No-Reference Algorithms for Video Quality Assessment based on Artifact Evaluation in MPEG-2 and H.264 Encoding Standards

J. P. López, D. Jiménez, A. Cerezo, J. M. Menéndez

Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universidad Politécnica de Madrid
Madrid, Spain
E-mail addresses: {jlv, djb, ace, jmm}@gatv.ssr.upm.es

Abstract— **In this paper, we propose two algorithms to assess quality in video sequences when no reference is available. The algorithms are the result of measuring artifacts such as blurring or tiling to the decoded image; firstly tested in MPEG-2 compression standard, and subsequently, analyzing the passage from this standard to its evolution H.264, taking into account its advanced techniques to mask the degradation, for example, the deblocking filtering or the variation in macroblocks size. The result of each metric affects in a different way depending on the spatial and temporal complexity of the video sequence. Several tests have been developed in a collection of representative sequences to demonstrate the efficiency of the algorithms.**

Keywords— Video Quality Assesment, No-Reference, H.264, Tiling, Blurring.

I. INTRODUCTION

The volume of multimedia content transmission is increasing nowadays. Every year, the traffic of video through the network is much bigger [1], especially because of the adoption of devices such as *smartphones* or tablets. For that reason, the end users demand a specific quality, and that is why the service providers need to fulfill their expectations. To ensure quality, video assessment is developed in order to provide tools to manage the quality issue in new media services, available for the different resolutions and environments, surrounding the observer.

The most effective method is developing subjective studies, by asking the observers to score video sequences with different qualities. The problem with these studies is that they require much time and are expensive. To replace them, objective metrics are used, but as the multiplicity of encoding standards and variety of formats and sequences are higher, the strength of the algorithm must be improved, to assure their validity in different environments. Objective metrics are software tools that provide quantitative results of image quality or image distortion while achieving a reduction of cost. Nevertheless, the results of these studies are only suitable when they are well-correlated to the subjective studies.

Tests have been performed in standards such as MPEG-2 and H.263, and efforts to allow the extension to other standards and its consequent usage, such as H.264/AVC. The basis of these three standards is the same, but the differences among them are enough to make necessary the individual analysis of each standard.

This paper explains the studies developed in NR objective assessment. The work started with low resolution video related to mobile video environment for tablets and *smartphones*, and resolutions were increased up to high-definition video. Extensive work has been developed in this area, but the models we present offers a sensitive improvement in detection of the most frequent video distortion and artifacts, which is confirmed by the high correlation in comparison with subjective assessment.

Thus, a collection of video sequences with different types of artifacts has been compiled. This set provides the most effective method for testing and evaluating the developed artifact metrics. These sequences have been generated by simulating transmission effects on the original video, or by compressing image at different bitrates or settings.

This set has allowed the validation of a collection of particular metrics that had been selected for each of the main artifacts.

The framework in which this research started with sequences whose definition was QCIF (176x144 pixels), i.e. low resolution interlaced video, encoded in H.263 standard, at average bitrates of 128Kbps, as required from a mobile company. The study has been extended to more efficient compression standards and higher resolutions, to admit it as a complete research work in this field to extrapolate the results obtained to other environments.

Subjective assessment with observers was also conducted, in order to obtain subjective results that could be compared to the ones obtained by our mathematical algorithms in the objective assessment process.

II. RELATED WORK

Image quality assessment is a difficult process which plays a major role in various processing applications [2]. A lot of work has been developed in this field, defining metrics and algorithms to predict the quality of a video sequence.

The usual metrics developed for video assessment, such as PSNR and MSE requires a video reference, as a way of comparison to detect the degradations on the image, which is called Full-Reference (FR) assessment. An overview of the extensive and most interesting work in objective quality assessment is collected in [3], [4] and [5].

The problem of these algorithms is the lack of availability of a reference in all the environments, especially related to internet

978-3-901882-50-0 ©2013 IFIP

and streaming transmissions. That is the importance of No-Reference (NR) assessment and the interest in this working field. No-Reference metrics are usually based on the measurement of artifacts in sequences containing only the desired artifact signal or a combination of artifact signals, as in [6], with description of techniques to detect blurring, blockiness and ringing.

Other works are specialized in the measurement of one or more artifacts with NR metrics. There are different studies developed in detecting blocking artifact with different techniques, such as [8], [9] and [10].

Also there are works in which the primary objective is detecting blurring artifact [11] or a combination of artifacts [10] to obtain a global algorithm developed through individual metrics.

Specific works are developed in the field of video encoding standard H.264/AVC, which is one of the main aspects in research in this paper, proposing new techniques of working with this kind of compression [12].

III. TEST DATASET

Aligned with the objectives, one of the first steps is the selection of contents used to test the metrics, which must be suitable to represent all the range of contents which normally are broadcasted on a conventional television channel and streaming supplier, including images of documentaries, news or talking head, cartoons or music content. An example of each individual sequence used is shown below in Fig. 1.

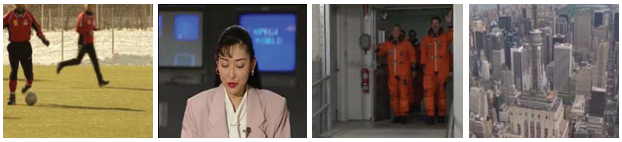


Fig. 1. Example of the main sequences used in the tests: "Soccer", "Akiyo", "Nasa" and "City"

Example of sequences "Dixon" and "Stewie" are not included, because the content was provided by a broadcaster under a non-disclosure agreement.

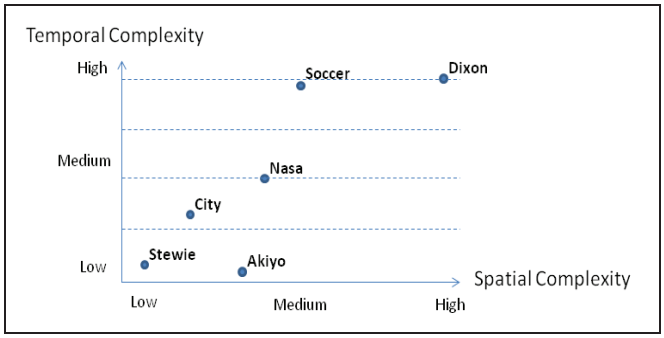


Fig. 2. Distribution of sequences depending on their complexity

Most of the sequences are available in database [13]. The dataset covers a variety of sequences with different degree of complexity, both temporal and spatial, as seen on Fig. 2, selected among a large collection of reference video sequences.

The human eye tends to detect the artifacts in a different way, depending of the motion of the frames, and the level of detail contain on a single frame, that is why the basis of the algorithm is fully related to the spatial and temporal complexity. The calculation of spatial complexity is based on measurement of variance for macroblocks. Moreover, the calculation of the temporal complexity is based on measurement of the temporal variation of the image, in formal difference between frames, and longitude of the motion vectors.

Sequences from SVT [14] have also been used for testing higher resolutions: "Crowd", "Ducks", "OldTown" and "ParkJoy".

IV. DESCRIPTION OF THE ALGORITHMS

A. Tiling (or Blocking)

The artifact of tiling (also named blocking), a block boundary effect, is usually present in conventional block-by-block transform coding techniques when a severe reduction in bitrate is applied, such as used in MPEG-2 or H.264. When compression is high, the tiling effect increases, making it more noticeable to the human eye.

As a conclusion of subjective tests developed previously, the blocking effect is more visible in some specific areas of the image, rather than the edges of the 8x8 blocks used in DCT-based H.263 and MPEG-2 coding, so the measurement is concrete pixels in the contour of each individual macroblock.

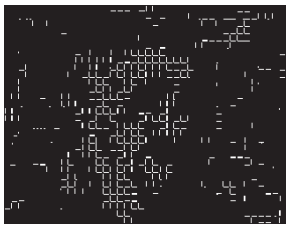
Three conditions must be fulfilled to consider a pixel as belonging to the tiling region of the image. First, the gradient of the image must be estimated in every pixel by means of the Sobel operator [15] in its horizontal and vertical directions. It is necessary to know both module and angle to obtain the best result. The angle of the gradient of the pixel must have an orthogonal direction, and correspond with a clear border at the block edge. Second, the module of the angle must be higher than an empirically fixed threshold value to be visible by the human eye in order to decide if that pixel belongs to the tiling region or not.

Lastly, the grating mask is of vital importance in order to distinguish the real pixels in which the tiling highly affects the video quality, discarding the pixels not belonging to the meaningful tiling regions. For this purpose, properties of a variety of transforms, such as Hadamard or DCT, are used after compression of the video data to be transmitted through a communication channel. Therefore, the pixel must be on the grid of 8x8 pixels among macroblocks. Depending on the coding standard, gratings are different. Some encoding processes can use different sizes for the macroblocks (for example H.264).

Summarizing the previous three conditions:

$$Pixel(i,j) \in Tiling \begin{cases} \varphi Grad(i,j) = 0^{\circ}, 90^{\circ}, 180^{\circ}, 270^{\circ} \\ |Gradient(i,j)| > Threshold \\ Pixel(i,j) \in Grid(MPEG-2/H.264) \end{cases} \quad (1)$$

a) Image with a high tiling



b) Resulting tiling region in white

Fig. 3. Process of tiling detection on a H.263 sequence

The main difference between MPEG-2 and H.264 lies in the variability of macroblock size and the use of deblocking filter, both techniques involving the reduction of the blocking effect, but do not remove it completely.

As in MPEG-2 the macroblocks have fixed size of 8x8 pixels, in the case of the H.264, the size of the macroblock is dependant of the homogeneity of the area in which is located.

On the other side, the deblocking filter improves the global quality at the expense of increasing the blur on an image.

A study has been developed in order to compare the blocking effect in different environments, and consequently define the principles to detect this kind of artifact in video sequences.

The modulus of the gradient applying Sobel masks has been obtained for different sequences. In the original sequence, it is possible to see that although the area seem homogeneous in the clouding area of the sequence "Crowd", but the pixels are variable, so the resulting image reflects the real heterogeneity of the values, as seen on Fig. 4.

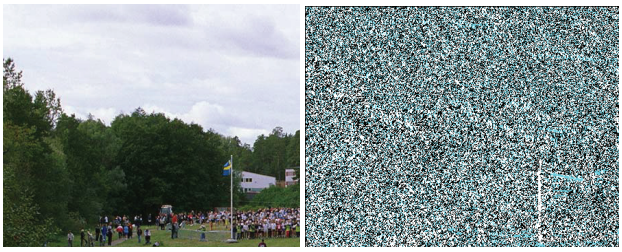


Fig. 4. Detail of image "Crowd" (left) with its gradient representation (right)

When encoding with standards MPEG-2 (Fig. 5) or even H.263, the grid is visible in little squares which reveal the presence of blocking artifact.

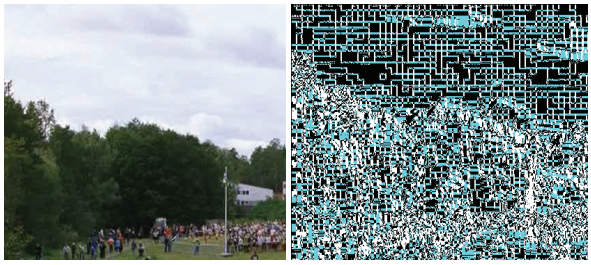


Fig. 5. Image "Crowd" (left) with its gradient representation (right), with MPEG-2 encoding.

In H.264, the deblocking filter [16] decreases the effect of tiling in small heterogeneous areas (Fig. 6), but in big areas, such as the clouds area, the artifact is still present and visible for the human eye [17]. As a conclusion, to detect the effect of this artifact and quantify its value, it is necessary to concentrate in the big homogeneous areas, and analyze the macroblocks of biggest sizes, especially the ones of 16x16, which must be considered.

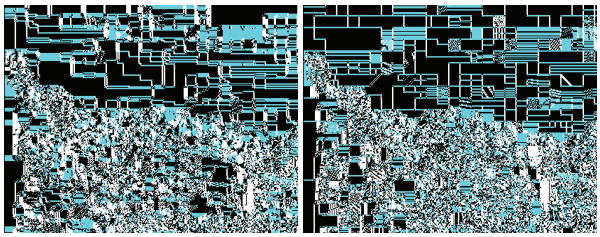


Fig. 6. H.264 encoded image with deblocking filter (left) and without deblocking filter (right) in big homogeneous areas

The detection of blocking in MPEG-2 is based on searching structures of this size (8x8) in homogeneous areas, while in H-264, the detection should be also extended to macroblocks of bigger sizes, such as 16x16, 16x8 or 8x16, independently of the use of the deblocking filter.

B. Blurring

The blurring artifact also appears when reducing the video encoding bitrate, presenting a defocus sensation on the image. Blurring is associated to blocking effect, because this artifact implies a loss of energy in edges of the image. The effect has a higher impact in highly detailed areas and high frequency components and, as a consequence, in edges. These are obtained by applying a Sobel filter to estimate the gradient [15]. By analyzing the energy loss of the image, a blurring estimation is provided, as seen in Fig. 7.

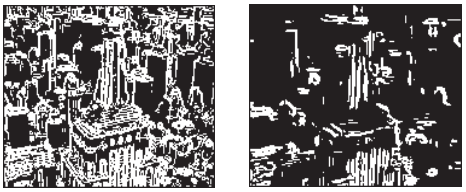


Fig. 7. Gradient in sharp (left) and blurred (right) images

Transforms, such as DCT or Hadamard [18] generate a matrix of transformed coefficients with the same size of the original macroblock. The coefficients represent the distribution of energy in frequency, the first corresponding to the DC component, the following coefficients to the lower frequencies, and progressively the higher frequencies information, with the details of edges and textures. For our purpose, an exhaustive analysis of macroblock size variation was carried out (8x8, 4x4, 2x2) transforms have been applied to get an estimation of energy distribution within the frame.

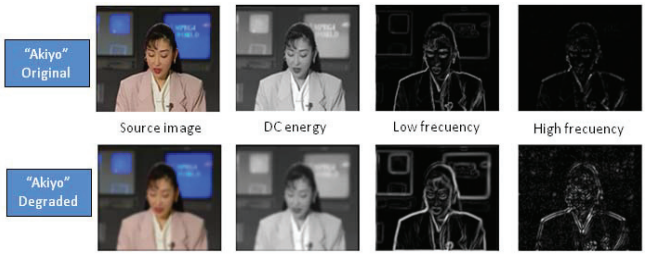


Fig. 8. Energy distribution for the original/degraded image

The Hadamard transform is selected to obtain an estimation of the distribution of energies in the image. It allows distinguishing three operational divisions: DC and low frequencies energy, medium and high frequencies energy.

The total accumulation of energy is obtained from the next equation, being "*n*" the dimension for the squared matrix used in the transform. This transform could be Hadamard or DCT, but Hadamard offers results faster than DCT, so it was chosen for the evaluation process.

$$E_{total}[T_{HAD}](x) = \sqrt{c_{00}^{\ 2} + c_{01}^{\ 2} + c_{10}^{\ 2} + \ldots + c_{nn}^{\ 2}}$$

The equation varies depending on the size of the matrix. As seen, there are three possible options for the Hadamard transform Matrix, and the computation of energies.

The formula that calculates the blurring (2) is the relation between the energy produced by high frequencies on the sequence and the energy by the DC component, when E(HF) is the energy related to high-frequency coefficients, and E(DC) the energy related to the continue coefficients.

$$Blur = \frac{E_{HF}}{E_{DC}} \qquad (2)$$

The tests were developed in H.263 and MPEG-2 but are also available for H.264 encoding. As seen in Fig. 9 percentage of high-frequency energy decrease when encoding in H.264. Also the effect of filters, such as deblocking filter, reduces the energy in edges, while producing some blurring effect.

The sequence "Crowd" was used to demonstrate it. This sequence possesses a high amount of edges which are decisive to the evaluation the level of blurring in the image.

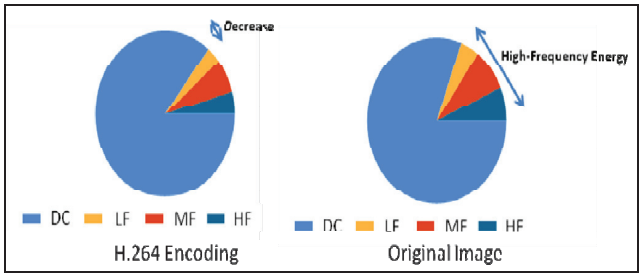


Fig. 9. Energy decrease in H.264/AVC encoding applied to a single frame in sequence "Crowd"

CONCLUSIONS

Metrics for measuring quality and assessing the level of artifacts such as blurring or tiling have been developed, in environments in which the reference was not available. The work started analyzing the Full-Reference algorithm and use the knowledge derived from it to obtain efficient No-Reference metrics.

The metrics tested in compression standards such as MPEG-2 or H.263 were extended to be used in H.264, with adequate results.

The results of the test varying the bitrates from high bitrates to low bitrates in MPEG-2 and H.263, and mean a high improvement in detecting these artifacts with H.264/AVC encoding. The concrete settings of each compression algorithm, as well as the advances in techniques used to reduce the volume of data, difficult the generalization of methods to detecting artifacts. Therefore, every step is important when considering the passage from full-reference to no-reference assessment, or from MPEG-2 to H.264/AVC.

ACKNOWLEDGMENT

This work has been partially funded by the National Project TEC2009-14219-C03-01 AMURA, and the Spanish telecommunication company Telefónica in order to obtain results in transmission of video contents through their mobile network. The work environment was defined by the company in the previous features of the project (resolution, encoding standard or mobile phones used for subjective tests)

REFERENCES

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010–2015. White Paper, February 2011

[2] Z.Wang, A.C.Bovik, and L.Lu, "Why is image quality assessment so difficult?," IEEE International Conference on Acoustics, Speech, & Signal Processing, vol, 4, pp. 3313-3316, May 2002.

[3] S. Winkler. "Digital Video Quality: Vision Models and Metrics". Ed. Wiley. March 2013. ISBN-13: 978-0470024041

[4] H.R. Wu, K.R. Rao, "Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)". CRC Press . November 2005. ISBN-13: 978-0824727772.

[5] Z. Wang, A. Bovik. "Modern Image Quality Assessment (Synthesis Lectures on Image, Video, & Multimedia Processing)". Morgan & Claypool Publishers . February, 2006. ISBN-13: 978-1598290226

[6] M. C. Q. Farias and S. K. Mitra. "No-Reference Video Quality Metric Based on Artifact Measurements". IEEE International Conference on Image Processing, 2005. ICIP 2005.

[7] Yamada, T.; Nishitani, T.; , "No-reference quality estimation for compressed videos based on inter-frame activity difference," Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on , vol., no., pp.2325-2328, 25-30 March 2012

[8] Vlachos, T., "Detection of blocking artifacts in compressed video," Electronics Letters, vol.36, no.13, pp.1106-1108, 22 Jun 2000.

[9] Uzair, M.; Fayek, D., "An efficient no-reference blockiness metric for intra-coded video frames," Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on.

[10] Mittal, A.; Moorthy, A.K.; Bovik, A.C.; , "No-Reference Image Quality Assessment in the Spatial Domain," Image Processing, IEEE Transactions on , vol.21, no.12, pp.4695-4708, Dec. 2012

[11] Liu Debing; Chen Zhibo; Ma Huadong; Xu Feng; Gu Xiaodong; , "No Reference Block Based Blur Detection," Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on , vol., no., pp.75-80, 29-31 July 2009

[12] Romaniak, P.; Janowski, L.; Leszczuk, M.; Papir, Z.; , "Perceptual quality assessment for H.264/AVC compression," Consumer Communications and Networking Conference (CCNC), 2012 IEEE , vol., no., pp.597-602, 14-17 Jan. 2012

[13] Xiph.org Video Test Media. http://media.xiph.org/video/derf/

[14] L. Haglund, "The SVT High Definition Multi Format Test Set." [Online]. Available at: ftp://vqeg.its.bldrdoc.gov

[15] González, R.C., Wintz, P. (1996). Procesamiento digital de imágenes.

[16] Mahjoub, W.H.; Osman, H.; Aly, G.M.; , "H.264 deblocking filter enhancement," *Computer Engineering & Systems (ICCES), 2011 International Conference on*, vol., no., pp.219-224, Nov. - Dec. 2011

[17] Woo-Seok Seo; Kwon-Yeol Choi; Min-Cheol Hong;, "Spatially adaptive gradient-projection algorithm to remove blocking artifacts of H.264 video coding standard," *Communications, 2008. APCC 2008. 14th Asia-Pacific Conference on*, vol., no., pp.1-5, 14-16 Oct. 2008.

[18] Aye Aung; Boon Poh Ng; Shwe, C.T.; , "A new transform for document image compression," *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, vol., no., pp.1-5, 8-10 Dec. 2009