World Academy of Science, Engineering and Technology
International Journal of Information and Communication Engineering
Vol:7, No:7, 2013

Video Quality Control Using a ROI and Two-Component Weighted Metrics

Petra Heribanová, Jaroslav Polec and Michal Martinovič

Open Science Index, Information and Communication Engineering Vol:7, No:7, 2013 publications.waset.org/16328/pdf

***Abstract*—**In this paper we propose a new content-weighted method for full reference (FR) video quality control using a region of interest (ROI) and wherein two-component weighted metrics for Deaf People Video Communication. In our approach, an image is partitioned into region of interest and into region "dry-as-dust", then region of interest is partitioned into two parts: edges and background (smooth regions), while the another methods (metrics) combined and weighted three or more parts as edges, edges errors, texture, smooth regions, blur, block distance etc. as we proposed. Using another idea that different image regions from deaf people video communication have different perceptual significance relative to quality. Intensity edges certainly contain considerable image information and are perceptually significant.

***Keywords*—**Video quality assessment, weighted MSE.

I. Introduction

MORE techniques and metrics for objective evaluation of video quality, where the main criterion is "lovely" of video regardless of content. 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 new objective video quality metrics, i.e. weighted normalization mean square error with non or unnormalized mean square error evaluating by subjective results. We used set of 41 video sequence presented in [9]. These content-based metric were correlated with the objective "subjective" methods. In Section II we described the objective metric, which was used for testing. The region recognition and classification is described in Section III. Then the results and discussion of our metrics are presented in Section V. The final section concludes our proposed work.

However in the evaluating of video with different methods of implementation of augmentative and alternative communication (AAK) [10] we cannot ignore content – intelligibility of video. The main difference between the terms quality and intelligibility is that the term "quality" describes the appearance of decoded video signal ("how" the viewer sees it) and the "intelligibility" is just one aspect of quality saying if the received information gives any sense ("what" the viewer sees in it). High-quality video signal is likely to be intelligible. Conversely, of course it may or may not apply. Anyway, unintelligibility is an indicator of poor quality. In the acoustics, intelligibility threshold is defined as a point, after which one does hear, but one does not understand [5]. Generally: IQA algorithms generally operate without attempting to take into account image content. Since algorithms for image content identification remain in a nascent state, IQA algorithms that succeed in assessing quality as a function of content will await developments in that direction. For example, intensity edges certainly contain considerable image information, and are perceptually significant [6].

Subjective tests show that sound tends to reduce people's ability to recognize video image degradation. Hearing-impaired people do not rely that much on video quality, as the most important thing to them is whether they are able to understand the meaning.

II. Correlation Based On Video Quality Assessment

A. MSE

The mean square error is defined as:

$$MSE = \sum_i \left(X_i - \hat{X}_i \right)^2 \tag{1}$$

B. Pearson Correlation Koeficient Measurement

The Pearson correlation coefficient is defined as:

$$R_P = \frac{\sum_i \left[\left(X_i - \bar{X} \right) \left(Y_i - \bar{Y} \right) \right]}{\sqrt{\sum_i \left[\left(X_i - \bar{X} \right)^2 \left(Y_i - \bar{Y} \right)^2 \right]}} \tag{2}$$

P. Heribanová is with the Department of Algebry, Geometry and Didactic of Mathematic, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava (e-mail: petra.heribanova@gmail.com).

J. Polec is with the Institute of Telecommunication, Slovak University of Technology, Ilkovičova 3, 812 12, Bratislava, Slovak Republic (phone: +421268279409, email: polec@ktl.elf.stuba.sk).

M. Martinovič is with the Institute of Telecommunication, Slovak University of Technology, Ilkovičova 3, 812 12, Bratislava, Slovak Republic (email: michal.martinovic@gmail.com).

Open Science Index, Information and Communication Engineering Vol:7, No:7, 2013 publications.waset.org/16328/pdf

III. The Proposed Method

A. ROI

Since the video contains areas that are for us in terms of intelligibility finger alphabet irrelevant, we divided the video into regions of interest according to their importance.

The ROI determination for each frame is performed in two steps - hand tracking and segmentation:

1) Mean shift based tracking extracts the color distribution of target appearance, and is implemented using kernel histogram.
2) A single threshold value is assigned to every image.

For experiment is used simple definition of ROI (just dominant region is bounded as sole bar around dominant hand and face) [7], [8].

B. Gradient and Background Mask

For gradient calculation is used conventional Sobel operator and threshold defined by [4].

C. Normalized and Weighted MSE

Let MSE value are in the range $\langle a,b \rangle$, then NMSE value are in the range $\langle 0,1 \rangle$.

$$NMSE = \frac{MSE - a}{b - a} \tag{3}$$

Then (3) can rewrite

$$NMSE_e = \frac{MSE_e - a_e}{b_e - a_e} \tag{4}$$

$$NMSE_b = \frac{MSE_b - a_b}{b_b - a_b} \tag{5}$$

Normalized MSE value of edge and background are weighted:

$$WNMSE = \alpha \cdot NMSE_e + \beta \cdot NMSE_b \tag{6}$$

A diagram depicting calculation of ROI-2-NMSE (WNMSE) is shown in Fig. 1.

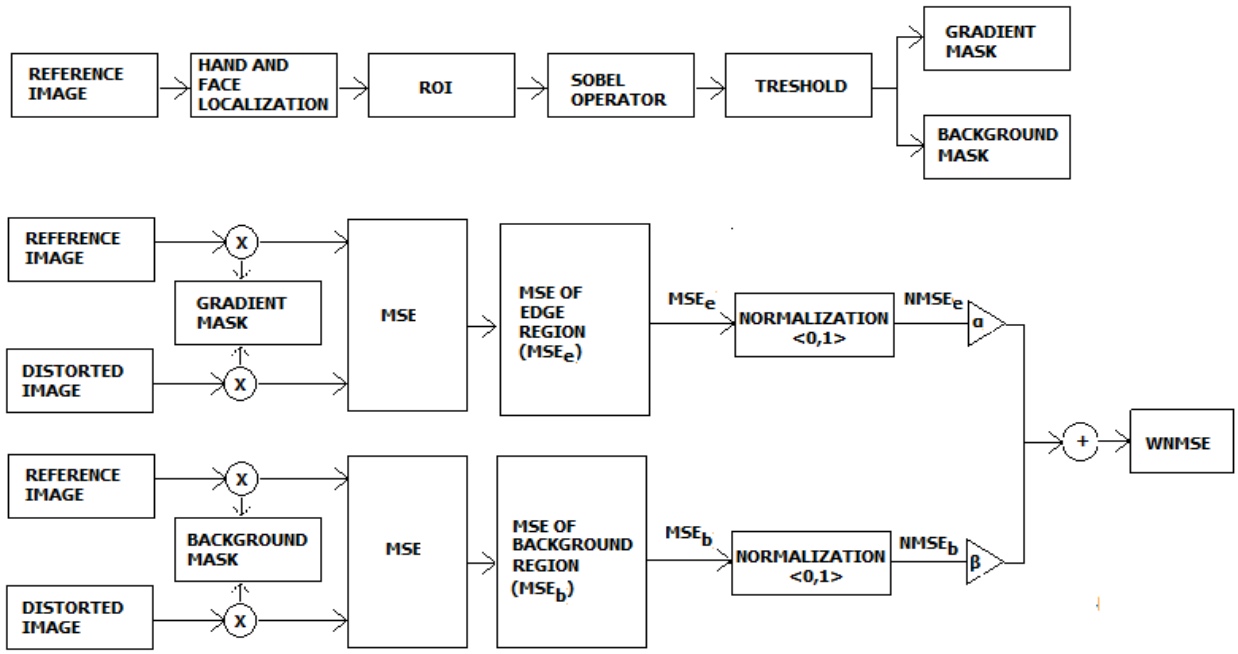


Fig. 1 Block diagram of the proposed quality control system

IV. Experiment

We work on the basis of full reference method (FR) with differential metrics to evaluate image quality and video between the original and processed video. As the original videos we used the primal videos in format "4-cif" (704x576). The processed video stream in other format "cif" (352x288) and "qcif" (176x144) we resized bilinear interpolation to size of format "4-cif".

We created the following experiment. We produced video preview with seven different logatoms in Slovak single-handed finger alphabet with 41 spells. The length of the video previews is about one minute.

For the whole experiment we used different video formats with 25 frames per second. Subsequently, these recordings were encoded by the H.264 codec in various bit rates (QP =

Open Science Index, Information and Communication Engineering Vol:7, No:7, 2013 publications.waset.org/16328/pdf

30, 40, 50 that corresponds to rates from 390 kbit/s to 4.5 kbit/s respectively). Testing was realized according to subjective ACR method on groups of hearing impaired volunteers. A random sequence of consonants is quite hard to remember; therefore some sequences were shown multiple times to the same people (in different bit-rate and/or video format) without mentioning it in advance.

Respondent had to rewrite the consonants organized into logatomes to the letters of the Slovak alphabet. While the sentence intelligibility evaluation was based on subjective rating, the logatom recognizability expresses the correctness of all consonants in logatom in percents.

The results obtained percentages evaluation of recognizability as used in acoustics.

Based on these results of objective evaluation of spell recognizability with respondent involvement we test the MSE method, and new WNMSE method, which correlate best with intelligibility, and therefore could represent a method for automatic evaluation of video intelligibility with finger alphabet.

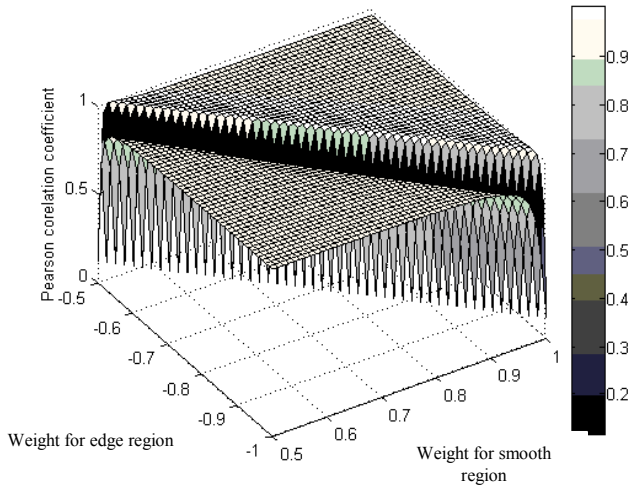


Fig. 2 Weights find

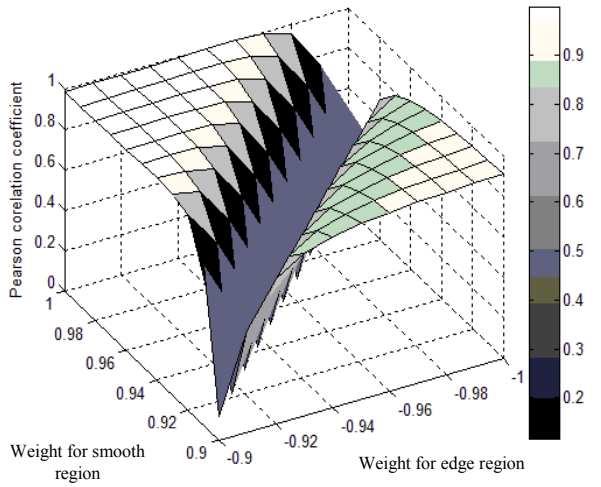


Fig. 3 Zoom of weights find



Fig. 4 One image finger



Fig. 5 Edge mask for image finger of Fig. 4

TABLE I
SETTING PARAMETERS (4CIF FORMAT)

Spell recognizability [%]	NMSE Edge	NMSE Background	WNMSE	MSE
94,840	0	0	0	6,83504
73,400	0,3331	0,3442	0,024091	21,9895
59,230	1	1	0,04	57,1223
Correlation	-0,9530	-0,9567	-1	-0.9418

N=normalized, W=weighted

TABLE II
EVALUATION (CIF FORMAT) (α,β)=(-0.93, 0.97)

Spell recognizability [%]	NMSE Edge	NMSE Background	WNMSE	MSE
91,46	0	0	0	18,24166
49,16	0,3615	0,3949	0,04686	39,47023
39,39	1	1	0,04	85,39771
Correlation	-0.8753	-0.8930	-0,9513	-0,8497

N=normalized, W=weighted

V. RESULTS

The Pearson correlation coefficient for setting video in 4CIF video format is equal to -0.9530 for edge region, -0.9567 for background region and only -0.9418 for conventionally quality evaluated (assessment) video. The normalized range of MSE values and setting weighting for edge region and smooth region maximized the correlation between objective evaluation

Open Science Index, Information and Communication Engineering Vol:7, No:7, 2013 publications.waset.org/16328/pdf

(inteligibillity) and objective video quality assessment with metric. The ideal Pearson correlation equal to -1 is achieved when the weight of edges is equal to -0,93 and the weight of smooth region (background) is equal to 0,97. This setting was made from 4CIF video format. Correlation (with intelligibility) results for these settings are demonstrated for 4CIF video format in Table I. When we use this setting in CIF video format, the Pearson correlation with intelligibility is improve - see the Table II. The Pearson correlation coefficient for edge region is equal to -0.8753for smooth region -0.8930 and for conventionally quality evaluated (assessment) video only - 0.8497.

New metric (WNMSE - Weighted Normalized Mean Square Error) improve correlation with intelligibility - in our experiment from 0.9418 for setting sequence to 1.0000 in ideal condition, and from 0.8497 to 0.9513 for slovenly test setting of different spatial format of same video.

VI. Conclusion

Experimental results show:

- the proposed algorithm outperforms existing video quality assessment method MSE.
- better overall performance is achieved by combining information of ROI tracking and weighting of metric results for different class image pictures.
- that it is possible find weights, evenly correlation with intelligibility is maximum, equal to 1 (other -1). Previously published works used values of weights in range from 0 to 1. We show that value of weight can be negative, too.

Acknowledgment

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

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] 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.

[5] F.Makáň: Elektroacoustics (in Slovak), Publisher STU Bratislava, 1995.

[6] Ch. Li, A. C. Bovik, “Content - weighted video quality assessment using a three - component image model.” In: *Journal of Electronic Imaging*, vol. 19, 2010, no.1, pp. 011003 - 1 – 011003 - 9.

[7] Beniak, M., Pavlovičová, J., Oravec, M., "3D Chrominance Histogram Based Face Localization", In: *Int. Journal of Signal and Imaging Systems Engineering* (IJSISE), Vol. 4, No.1 pp. 3 - 12, 2011, www.inderscience.com/ijsise

[8] Ban, Jozef - Pavlovičová, Jarmila - Féder, Matej - Omelina, Ľuboš - Oravec, Miloš: Face Recognition Methods for Multimodal Interface. In: Proceedings of 2012 5th Joint IFIP Wireless and Mobile Networking Conference: Bratislava, Slovakia, September 19-20, 2012. - Piscataway: IEEE, 2012. - ISBN 978-1-4673-2994-1. - S. 110-113

[9] P. Heribanová, J. Polec, J. Poctavek, A. Mordelová: "Intelligibility Threshold for Cued Speech in H.264 Video Conference". In: *International Journal of Electronics and Telecommunications*. - ISSN 0867-6747. - Vol. 57, Iss. 3 2011, pp. 383-387

[10] Tarcsiová, D.: The communication system for deaf and ways to overcome their communication barriers. (In Slovak) Sapientia: Bratislava, 2005. ISBN 80-69112-7- 9.

J. Polec was born in 1964 in Trstená, Slovak Republic. He received the M.Sc. and PhD. degrees in telecommunication engineering from the Faculty of Electrical and Information Technology, Slovak University of Technology in 1987 and 1994, respectively. From 2007 he is professor at Department of Telecommunications of the Faculty of Electrical and Information Technology, Slovak University of Technology and at Department of Applied Informatic of Faculty of Mathematics, Physics and Informatic of Comenius University. His research interests include Automatic-Repeat-Request (ARQ), channel modeling, image coding, reconstruction and filtering.

P. Heribanová was born in 1986 in Kremnica, Slovak Republic. She received M.Sc. degree in Geometry from the Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava in 2010. She is a PhD. student of Geometry and Topology at the same university. Her research interests include image coding, reconstruction and quality evaluation.

M. Martinovič was born in 1987 in Krupina, Slovak Republic. He received his Bc. (bachelor degree) and Ing. (MSc.) in Telecommunication from the Slovak University of Technology (SUT) in Bratislava, in 2010 and 2012, respectively. From 2012 until the presence, he has been an internal PhD. student at the Institute of Telecommunication, (SUT). His research interests are concentrated on Automatic-Repeat-Request (ARQ), error control coding for wireless communications and also signal processing. From 2011 until the presence, he has been with the Intech Control spol. s r.o.