Image communication quality assessment based on watermarking, an overview.

Tomás Brandão

Internal Report, v0.02 – Instituto de Telecomunicações, April 2005.

1. Introduction

Digital watermarking techniques were originally proposed as means of intellectual properties management regarding multimedia products such as still images, video and audio. In the past few years, however, watermarking techniques have been proposed as possible solutions to other different problems, such as: authentication, identification and indexing, content verification, broadcast monitoring and fingerprinting.

Other application where digital watermarking can be potentially used to solve some problems is the automated quality monitoring of multimedia transmission. The perceived quality at the users end is a relevant topic, especially due to increasing transmission of multimedia contents over the internet and over 3G mobile networks. It is expectable that content providers should be able to automatically monitor the quality of the received media, in order to optimize streaming services and bill end users proportionally to their perceived quality of service.

The most reliable metrics to evaluate the quality of media contents received by end-users are subjective metrics, i.e., it is the users that evaluate the quality of the received video or audio. Although they are the most realistic measurements, subjective metrics are difficult to obtain, since they require the organization of multiple tests with several subjects, and it is impossible to adjust in real-time some transmission parameters.

An alternative is the use of objective metrics. The main purpose of objective media quality assessment is to provide a set of quality metrics that can predict the perceived quality from the user's point of view. The ultimate goal is to develop a metric that exhibits the same behavior of a human observer without requiring access to the original media.

Since it is desirable to perform the evaluation without access to the original media, watermarking techniques could give a significant contribution to achieve this goal. The idea is to embed a hidden reference signal – the watermark – in the original media. This signal will be subject to the same degradation (e.g. compression, channel errors) applied to the watermark media.

This document intends to be an overview of existing objective quality evaluation techniques based in watermarking technology.

2. Objective metrics

2.1. System overview

Objective quality metrics can be classified according to the amount of side information required to compute a given quality measurement. Using this criterion, three generic classes of objective metrics can be described:

- ***Full reference metrics* (FR)** – the evaluation system has access to the original media. Typical metrics within this class are the PSNR (Peak Signal-to-Noise Ratio) and the MSE (mean square error), due to simplicity of their computation.

- ***Reduced reference metrics* (RR)** – the evaluation system has access to a small amount of side information regarding the original media, i.e. features or descriptors extracted from the original.

- ***No-reference metrics* (NR)** – the evaluation system has no reference to any side information regarding the original media. This kind of metrics is the most promising in the context of video broadcast scenario, since the original images or video are in practice not accessible to end users.

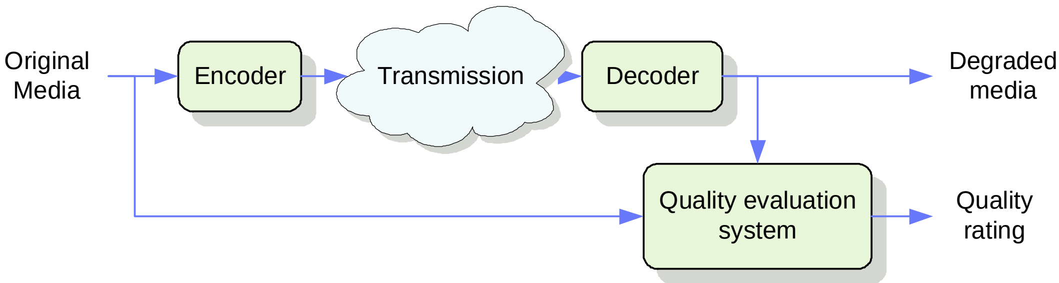

Figure 1 – Full reference metrics system.

Figure 1 depicts a general structure of a full-reference quality evaluation system embedded in a video distribution system. As can be observed from the figure, the original (reference) media is required at the receiver side. Full-reference metrics can also be sub-divided according to the computation required to attain the metric: *simple* or *complex* objective metrics.

Simple objective metrics are attractive because they are computed in a fast way, while attaining a minimally feasible metric for the fidelity of images and video. Probably, the most relevant example of a simple objective metric is the PSNR, which is widely used to perform a fast (and simple) quality evaluation. When applied to images or video, the PSNR can be defined as:

$$\mathrm{PSNR} = 10\log_{10}\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2} = 10\log_{10}\frac{L^2}{\mathrm{MSE}},$$

where X_i and Y_i are the i^{th} pixels of the reference and distorted media, respectively. N is the total number of pixels under analysis and L is the maximum value possible in each pixel.

As can be seen, the PSNR is easily obtained and also has a clear mathematical meaning, which can be used for optimization purposes.

However, simple objective metrics like PSNR (or MSE) have been criticized because they don't correlate well with perceived quality metrics. The main reasons for are:

- Two images with the same PSNR (or MSE) values can have different quality scores.
- The sensitivity of HVS to errors is different for different types of errors, and may also vary with visual content. PSNR treats all errors equally, regardless of their types.

Due to these reasons, a lot of effort has been made to develop objective quality metrics that incorporate perceptual characteristics of the human visual system. These metrics are also designated as complex objective metrics.

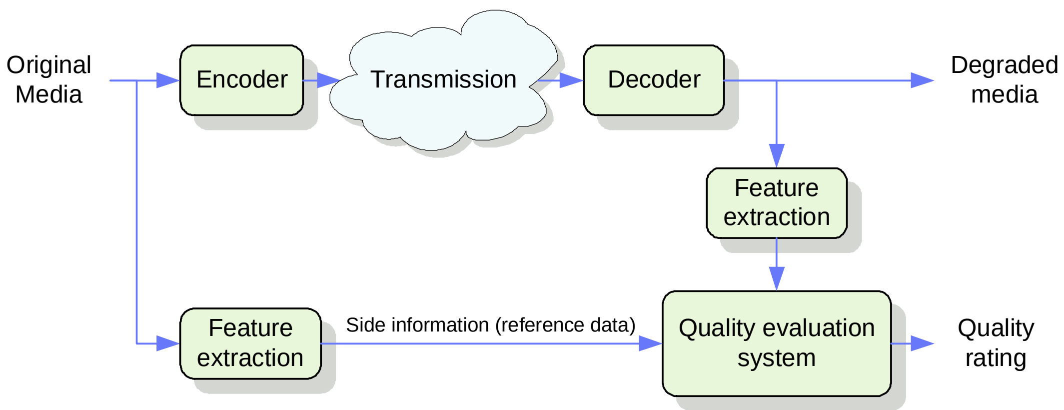

Figure 2 – Reduced reference metrics.

In a reduced reference metrics scenario, the content provider transmits additional information together with the video. This class of metrics requires additional bandwidth (or an additional channel) to transmit the side information. Generically speaking, side information usually consists of relevant features extracted from the original media which are transmitted and compared with the analogous features extracted from the degraded media. The amount of additional information that is transmitted through the side channel is highly dependent of the design of the system. The general architecture of the system is depicted in figure 2.

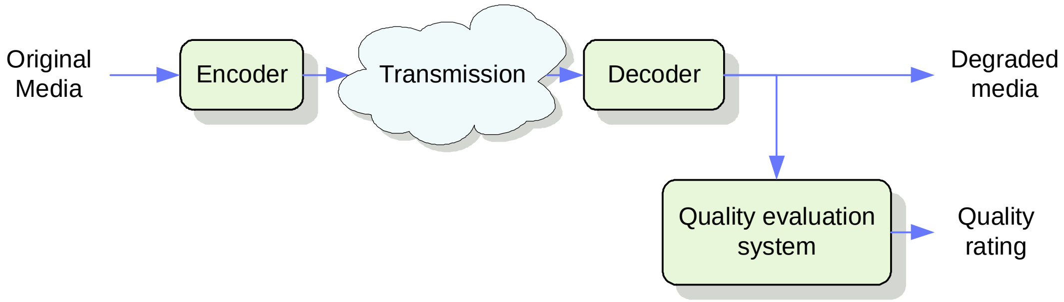

Figure 3 – No-reference metrics.

As an example, Webster et al. [20] propose a reduced reference system in which the side information consists of two distinct types of measurements: spatial measurements extracted from the frames edges, and temporal measurements extracted from frames differences. This work is extended by Wolf et al. [21], where edge activity is further analyzed.

In the no-reference objective metrics scenario (figure 3), quality rating is attained through analysis of the received media only. No-reference objective metrics are relatively rare in literature, but some proposals have been made. Generally, the proposed algorithms evaluate some specific quality features that result from image or video transmission, like block effect in block-based DCT compression methods, edge discontinuity, etc. This kind of analysis is possible by taking into account both human visual system models and natural image models.

2.2. Applications and requirements

Reduced and no-reference metrics could potentially development of useful new applications such as [12]:

- **Branding protection** – content providers should be able to verify that end users are receiving multimedia content with adequate quality.

- **Scalable billing** – users billing should be proportional to the perceived quality of the contents received. In order to introduce more fairness in the multimedia delivering system, users that receive poor quality media data should pay less than users that receive media with higher quality.

- **Quality-based real-time adaptation of streaming services** – the streaming server could automatically adjust some transmission parameters (such as bandwidth or error correction coding) in order to deliver content with an adequate perceived quality, while optimizing resource usage.

- **Quality-aware transcoding** – automated quality monitoring could be used to ensure that resulting quality is bounded by certain criteria during encoding or transcoding of multimedia data.

In order to enable the described applications, an automated quality evaluation system should be able to [12]:

- **Compute a global distortion metric** – the system should be able to evaluate the extent of the global distortion due to channel noise or encoding / transcoding. The measurement should, at least, correlate well with simple distortion metrics (e.g. PSNR), but should preferentially be computed by weighting the characteristics of the human visual system, in order to achieve greater reliability.

- **Compute localized distortion measurements** – since distortion is probably not homogeneous throughout the media, it is important that the system should also be able to identify regions where distortion exhibits abnormal values (e.g. data loss, "block" effect due to MPEG/JPEG compression, "ringing" effect due to JPEG2000 compression).

- **Compute the metrics with reduced or no reference to the original** – the applications described assume that the evaluation can be performed in the user side, and it is assumed that access to original (reference) media is not viable.

3. Watermarking and quality evaluation

3.1. General architecture

Quality assessment of images and video based on watermarking techniques can be considered a mixture between reduced reference methods and no-reference methods.

When compared with reduced reference, the watermark can be considered a reduced reference to the original and the side information channel is the host image or video itself. It can be possible for the watermark to carry additional information regarding features extracted from the reference media, but that is not the usual approach, which constitutes the main difference from reduced reference methods. The watermark usually consists of a binary signal that is subject to the same distortion as the host. Based on distortion measurements of the watermark, it is possible to estimate the distortion in the host.

Watermarking techniques also have some aspects more similar to no-reference metrics – algorithms could be designed in order to achieve independency of the host signal so any information concerning the host signal may be absent. The watermark also dispenses the use of additional bandwidth or channels in server-user downlink, which is also one of the main characteristics of no-reference metrics.

Although research in this area is in the beginning, some watermarking algorithms have already been proposed in the literature. In general, watermarking for quality assessment of video transmission can be performed according to figure 4.

- Embed the watermark in the original media – different methods have been proposed, but the majority of the algorithms use spread-spectrum [11, 15] or quantization based approaches [6, 7].

- Transmission of the watermarked media – this is the part of the communication process where the watermarked media is subject to distortion. Distortion can be caused by a variety of factors, including encoding transcoding of media, data loss, gaussian noise, etc.

- Extraction of the watermark – the watermark signal, or the watermark information bits, is extracted from the corrupted media.

- Comparison between the extracted mark and the reference mark – it is assumed that a reference watermark signal can be generated in the receiving side. The extracted mark is compared with the reference mark and a metric is generated.

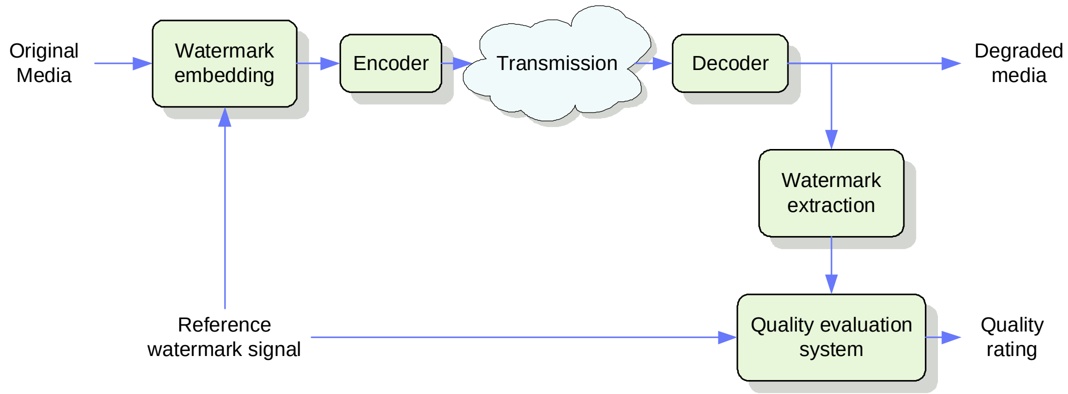

Figure 4 – Watermarking-based reference metrics.

As for the metrics extracted from the watermark signal to quantify media quality degradation, different proposals have been made:

- **Watermark error rate** – probably the most simple but also the most inaccurate metric. This is the metric proposed by Wang et al. [18] and by Zheng et al. [22]. As expected, it is shown that watermark error rate increases with increasing distortion, but there is lack of quantification for the degradation.

- **Watermark signal MSE** – the mean square error between extracted and reference watermark signals provides greater accuracy than the extracted mark error rate, since it exhibits better correlation with the host video MSE. This kind of measurements (or similar) are proposed by the majority of the authors: Campisi et al. [3,4,5], Farias et al. [9,10], Saviotti et al. [16] and Sugimoto et al. [17].

- **Watermark correlation** – another proposed alternative is to compute the correlation between extracted watermark signal and the reference watermark signal. This is the metric used in the work of Bossi et al. [2] and also one the metrics under analysis in the paper by Holliman et al. [12].

All these metrics make an explicit use of the watermark signal, i.e., quality rating of the received media is estimated directly from the watermark degradation. However, the nature of the problem also suggests an implicit use of the watermark signal – the watermark could carry information to be used as side information for the quality evaluation system, following a close approach to reduced reference objective metrics, with the advantage of not requiring additional bandwidth or extra channels to carry side information.

Probably the only proposal in literature that follows this approach is the work by Holliman et al. [12]. In this paper, it is suggested the use of a watermark that comprises information regarding the maximum distortion allowed at each image point. This is accomplished by using a set of quantizers (the quantization step can be weighted according to characteristics of the HVS), whose results are used to generate the watermark. During the extraction phase, out-bounded local distortions will originate errors in the watermark, thus distortion can be measured. This method, however, requires an extremely robust watermark in order to attain feasible results under extreme distortion (low bit rates, high noise, etc.).

3.2. Some Proposed algorithms

Probably, one of the first watermarking techniques used to provide quality evaluation of transmitted MPEG video is the work proposed **Sugimoto et al.** [17]. The authors do not explicitly use the word *watermark*, using the word *markers* instead (which indicates that they are not very familiar with watermarking terminology).

The embedding method consists of adding a pseudo-random binary sequence transformed to the frequency domain (fourier transform) of the host signal. The embedding is block-based and 2 bits are embedded in each block through the use of quantization-based embedding method.

It is interesting to note that watermark embedding strength results from the relation between the host signal power and some reference host power (it seems that the goal is to make the evaluation scheme independent of the host signal).

Quality in the receiving side is estimated from the detection error rate attained in the extracted mark. The authors claim that the watermark detection error rate shows a good correlation with the received media RSNR, which is given by:

$$\mathrm{RSNR} = 20\log\sqrt{\frac{\overline{\sigma_{mb}^2}}{\mathrm{MSE}}},$$

where $\overline{\sigma_{mb}^2}$ represents de mean value of the variances of each macroblock.

Maybe this scheme could be improved if a better watermarking distortion metric was used.

Campisi et al. [3,4,5] propose a spread-spectrum based watermarking scheme to evaluate the quality of compressed video. The watermark is embedded in the middle-high frequency coefficients of the 8x8 block based DCT-transform. No perceptual model is used – there is only a parameter that regulates embedding strength, whose value is obtained empirically in order to preserve imperceptibility of the watermark.

The evaluation metric is attained in extraction by computing the total mean-square error between reference and received watermark (after watermark dispreading). The authors illustrate the effectiveness of their approach with several results. In [5] it is also suggested interesting possibilities to implement client-server feedback strategies, for scaling bill purposes over mobile networks.

Farias et al. [9,10] propose the use of a watermark that consists of a binary sequence (can be a predefined pattern or a binary logo). During embedding, this sequence is multiplied by a pseudo-random sequence with values in [-1;1], and the result is added the DCT domain of the image.

The metrics used to evaluate quality are related to the square error (MSE and TSE) resulting from the difference between the extracted watermark signal and the reference mark signal (which is assumed to be known in the reception).

In [9] the authors also organized a subjective evaluation of video sequences in order to validate the results and to demonstrate the effectiveness of the proposed algorithm as a good quality evaluation scheme. The algorithm's results were compared with the attained in the subjective experiments. The authors claim that the results are analogous.

In [10] the authors assume that motion is a key factor for perceived quality at a human observer. The watermark is also embedded in the DCT but only in blocks where motion is present – the remaining blocks are ignored. The metric used in quality evaluation is the total mean square error (TSE) between the extracted mark and the reference mark. The authors show good correlation between TSE and PSNR of a given video sequence. The main problem is that results are clearly dependent of the video sequence used in the tests. All plots follow similar evolutions but exhibit different numerical values, which can be a serious drawback.

The use of a semi-fragile watermarking technique is proposed by **Saviotti et al.** [16] in order to evaluate video quality in a digital TV environment. The watermark is embedded frame-by-frame in the block-based DCT domain (8x8) using a quantization technique. The watermark consists of a symbol sequence (each symbol corresponds to 2 bits) with values belonging to {0, 1, 2, 3}. Only one coefficient of the original media is modified per DCT block.

Assuming that the watermark is known in the receiver side, the quality of the compressed video is estimated by extracting the watermark, computing its PSNR (it is assumed and using a linear transform to get the estimated PSNR for the corrupted video i.e.,

$$PSNR_{video} = a \cdot PSNR_{watermark} + b,$$

where a and b are empirical parameters, computed from tests using different video sequences compressed at different rates. It remain unclear that these parameters are independent of the host signal.

The main drawback of this paper is the fact that the authors do not show exactly how watermark PSNR is computed.

Holliman et al. [12] analyze different watermarking approaches (i.e., spread spectrum and quantization) to the quality evaluation of video. They conclude that following a quantization approach, the watermark degrades too fast (cliff-like), which introduces difficulties in estimating video quality when degradation is high. On the other hand, with the use of spread spectrum approaches, some issues arise when transcoding multimedia streams.

The authors propose a hybrid approach where the watermark is embedded/extracted following a spread spectrum approach (in order to achieve greater robustness) but the watermark to embed consists of a pseudo-random sequence whose generation key depends on the quantization of image samples (each image location could have a different quantizer, adjusted in order to weight perceptual errors). In this way, it is expected that watermark errors will occur in locations where distortion is greater than the largest quantization step used to generate the mark signal in that location.

The results are compared with the ones attained by using Watson's metric for perceptual error in JPEG compression, showing a good relation.

Another algorithm that provides estimation for the quality of MPEG compressed video is proposed by **Bossi et al.** [2]. The algorithm uses a reference semi-fragile watermark that is subject to the same degradation that the originally watermarked video. The measured correlation between original and extracted watermarks is used to estimate video quality.

The watermark is embedded in the block based DCT domain of the original video (only in the luminance component and only in the med-high frequencies of each DCT block). The watermark consists of a pseudo-random sequence that depends on a parameter β, which results from the difference between maximum theoretical correlation (C_{max}) between watermark and watermarked frame and correlation empirically obtain (the authors don't explain how it is computed).

The correlation value used as quality measure (C_{ext}) is computed from the correlation between the watermark and received frame. This correlation value is then compensated by considering parameter β and C_{max}. The resulting value (C_f) can then be used to estimate the PSNR of the video sequence. C_f defined as:

$$C_f = C_{ext} + \left(1 - \frac{C_{ext}}{C_{max} - \beta}\right) \cdot \beta$$

The English language quality of this paper is very poor. There are also some obscure parts, where the authors do not clearly show what they have done – and this affects very important parts of the paper, namely the computing of the important parameter β for watermark embedding.

4. Conclusions and future directions

After literature analysis, some conclusions and future directions can be proposed.

Some conclusions can be made regarding the requirements of the watermarking algorithm – the following requirements should be met:

- **Robustness** – The response of the watermark detector should decrease with increasing distortion. This goal suggests the use of robust or semi-fragile watermarks whose degradation follows the host signal degradation.

- **Localization** – the algorithm should be able to detect the location of perceptual errors. This location should be spatial and temporal – consequently, the watermark should be computed in independent regions of the host signal. This requirement is important because perceptual errors are probably not perceived equally throughout the whole video sequence.

- **Blindness** – As the original media is absent, oblivious detection of the watermark is imperative.

- **Scalability** – Resolution and granularity of the transmitted video can vary due to different types of terminal equipment (mobile phones displays, computer monitors, TV, etc.) – due to these reason, the watermark should ideally be fully scalable. Some authors consider that geometric transforms (spatial scalability) can be compensated prior to watermark detection [12].

- **Security** – The algorithm should not be constraint due to security reasons. Since the watermarking system works for clients benefit, there is no real interest in remove or forge a watermark. Therefore, there is no need of cryptography, and a simple key management system for embedding / extraction is sufficient. Since security is not an important issue, all the algorithms found in literature neglect it.

- **Watermark domain** – The choice of watermarking domain should be made in order to produce computationally simple results. Since compressed video is generally transmitted in MPEG-2 format, the obvious choice for watermarking domain is the block-based 8x8 DCT domain, as proposed by the majority of the authors [2,3,4,5,9,10,16,17,22]. It can also be found in literature a proposed algorithm that works in the DWT domain [18]. If the system is real–time constrained, probably the best choice for the watermark domain is the 8x8 block-based DCT domain. If real–time is not required, any domain can be used.

An issue that should also be analyzed is which class of watermarking algorithms performs better to achieve the quality-rating goal. Two main classes are generally proposed in literature: spread-spectrum and quantization based watermarking. Concerning this topic, in [12] the author noticed that when using quantization based watermarking algorithms, the watermark degrades too fast with increasing media distortion (in a cliff-like fashion), making it difficult to establish a relation between watermark distortion and media distortion. This is the main reason why spread-spectrum based watermarking schemes are probably the most adequate.

Another question that arises from literature analysis is how to make independent of the host signal the metric that results from watermark distortion, i.e., how to make the scheme a good quality estimator regardless the input video sequence to evaluate? This is a question that can be a topic for future work.

Bibliography

1. Mauro Barni, Franco Bartolini, Alessandro Piva, *Improved wavelet-based watermarking throughpixel-wise masking*, IEEE Transactions on Image Processing, Vol. 10, N. 5, May 2001.

2. Stefano Bossi, Francesco Mapelli, Rosa Lancini, *Objective video quality evaluation by using semi-fragile watermarking*, Proc. of Picture Coding Symposium 2004, S. Francisco, USA, December 2004.

3. Patrizio Campisi, Marco Carli, Gaetano Giunta, Alessandro Neri, *Tracing watermarking for multimedia communication quality assessment*, Proc. of IEEE International Conference on Communications, New York, USA, May 2002.

4. Patrizio Campisi, Gaetano Giunta, Alessandro Neri, *Object-based quality of service assessment using fragile watermarking in MPEG-4 video cellular services*, Proc. of IEEE International Conference on Image Processing, Rochester, New York, USA, Vol. II pp. 881-884, September 2002.

5. Patrizio Campisi, Marco Carli, Gaetano Giunta, Alessandro Neri, *Blind quality assessment system for multimedia communications using tracing watermarking*, IEEE Transactions on Signal Processing, Vol. 51, N. 4, April 2003.

6. Brian Chen, Gregory W. Wornell, *Dither modulation: a new approach to digital watermarking and information embedding*, Proc. of SPIE Security and Watermarking of Multimedia Contents, S. Jose, EUA, January 1999.

7. Brian Chen, Gregory W. Wornell, *Quantization index modulation: a class of provably good methods for digital watermarking and information embedding*, IEEE Transactions on Information Theory, Vol. 47, N. 4, May 2001.

8. Max H. M. Costa, *Writing on dirty paper*, IEEE Transactions on Information Theory, Vol. 29, N. 3, May 1983.

9. Mylène C. Q. Farias, Marco Carli, *Video objective metric using data hiding*, Proc. of IEEE 5th Workshop on Multimedia Signal Processing, St. Thomas, USA, December 2002.

10. Mylène C. Q Farias, Marco Carli, Alessandro Neri, and Sanjit K. Mitra, *Video quality assessment based on data hiding driven by optical flow information*, Proc. of SPIE Image Quality and System Performance, Vol. 5294, San Jose , January, 2004.

11. Frank Hartung, Bernd Girod, *Watermarking of uncompressed and compressed video*, Signal Processing, Vol. 66, N. 3, May 1998.

12. Matthew Holliman, Minerva M. Yeung, *Watermarking for automatic quality monitoring*, Proc. of SPIE Security and Watermarking of Multimedia Contents IV, S. Jose, EUA, January 2002.

13. Deepa Kundur, Dimitrios Hatzinakos, *Digital Watermarking for Telltale Tamper Proofing and Authentication*, Proceedings. of the IEEE, Vol. 87, N. 7, July 1999.

14. Rainer Lienhart, Igor Kozintsev, Yen-Kuan Chen, Matthew Holliman, Minerva Yeung, Andre Zaccarin and Rohit Puri, *Challenges in distributed video management and delivery*, Chapter 1 in *The handbook of video databases, design and applications*, CRC Press, 2003.

15. Christine I. Podilchuk, Wenjun Zeng, *Image-adaptive watermarking using visual models*, IEEE Journal on Selected Areas in Communications, Vol. 16, N. 4, May 1998.

16. Sergio Saviotti, Francesco Mapelli, Rosa Lancini, *Video quality analysis using a watermarking technique*, Proc. of WIAMIS2004, Lisbon, Portugal, April 2004.

17. Osamu Sugimoto, Ryoichi Kawada, Masahiro Wada, Shuichi Matsumoto, *Objective measurement scheme for perceived picture quality degradation caused by MPEG encoding without any reference pictures*, Proc. of SPIE Visual Communications and Image Processing, Vol. 4310, pp 932-939, 2001.

18. Sha Wang, Jiying Zhao, Wa James Tam, Filippo Speranza, *Image quality measurement by using watermarking based on discrete wavelet transform*, 22nd Biennial Symposium on Communications, Kingston, Ontario, Canada, May-June 2004.

19. Zhou Wang, Hamid R. Sheikh, Alan C. Bovik**,** *Objective video quality* – chapter 41 in *The handbook of video databases, design and applications*, CRC Press, 2003.

20. A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, *An objective video quality assessment system based on human perception Proc. SPIE*, vol. 1913, pp. 15-26, 1993.

21. S. Wolf and M. H. Pinson, Spatio-temporal distortion metrics for inservice quality monitoring of any digital video system, *Proc. SPIE*, vol. 3845, pp. 266-277, 1999.

22. Dong Zheng, Jiying Zhao, Wa James Tam, and Filippo Speranza, *Image quality measurement by using digital watermarking*, IEEE International Workshop on Haptic, Audio and Visual Environments and their Applications, Ottawa, Ontario, Canada, pp. 65-70, September 2003.