-0.6183
LoG 1.2	-0.474	0.7863	-0.7628	-0.6109
$h_1(t)$	-0.3387	0.364	-0.3574	-0.0983
$h_2(t)$	0.4026	-0.4822	0.4746	0.1545

two different impulse response functions. By using this HVS the number of oscillations in the video quality grow rapidly. From the comparison with other objective methods it can be seen that this model of HVS is not suitable to be used with mutual information even if the correlation coefficient is not the smallest. The best results are provided by LoG filter with parameter $\sigma = 1$ where the correlation between the mutual information and the VQM metric is above 0.8. It seems that the mutual information is sensitive on the massive blurriness in the frame and reacts on this fact by the degradation of the quality. The results show that implementing of simple HVS together with mutual information could provide good correlation so our metric could be useful in objective evaluation of quality. For future work, we would like to run a more complex set of experiments with different video sequences to prove the relevance of the proposed method. Furthermore subjective testing will be necessary to run to compare our method with human perceiving of video quality. It is know that different parts of video frame have different influence on evaluation of video quality by human so implementation of region of interests can possibly improve the obtained results.

REFERENCES

- [1] A. B. Watson, J. Hu, and J. F. McGowan, "Dvq: A digital video quality metric based on human vision," *Electronic Imaging*, 2001.
- [2] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union - Radio-communication 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," Sep. 2006.
- [4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [5] Y. Wang, "Survey of objective video quality measurements," EMC Corporation, MA, Tech. Rep. WPI-CS-TR-06-02, 2006.

- [6] S. Winkler, *Digital video quality vision model and metrics*. Chichester: John Wiley & Sons Ltd., 2005.
- [7] F. Xiao, "Dct-based video quality evaluation," Final Project for EE392J, Tech. Rep., 2000.
- [8] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vision*, vol. 40, pp. 99–121, November 2000. [Online]. Available: <http://portal.acm.org/citation.cfm?id=365875.365881>
- [9] S. Westen, R. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel hvs model," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 4, pp. 2351–2354, 1995.
- [10] I. Avcibas, I. Avcba, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, vol. 11, pp. 206–223, 2002.
- [11] R. Tiwari and A. Yardi, "Dental x-ray image enhancement based on human visual system and local image statistics." in *International Conference on Image Processing, Computer Vision, Pattern Recognition (IPCV)*, 2006, pp. 100–108.
- [12] K. Minoo and T. Nguyen, "A perceptual metric for blind measurement of blocking artifacts with applications in transform-block-based image and video coding," 2008, pp. 3152–3155.
- [13] Z. Yu and H. R. Wu, "Human visual system based objective digital video quality metrics," in *Proceedings of the International Conference on Signal Processing of IFIP World Computer Conference*, vol. 2, August 2000, pp. 1088–1095.
- [14] (2010) Full reference image and video quality assessment. Laboratory for Image & Video Engineering. [Online]. Available: <http://live.ece.utexas.edu/research/quality/frqa.htm>
- [15] R. Gray, *Entropy and information theory*. Springer-Verlag, 2009.
- [16] Z. Cernekova, "Temporal video segmentation and video summarization," Ph.D. dissertation, Comenius Univ., Bratislava, 2009.