Objective Quality Assessment of MPEG-2 Video Streams by Using CBP Neural Networks

Paolo Gastaldo, Stefano Rovetta, and Rodolfo Zunino, *Member, IEEE*

***Abstract*—The increasing use of compression standards in broadcasting digital TV has raised the need for established criteria to measure perceived quality. Novel methods must take into account the specific artifacts introduced by digital compression techniques. This paper presents a methodology using circular backpropagation (CBP) neural networks for the objective quality assessment of motion picture expert group (MPEG) video streams. Objective features are continuously extracted from compressed video streams on a frame-by-frame basis; they feed the CBP network estimating the corresponding perceived quality. The resulting adaptive modeling of subjective perception supports a real-time system for monitoring displayed video quality. The overall system mimics perception but does not require an analytical model of the underlying physical phenomenon. The ability to process compressed video streams represents a crucial advantage over existing approaches, as avoiding the decoding process greatly enhances the system's real-time performance. Experimental evidence confirmed the approach validity. The system was tested on real test videos; they included different contents ranging from fiction to sport. The neural model provided a satisfactory, continuous-time approximation for actual scoring curves, which was validated statistically in terms of confidence analysis. As expected, videos with slow-varying contents such as fiction featured the best performances.**

***Index Terms*—Circular backpropagation (CBP), compressed video, digital TV, objective quality assessment.**

I. Introduction

THE recent increasing success of digital TV has stimulated the research for objective automated methods to assess the user-end perception of broadcasting. The underlying technical problem is to estimate the effects of the visual artifacts brought about by digital encoding. In this sense, traditional techniques for analog data processing often prove ineffective in measuring the perceived quality of a digital compressed video.

Up to now subjective measurements [1] have been a fundamental instrument to characterize video quality, despite their complexity and variability of results. Subjective assessment methods attempt to evaluate the perceived quality by asking human assessors to score the quality of a series of test scenes. Objective quality assessment aims to emulate human response to perceived quality by extracting numerical quantities from video streams. As a result, this technique no longer requires inputs from human operators, as opposed to conventional subjective tests. The need for objective quality measures in digital TV has a commercial rationale, too, as quality may bias a customer's choices of advanced pay-on-demand services. In addition, the number of coders on the market will increase in the next years, hence both manufacturers and broadcasters will invariably face the problem of comparing the user-level quality of video.

A variety of methods for objective quality assessment of digital TV have been proposed in the literature [2], [3]. "No reference" approaches to objective assessment aim to estimate perceived quality by processing data extracted from video streams only. By contrast, "full reference" or "reduced reference" approaches involve both the encoded signal and the video source in the evaluation. Most methods are based on decompressed video: objective parameters are worked out by comparing pictures at the receiver end with original scenes. The comparison is made either in the feature space or in the picture domain by using differencing methods [4].

An attempt to relate objective measures to subjective assessments is described in [5]–[7], where linear mathematical models single out clusters of objective features that best fit subjective assessment results. Other approaches aim to emulate human perception explicitly: perceptual models process objective parameters from image segmentation [8]; in a structured approach, objective assessments stem from a three-layered picture structure (object, texture, and noise layers) supporting the human visual process [9]. Metric-based approaches to the emulation of human perception measure spatio-temporal distortion [10] as well as blurring and blockiness in decoded pictures [11]. A method that does not involve the original video is described in [12], where an algorithm extracts data from decoded frames to detect blockiness artifacts.

Most of the above papers implied some *a priori* simplifying hypotheses about the underlying mathematical model, which somehow affected the practical validity of most results. From a scientific perspective, those research works approached the problem of human perception of quality as a modeling one. A neural-based approach to motion picture expert group (MPEG) quality evaluation is described in [13]; that method operated at a granular level and employed conventional multilayer perceptrons (MLPs) [14] for a "full reference" evaluation schema.

As compared with those approaches, this work presents a method using the circular backpropagation (CBP) neural network (NN) [15] for automatic evaluation of subjective assessment in a "no reference" environment. The network operates

Manuscript received April 12, 20001; revised October 29, 2001.

P. Gastaldo and R. Zunino are with the Department of Biophysical and Electronic Engineering—DIBE, University of Genoa, 16145 Genova, Italy (e-mail: gastaldo@ dibe.unige.it, zunino@ dibe.unige.it).

S. Rovetta is with the Department of Computer and Information Sciences—DISI, University of Genoa, 16146 Genova, Italy, and INFM (e-mail: ste@disi.unige.it).

Publisher Item Identifier S 1045-9227(02)04430-2.

on compressed data only; this removes the need for any information about either the original video or the decoding process. From an engineering standpoint, the adaptive neural framework decouples the evaluation task from the specific video source and from decoder issues as well.

The present approach partly disregards the objective of gaining a deeper insight into some aspects of quality perception. Rather, the aim is to produce a method to mimic such perception. As an immediate consequence, many simplifying assumptions, useful to enable one to understand the perception mechanism, are discarded in that the resulting model is not sufficiently powerful. This in turn requires that a potentially complex mathematical model be used.

Section II briefly summarizes the neural model (CBP) adopted, highlighting the advantages in using this network for the specific multimedia application. Section III describes the neural-based system for video-quality evaluation, showing the criteria driving feature selection, the experimental setup, and neural training. Section IV reports on experimental results, demonstrating the method operation under different conditions and for different input sources. Some concluding remarks are made in Section V.

II. Feedforward CBP Architectures

Feedforward NNs provide a straightforward paradigm to map feature vectors (describing video frames) into the corresponding quality assessments. Such a problem setting treats the quality scorings used for training as an ordered discrete set of labels, whereas any intermediate values in the associate network output are allowed. In this sense, efficiency requirements as well as generalization issues ultimately lead to the problem of properly sizing the number of neurons in the NN.

MLPs can efficiently tackle problems in which the target-mapping function can be supported by few units with global scope; in MLPs, those elements are encoded by the sigmoid functions within hidden units. Conversely, if the target mapping can be best expressed as a superposition of locally tuned components, radial basis function (RBF) networks will typically perform much more efficiently. As a result, the unknown characteristics of the problem-related target mapping further complicate the problem of selecting the nature and the number of hidden units.

A solution to this specific problem has been proposed in [15]. The CBP network extends the multilayer perceptron by including one additional input with its associated weight. Such an input just sums the squared values of all the other network inputs. As proved by CBP theory, the additional unit allows the overall network to adopt the standard, sigmoidal behavior, or to drift smoothly to a bell-shaped radial function, which approximates—but is not—a Gaussian. At the same time, the limited increase in the network parameters does not affect its expected generalization performance, as it has been proved that the Vapnik–Chervonenkis dimension (VC-dim) [16] of the augmented circular perceptron increases by one unit [15].

The CBP model adopted for this research can be formally described as follows. An MLP architecture combines two functional layers (Fig. 1) including n_h and n_o units, respectively.

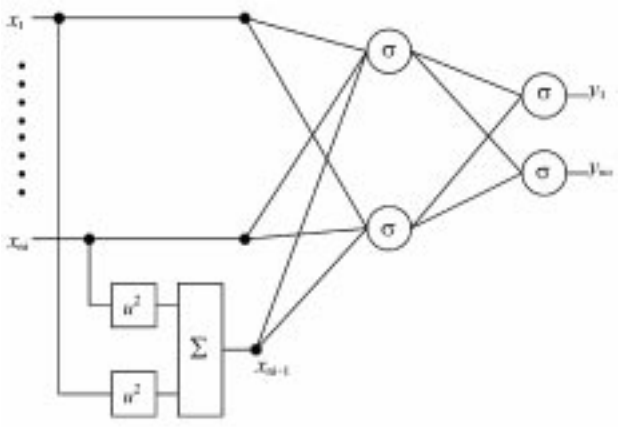


Fig. 1. The CBP model includes one additional input to the standard MLP.

The conventional sigmoidal function is denoted by $\sigma(x) = (1+e^{-x})^{-1}$.

The input layer connects the n_i input values to each unit of the hidden layer. The jth "hidden" neuron performs the following transformations on the input values:

$$r_j = w_{j,0} + \sum_{i=1}^{n_i} w_{j,i} x_i + w_{j,n_i+1} \sum_{i=1}^{n_i} x_i^2, \quad j = 1, \ldots, n_h \quad (1)$$

where $a_j = \sigma(r_j)$. The input features $x_i (i = 1, \ldots, n_i)$ combine with the associated weights $w_{ji} (i = 0, \ldots, n_i + 1)$ and feed the jth hidden unit. The terms r_j and a_j denote the neuron *stimulus* and *activation*, respectively. The last term in expression (1) actually augments the conventional MLP up to the CBP model.

The *output* layer provides the actual network responses, y_k, by the following transformations:

$$r_k = w_{k,0} + \sum_{j=1}^{n_h} w_{k,j} a_j; \quad y_k = \sigma(r_k); \quad k = 1, \ldots, n_o. \quad (2)$$

Theory proves [15] that this model is the most efficient polynomial extension of MLPs with linear stimulus, and formally encompasses the RBF network model as well. The strict relationship of CBP to vector quantization (VQ) networks has been analyzed in [17], showing that the model ensures a notable representation effectiveness with a very small increase in the number of parameters. Previous experimental verifications on real testbeds always confirmed that such theoretical properties actually witness a satisfactory practical effectiveness.

The crucial feature that makes the CBP model suitable for the video quality-assessment task is its ability to switch autonomously between the different representation paradigms (MLP or RBF), as conventional backpropagation algorithms [18] can be adopted for weight adjustment. The resulting weight configuration ultimately sets the most suitable representation setting for the mapping problem, and is only driven by training data, independently of any *a priori* assumption on the observed domain.

III. Neural–Network–Based Assessment of Video Quality

Avoiding inputs from human subjects can lead to deterministic models, yet objective systems should keep human scores as references to ensure consistency with subjective results. The present approach applies CBP feedforward networks to the automated quality evaluation of MPEG-2 [19] video streams; the single-ended no-reference paradigm need not know uncompressed original videos.

Fig. 2 shows a schematic representation of the overall system. Objective features are worked out directly from MPEG-2 bitstreams (i.e., without any decoding), and feed the NN to obtain quality ratings. The system operates on a frame-by-frame basis and yields a continuous output; as such, it provides a real-time monitoring tool for displayed video quality. Thus, the NN is entrusted to mimic the subjective, single-stimulus continuous quality evaluation (SSCQE) method [20], recording continuous assessments of picture quality provided by human observers.

The crucial advantage of the approach lies in generating quality ratings without decoding the video stream. Indeed, the objective metric supported by the neural system relies entirely on a representation format—the compressed bitstream—that bypasses the need for human assessors' rating process altogether. This greatly improves the method's effectiveness especially in terms of real-time performance, as one can get an estimate of perceived quality at transmission time.

For the reader's convenience, we recall that MPEG-2 attains still-image quality by standard discrete cosine transform (DCT) compression; motion information is treated by dividing each frame (picture) into several macroblocks (holding 16×16 pixels each), and by encoding the apparent movement of macroblocks within time-consecutive frames.

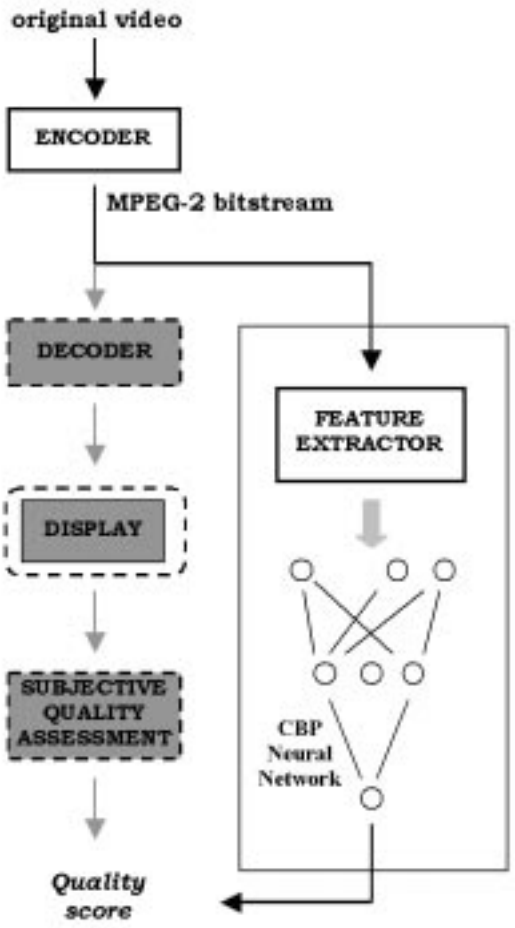


Fig. 2. The proposed single-ended system for automated quality assessment.

A. Features for Objective Quality Assessment

The set of processed features play a crucial role for the effectiveness of the overall methodology. A single-ended paradigm avoiding *a priori* assumptions requires quite a large set of parameters to be extracted from video streams, for the purpose of collecting as much information as possible. Truly significant features are then sorted out by a conventional statistical analysis. Appendix A lists the objective features worked out from the MPEG-2 compressed stream.

In principle, one expects that a considerable number of all the above features will be discarded, either because they do not carry significant information or because they are mutually correlated. Since the present approach does not imply any *a priori* assumption on the significance of the encoding parameters, an *a posteriori* statistical analysis drives the feature-selection criterion.

First of all, a percentile-basis analysis is required to remove outliers from input data. To this end, for each feature, an elementary preprocessing phase set a 0.05 percentile threshold to each tail of the distribution of empirical values. The specific threshold value was determined empirically, but does not affect the method generality thanks to the quite narrow distribution of measured samples, which appear concentrated around their mean values.

Secondly, the statistical analysis assumes that nonnormally distributed features carry most information. The method adopts third- and fourth-order moments of the distributions of values as normality indicators. *Skewness* (i.e., a measure of the degree of symmetry in a variable distribution) and *kurtosis* (i.e., a measure of the relative *peakedness/flatness* of a distribution) are used to characterize the statistical activity of each feature. Only features having skewness and kurtosis significantly different from normality are considered for the neural-network modeling. A threshold scheme drives that selection process; specific threshold values have been set by averaging over several samples in different contexts. The following basic quantities are defined:

- Ψ is a library $\{\Psi_1, \ldots, \Psi_L\}$ of L test streams, composed of P frames each;
- F_k is the set of objective features $(k = 1, \ldots, N_f)$;
- $F_{k\Psi_i}^{(j)}$ is the value measured by F_k for the jth frame of the ith stream Ψ_i.

The feature-selection algorithm can be outlined as follows:

0. *(Input)*—sets of features, Φ_k, associated with each sequence frame ($k = 1,\ldots,N_f$).

$$\Phi_k = \left\{ f^{(1)}_{k\Psi_1},\ldots,f^{(P)}_{k\Psi_1},\ldots,f^{(1)}_{k\Psi_L},\ldots,f^{(P)}_{k\Psi_L} \right\}. \quad (3)$$

1. *(Rescaling)*
For $k = 1,\ldots,N_f$:
—compute the 0.05 and the 0.95 percentiles, $x^{(k)}_{0.05}$, $x^{(k)}_{0.95}$, respectively, for the values in Φ_k;
—build up a set $\underline{\Phi}_k$ by rescaling each element of Φ_k into the range $[-1,1]$:

$$\underline{\Phi}_k = \left\{ \underline{f}_{kz} \right\}; \quad z = 1,\ldots,P,\ldots,P(L-1)+1,\ldots,P\cdot L \quad (4)$$

where

$$\underline{f}_{kz} = 2\frac{\left(f^{z-Pq}_{k\Psi q} - x^k_{0.05}\right)}{\left(x^k_{0.95} - x^k_{0.05}\right)} - 1, \quad q = \left\lfloor \frac{(P+z-1)}{P} \right\rfloor. \quad (5)$$

Rescaling $f^{(j)}_{k\Psi_i}$ by using $x^{(k)}_{0.05}$ and $x^{(k)}_{0.95}$ as its lower and upper bound, respectively, supports the outlier-removal process previously anticipated.
2. *(Descriptive statistics)*
Create two sets:
– $\Sigma = \{\mathrm{skew}_k\}$ where $\mathrm{skew}_k = \mathrm{skewness}(\underline{\Phi}_k), k = 1,\ldots,N_f$;
– $K = \{\mathrm{kurt}_k\}$ where $\mathrm{kurt}_k = \mathrm{kurtosis}(\underline{\Phi}_k), k = 1,\ldots,N_f$.
3. *(Threshold setting)*
Compute the threshold values, $\mathrm{skew}_{\mathrm{thr}}$ and $\mathrm{kurt}_{\mathrm{thr}}$, as:
– $\mathrm{skew}_{\mathrm{thr}}$ is the 0.5 percentile of Σ;
– $\mathrm{kurt}_{\mathrm{thr}}$ is the 0.5 percentile of K.
4. *(Feature selection)*
Compile the feature set, Z, holding the objective features that satisfy (for $k = 1,\ldots,N_f$):

$$F_k \in Z \Leftrightarrow (\mathrm{skew}_k > \mathrm{skew}_{\mathrm{thr}}) \ \texttt{AND} \ (\mathrm{kurt}_k > \mathrm{kurt}_{\mathrm{thr}}). \quad (6)$$

As a result of the above procedure, the set Z includes the features that, due to their asymmetrical distribution, are unlikely to stem from a Gaussian distribution. The purpose is to single out the statistically significant objective descriptors, under the (practically reasonable) assumption that noninformative quantities most often exhibit a Gaussian distribution.

The described algorithm has been preferred to alternative approaches such as the principal component analysis (PCA) [21] mainly because of the high data dimensionality involved. The complexity of working out eigenvectors due to numerical precision issues may sometimes affect the performance of PCA when applied to huge multidimensional data. Conversely, exploratory projection pursuit (EPP) [22], [23] represents a method that follows the same paradigm of the proposed algorithm. EPP is a powerful methodology to derive a feature set to describe the original data set; it seeks for a coordinate system such that the resulting distribution of values along each axis is as much distant from a Gaussian curve as possible. EPP, on the other hand, is a computation-intensive method, which might prove difficult to tune in nonlinear domains.

B. Feature Run-Time Sampling

The objective assessment system should generate continuous-time quality ratings. In principle, one might feed the CBP network with the feature values continuously extracted from each sequence frame. In fact, the mechanism generating the input features x_i must take into account known mechanisms specific for human perception.

In more detail, one has to consider that: 1) assessor's reaction times are subject to delays [24]–[26]; 2) time-consecutive frames tend to interfere with one another [27], and 3) the most recent segments of a sequence have a greater effect on the overall quality rating [3], [28]. In the literature, such peculiarities are known as "the assessor's response time," "masking phenomenon," and "recency effect," respectively.

The following quantities are used to parameterize these settings (Fig. 3). To compensate for temporal averaging, a set of N frames contribute to generating a single score. Within this set, groups of W consecutive frames yield a single feature vector $\vec{x}$, according to the masking phenomenon. The input vector $\vec{x}$ includes n_i features $\overset{\frown}{f}{}^{(j,W)}_{k\Psi_i}$ $(F_k \in Z)$ defined as follows:

$$\overset{\frown}{f}{}^{(j,W)}_{k\Psi_i} = \rho\left(f^{(j)}_{k\Psi_i} \ldots f^{(j+W-1)}_{k\Psi_i} \right) \quad (7)$$

where $\rho(f_1,\ldots,f_k)$ is a family of operators, with ρ_h, ρ_s and ρ_m, respectively, the highest, the smallest and the mean values over the interval. The parameter Δ refers to the delay between the subjective judgment and the last frame that has influenced it.

C. The Neural-Network Approach

Several features characterizing video streams jointly affect subjective judgments; possibly nonlinear relationships and partly unknown mechanisms may complicate the process modeling. These effects actually seem to have sometimes been underevaluated in the literature, and the major advantage of a neural-network approach lies in the ability to deal with multidimensional data representing complex relationships. By decoupling the feature-selection task from the design of an explicit mathematical model, one obtains the crucial advantage of avoiding *a priori* assumptions on the significance of objective measures.

In the present approach, CBP networks map feature vectors into quality ratings. The mapping function is learned from examples by means of an iterative training algorithm, and a single output neuron in the NN yields the quality assessment for a given input vector. The network configuration (i.e., the number of hidden units) has been designed by using a specific initialization technique that exploits the equivalence of the CBP model to

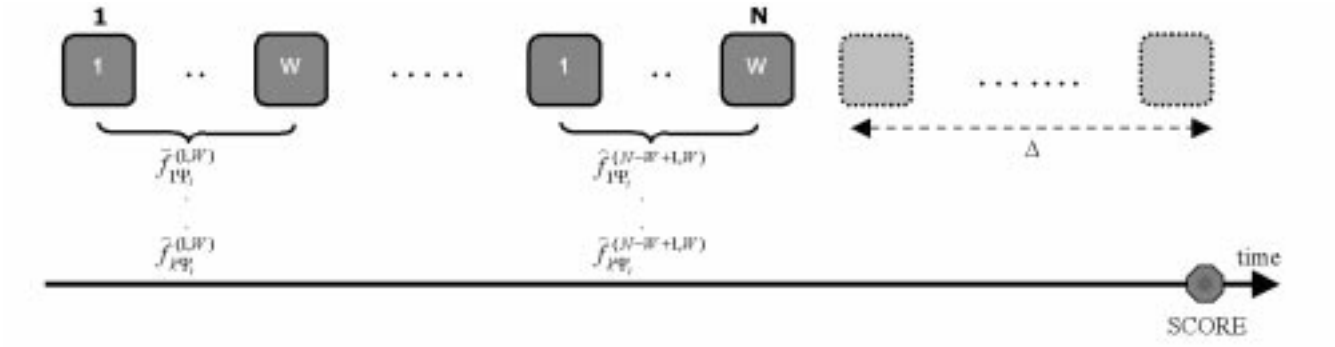


Fig. 3. Feature run-time sampling process according to perceptual mechanism.

VQ paradigms [17]. In particular, a VQ preliminary phase using the plastic neural gas algorithm [29]–[31] made it possible to assess the proper number of prototype vectors to represent the available sample distribution. In the subsequent network setup phase, the number and the space positions of those prototypes were mapped directly into the specific CBP network configuration according to the formalism described in [17]. Thus the initial setting of the network weights proved most effective in accelerating the convergence of the overall training process, as compared with a conventional random setting.

The CBP network training uses an accelerated variant [18] of the classical backpropagation algorithm. The possibility of using conventional techniques to train an advanced network structure is the major advantage of the CBP model. The network cost function is expressed as

$$e(\vec{w}) = \frac{1}{n_o n_p} \sum_{m=1}^{n_p} \sum_{k=1}^{n_o} \left(t_k^{(m)} - y_k^{(m)} \right)^2 \tag{8}$$

where n_p is the number of training patterns and t_k is the actual quality assessment derived experimentally from the human scoring panel. An alternative to (8) is the *threshold cost* function $e_T(\gamma)$:

$$e_T(\gamma) = \frac{1}{n_o n_p} \sum_{m=1}^{n_p} \sum_{k=1}^{n_o} g\left(\left| t_k^{(m)} - y_k^{(m)} \right| \right)$$
$$g(u) = \begin{cases} 0 \Leftrightarrow u \le \gamma \\ 1 \Leftrightarrow u > \gamma \end{cases} \tag{9}$$

where the network cost is expressed as the percentage of outputs y_k that differ from the expected score t_k in more than a fixed threshold γ.

IV. Experimental Results

The effectiveness of the CBP model for objective quality assessment was verified experimentally by using a library of MPEG-2 videos provided by the Research Center of the Italian Radio and Television Corporation (RAI). The testbed included twelve frame-coded sequences, each 70 s long; the picture size was 720 × 576 pixels. The sequence contents varied from fiction to sport and were encoded at different bit rates in the range [4,8] Mbits/s.

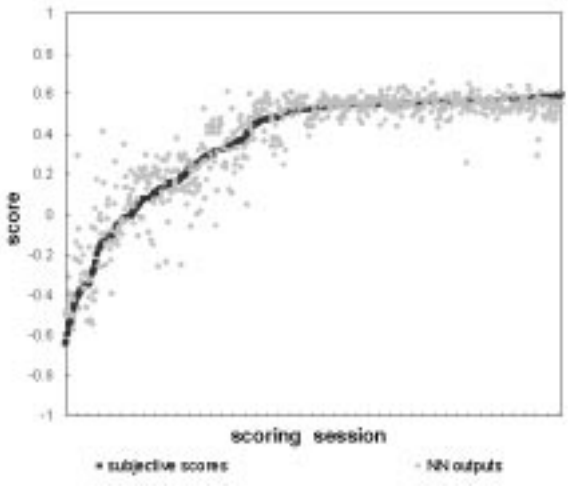
The assessments for each sequence were collected from nonexpert viewers; the subjective tests were performed with an SSCQE technique at a sampling rate of two scores per second. The quality ratings were represented by a continuous scale ranging in [−1, 1].

TABLE I
Test Results

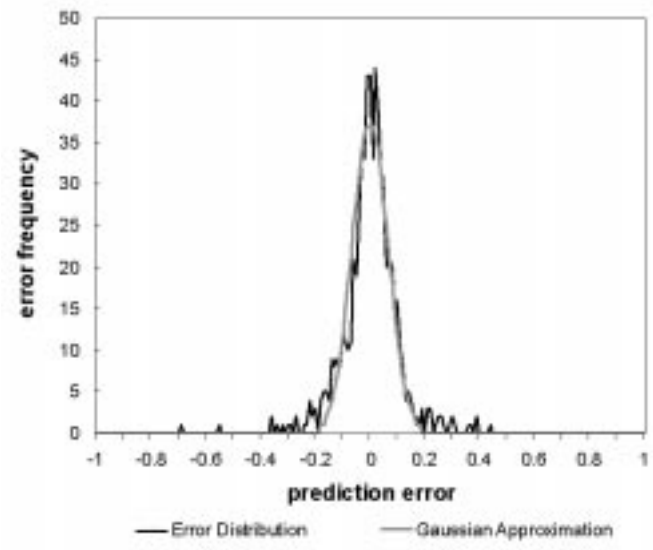
	Test videos		
	complete set	no-sport	sport
$\hat{\mu}_{err}$	0.001	0.0005	0.003
$\hat{\sigma}_{err}$	0.0665	0.045	0.11
e	0.0104	0.0038	0.0251
$e_T(0.15)$	0.1022	0.0308	0.3011

TABLE II
Features Worked Out From MPEG Stream

PERCENTAGES (MACROBLOCKS)	
Pmb_no_pred	n_{mb} = macroblocks with no motion vectors
Pmb_fwd	n_{mb} = macroblocks with forward mv only
Pmb_back	n_{mb} = macroblocks with backward mv only
Pmb_bidir	n_{mb} = macroblocks with both forward and backward mv
Pmb_I	n_{mb} = intra macroblocks
Pmb_skipped	n_{mb} = skipped macroblocks
PERCENTAGES (BLOCKS)	
Pb_sk_luma	n_b = skipped luminance blocks
Pb_sk_chroma	n_b = skipped chrominance blocks
STATISTIC	
Smv_mean	mean - p_i = |motion vector|
Sq_scale_mean	mean – p_i = *q_scale*
Senergy_mean	mean – p_i = *energy*
Smv_dev_std	standard deviation - p_i = |motion vector|
Sq_scale_dev_std	standard deviation – p_i = *q_scale*
Senergy_dev_std	standard deviation – p_i = *energy*
Smv_var	variance - p_i = |motion vector|
Sq_scale_var	variance - p_i = *q_scale*
Senergy_var	variance - p_i = *energy*
PERCENTILES	
Xmv(α)	p_i = mean of |motion vector|
Xq_scale(α)	p_i = *q_scale*
Xenergy(α)	p_i = *energy*
Xq_mv(α)	p_i = *q_mv*
Xe_mv(α)	p_i = *e_mv*

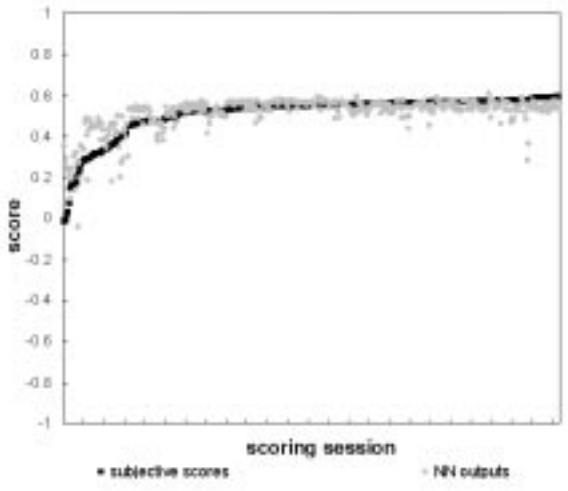


(a)

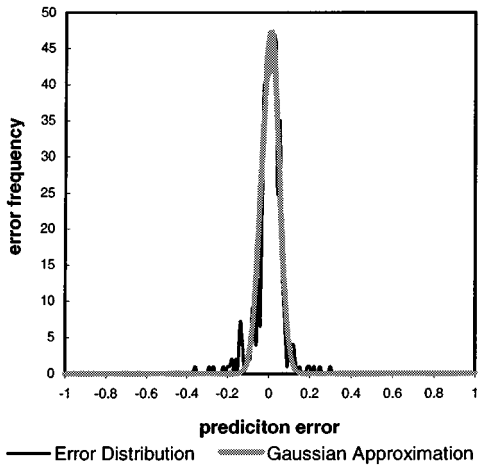


(b)

Fig. 4. Test results obtained with the four-dimensional space covered by the selected features. (a) Neural-network outputs compared with human quality ratings. (b) The associated error distribution.



(a)



(b)

Fig. 5. Test results obtained for videos included in *no-sport*. (a) Neural network outputs compared with human quality ratings. (b) The associated error distribution.

A. Experimental Setup

The neural-network training process involved the set of features Z that the statistical analysis selected from the global feature set F_k listed in Appendix A. The resulting feature space (a subset of $\Re^{24}$) included the quantities highlighted in bold face in Table II. The training patterns were generated by the run-time sampling process presented in Section III, with $N = 24$, $W = 6$ and $\Delta = 17$.

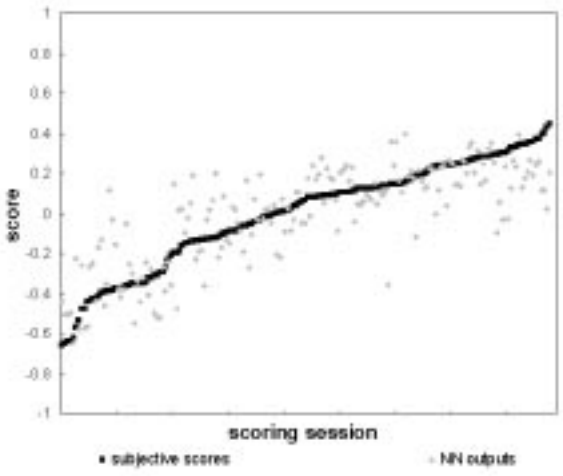
In order to enhance the CBP network generalization performance, the dimensionality of the input data space was further reduced with the feature-selection technique described in [32]. The eventual four-dimensional feature space covered the quantities: $\rho_s(Nn\_\text{bits})$, $\rho_h(Xq\_\text{scale}(1))$, $\rho_h(Xmv(1))$, and $\rho_h(Smv\_\text{dev\_std})$. The plastic VQ algorithm processed the training samples to design the neural-network configuration; the resulting value $n_h = 15$ set the number of hidden units in the feedforward structure.

B. Results

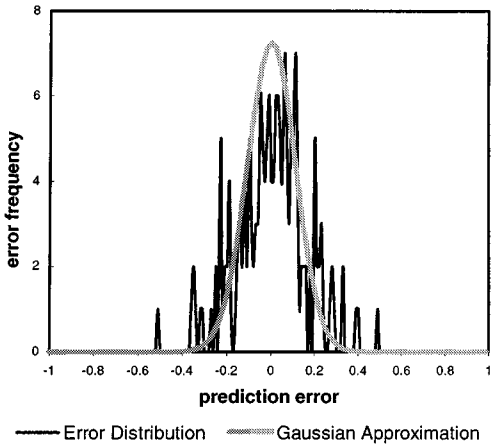
Fig. 4 shows test results obtained for the selected feature set. Fig. 4(a) compares the quality ratings by human assessors with the corresponding outputs of the NN; for display clarity, the actual ratings are sorted in increasing order, each point on the x axis representing a single evaluation event. The graph shows an asymmetric distribution of subjective scores, as 44% of the original scores exceed 0.5. Since the lower scores appear subsampled, they are subject to greater errors due to the lower statistical confidence. Nevertheless, the CPB NN attained an average error $\hat{\mu}_{\text{err}} = 0.001$ on the test set.

Fig. 4(b) plots the error distribution together with the related best-fitting Gaussian approximation ($\hat{\mu} = 0$, $\hat{\sigma} = 0.066$). A chi-square test verifying the correctness of the Gaussian assumption did not detect a satisfactory match, mainly due to the apparent undersampling phenomenon. However, a more robust Kolmogorov-Smirnov (KS) normality test satisfied the null hypothesis to a high degree of confidence ($p > 0.95$).

Figs. 5 and 6 present the results obtained by letting the CBP network evaluate two subsets of the original data set not



(a)



(b)

Fig. 6. Test results obtained for videos included in *sport*. (a) Neural-network outputs compared with human quality ratings. (b) The associated error distribution.

used for training. The two subsets (*"sport"* and *"no-sport"*, respectively) differed in their video contents. A comparison of Fig. 5(a) with Fig. 6(a) points out that human quality ratings show a higher variance for videos including sport contents only; in addition, sequences with sport contents are a small subset (27%) of the test library, hence the NN suffered from larger errors due to the lower statistical confidence.

Figs. 5(b) and 6(b) confirm these achievements by fitting error distributions with the associated best-approximating Gaussians. The Gaussian parameters are $(\hat{\mu} = 0,\ \hat{\sigma} = 0.045)$ for *no-sport* and $(\hat{\mu} = 0,\ \hat{\sigma} = 0.11)$ for *sport*. The overall numerical results are summarized in Table I, also giving the costs c and $c_T(\gamma)$ derived from the neural-network test.

The graph in Fig. 7 plots the estimated confidence interval for the sample average error $\hat{\mu}_{\text{err}}$, and confirms the method effectiveness. Theory states [34] that, for large sample sizes n, the confidence interval for a distribution having expectation μ and variance σ^2 (both unknown) can be computed as

$$P(|\hat{\mu} - \mu| \leq \varepsilon) = 1 - \alpha \tag{10}$$

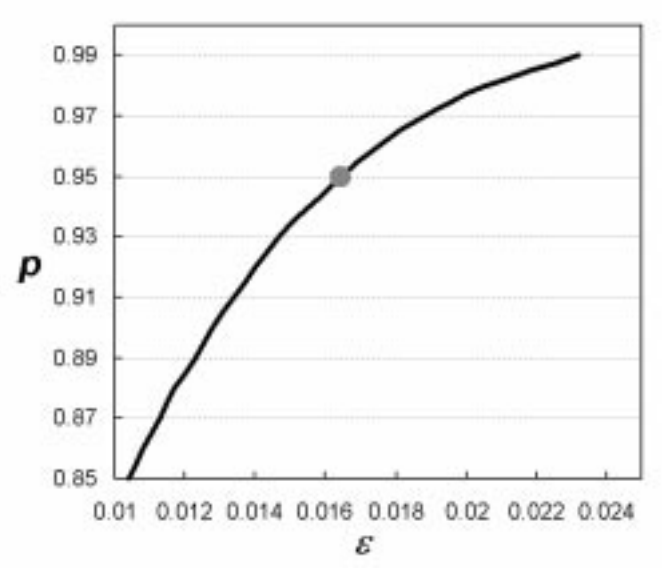


Fig. 7. Plot of ε versus the confidence $p(= 1 - \alpha)$.

where $(1 - \alpha)$ is the confidence level, and ε is defined as

$$\varepsilon = z_{\alpha/2}\frac{s}{\sqrt{n}}. \tag{11}$$

In (11), s is the sample standard deviation-unbiased estimator of σ—and $z_{\alpha/2}$ is the $(1 - \alpha/2)$ percentile of $N(0, 1)$. The curve in Fig. 7 plots (10) for the *complete set* test results $(n = 665)$ and shows that the neural-network system achieved $\varepsilon = 0.0175$ with confidence $p(= 1 - \alpha) = 0.95$.

V. Conclusion

Feedforward NNs can effectively support objective quality assessment of MPEG-2 videos. In this respect, the major result of the presented research is the possibility of reproducing human perception consistently by using quantitative data-driven models. The neural-network model is specifically tuned to learn the perceptual phenomenon from examples, and exploits a known effective augmentation of standard BP networks.

A crucial advantage of the proposed methodology is the system ability to handle compressed video streams. Avoiding the need for decompressed pictures enhances the method's effectiveness in real-time production applications.

The experimental setup involved a training phase with observations collected from evaluation panels, and generalization testing using sequences and the associated quality assessments not included in the training sets. Experimental evidence confirmed the validity of the approach, as the system always provided a satisfactory continuous-time approximation for the actual scoring curves related to test videos. A comparison with related works is complicated by the lack of a consolidated standard allowing reliable comparison among quality-evaluation methods; more importantly, this is even more true when considering that the approach presented in this paper treats no-reference objective quality assessment, which is new in the literature to the best of our knowledge.

APPENDIX

OBJECTIVE FEATURES

The following quantities are defined:

—

$$\text{energy} = \frac{1}{256}\sum_{i=0}^{16}\sum_{j=0}^{16}(mb_{\text{DCT}}[i][j])^2 \tag{12}$$

where $mb_{\text{DCT}}[i][j]$ are the DCT coefficients of a P or B macroblock. This quantity gives the energy of the correction to the predicted macroblock.

—

$$q\_mv = \frac{q\_\text{scale}}{1+\langle|m\_v|\rangle} \tag{13}$$

where q_scale is the quantiser-scale factor in a macroblock, and $\langle|m\_v|\rangle$ is the mean amplitude value of motion vectors in the same macroblock.

—

$$e\_mv = \text{energy}\cdot\langle|m\_v|\rangle \tag{14}$$

where e_mv is defined as the weighted energy of a macroblock.

An MPEG-2 bitstream has a hierarchical structure that allows one to get information at multiple levels: sequence, group of pictures, picture, slice, macroblock and block. Objective features have been designed to characterize the stream at the picture level. Table II lists the objective features worked out from the coded bitstream. Four classes of measures can be identified:

— Percentage of macroblocks—Features are defined as follows:

$$F_k = \frac{n_{mb}}{n_T^{(mb)}} \tag{15}$$

where n_{mb} is the number of macroblocks of the type specified in the second column of Table II, and $n_T^{(mb)}$ is the total number of macroblocks in the picture.

— Percentage of blocks—Features are defined as follows:

$$F_k = \frac{n_b}{n_T^{(b)}} \tag{16}$$

where n_b is the number of blocks of the type specified in Table II, and $n_T^{(b)}$ is the total number of blocks in the picture.

— Statistic features are defined as follows:

$$F_k = \begin{cases}\text{mean}(\vec{p})\\ \text{st. deviation}(\vec{p})\\ \text{variance}(\vec{p})\end{cases} \tag{17}$$

where $\vec{p}$ is a vector of values p_i computed on each macroblock of the picture; p_i is given in Table II.

— Percentiles—features are defined as follows:

$$F_k = x_\alpha(\vec{p}) \tag{18}$$

where x_α is the α percentile of $\vec{p}$.

The last feature included in the objective set is Nbits, i.e., the number of bits per picture.

ACKNOWLEDGMENT

The authors wish to thank P. Badino and P. Tedesco for their assistance in developing the method and performing the experiments described in the paper.

REFERENCES

[1] "Methodology for the Subjective Assessment of the Quality of Television Pictures," International Telecommunication Union, Geneva, Switzerland, Recommendation BT.500-10, 2000.

[2] S. Olsson, M. Stroppiana, and J. Baïna, "Objective methods for assessment of video quality: State of the art," *IEEE Trans. Broadcast.*, vol. 43, pp. 487–95, Dec. 1997.

[3] W. Y. Zou and P. J. Corriveau, "Methods for evaluation of digital television picture quality," in *IEEE Broadcast Tech. Soc.—4th Meet. G-2.1.6, Audio Video Techniques Committee G-2.1, Compression Processing Subcommittee*, Boulder, CO, May 1997, Doc.~G-2.1.6/28.

[4] D. K. Fibush, "Pratical application of objective picture quality measurements," *SMPTE J.*, vol. 108, no. 1, pp. 10–9, Jan. 1999.

[5] S. D. Voran and S. Wolf, "The development and evaluation of an objective video quality assessment system that emulates human viewing panels," in *Proc. IBC 1992—IEE Int. Broadcast. Conv.*, 1992, pp. 504–8.

[6] M. Ardito and M. Visca, "Correlation between objective and subjective measurements for video compressed systems," *SMPTE J.*, vol. 105, no. 12, pp. 768–73, Dec. 1996.

[7] G. W. Cermak, S. Wolf, E. P. Tweedy, M. H. Pinson, and A. Webster, "Validating objective measures of MPEG video quality," *SMPTE J.*, vol. 107, no. 4, pp. 226–35, Apr. 1998.

[8] A. Pessoa, A. Falcão, R. Nishihara, A. Silva, and A. Lotufo, "Video quality assessment using objective parameters based on image segmentation," *SMPTE J.*, vol. 108, no. 12, pp. 865–72, Dec. 1999.

[9] T. Hamada, S. Miyaji, and S. Matsumoto, "Picture quality assessment system by three-layered bottom-up noise weighting considering human visual perception," *SMPTE J.*, vol. 108, no. 1, pp. 20–6, Jan. 1999.

[10] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video systems," in *Proc. SPIE Int. Symp. Voice, Video Data Commun.*, vol. 3845, Boston, MA, Sept. 1999, pp. 266–77.

[11] K. T. Tan and M. Ghanbari, "A multi-metric objective picture quality measurement model for MPEG video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 1208–13, Oct. 2000.

[12] T. Vlachos, "Detection of blocking artifacts in compressed video," *Electron. Lett.*, vol. 36, no. 13, pp. 1106–8, June 2000.

[13] F.-H. Lin and R. M. Mersereau, "Rate-quality tradeoff MPEG video encoder," *Signal Processing: Image Commun.*, vol. 14, no. 4, pp. 297–309, Feb. 1999.

[14] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1986.

[15] S. Ridella, S. Rovetta, and R. Zunino, "Circular back-propagation networks for classification," *IEEE Trans. Neural Networks*, vol. 8, pp. 84–97, Jan. 1997.

[16] V. N. Vapnik, *Statist. Learning Theory*. New York: Wiley, 1998.

[17] S. Ridella, S. Rovetta, and R. Zunino, "Circular backpropagation networks embed vector quantization," *IEEE Trans. Neural Networks*, vol. 10, pp. 972–975, July 1999.

[18] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, "Accelerating the convergence of the back propagation method," *Biol. Cybern.*, vol. 59, pp. 257–263, 1988.

[19] *Information Technology—Generic Coding of Moving Pictures and Associated Audio Video Information: Video*, ISO/IEC 13 818-2, 1996.

[20] "Introduction of a New Method for Single Stimulus Continuous Quality Evaluation (SSCQE)," International Telecommunication Union, Draft revision of Rec.ITU-R BT.500-7, ITU-R SG 11/E Document 11/21, 1996.

[21] I. T. Joliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.

[22] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, vol. C-23, pp. 881–890, Sept. 1974.

[23] J. H. Friedman, "Exploratory projection pursuit," *J. Amer. Statist. Assoc.*, vol. 82, no. 397, pp. 249–266, 1987.

[24] R. Aldridge and D. Pearson, "A calibration method for continuous video quality (SSCQE) measurements," *Signal Processing: Image Commun.*, vol. 16, no. 3, pp. 321–32, 2000.

[25] H. de Ridder and R. Hamberg, "Continuous assessment of image quality," *SMPTE J.*, vol. 106, no. 2, pp. 123–8, Feb. 1997.
[26] N. Narita, Y. Sugiura, and I. Yuyama, "Time-sensitive evaluation of the quality of digital coded sequences," *SMPTE J.*, vol. 108, no. 1, pp. 32–8, Jan. 1999.
[27] H. R. Schiffman, *Sensation and Perception—An Integrated Approach*, 4th ed. New York: Wiley, 1996.
[28] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D. Pearson, "Recency effect in the subjective assessment of digitally-coded television pictures," in *Proc. MOSAIC Workshop Advanced Methods Evaluation Television Picture Quality*, Sept. 1995.
[29] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten, "Neural gas' network for vector quantization and its application to time-series prediction," *IEEE Trans. Neural Networks*, vol. 4, pp. 558–569, July 1993.
[30] S. Ridella, S. Rovetta, and R. Zunino, "K-winner machines for pattern classification," *IEEE Trans. Neural Networks*, vol. 12, pp. 371–85, Mar. 2001.
[31] S. Rovetta and R. Zunino, "Efficient training of neural gas vector quantizers with analog circuit implementation," *IEEE Trans. Circuits Syst. II*, vol. 46, pp. 688–98, June 1999.
[32] G. P. Drago and S. Ridella, "Pruning with interval arithmetic perceptron," *Neurocomput.*, vol. 18, pp. 229–46, 1998.
[33] A. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*. New York: McGraw-Hill, 1974.

Paolo Gastaldo received the Laurea degree in electronic engineering from Genoa University, Genoa, Italy, in 1998. He is currently with the Electronic Systems and Networking Group of the Department of Biophysical and Electronic Engineering, University of Genoa, and is pursuing the Ph.D. degree in space science and engineering.

His research interests include neural-network implementation and applications, advanced multimedia signal processing, and digital television.

Stefano Rovetta received the Laurea degree in electronic engineering in 1993 and the Ph.D. degree in models, methods and tools for electronic and electromagnetic systems in 1997, both from the University of Genoa, Genoa, Italy.

He held a position as a Postdoctoral Researcher with the Electronic Systems and Networking Group of the Department of Biophysical and Electronic Engineering, University of Genoa. He has been an Invited Professor of Operating Systems at the University of Siena. He is currently Assistant Professor at the Department of Computer and Information Sciences of Genoa University. His research interests include electronic circuits and systems, and neural-network theory, implementation, and applications.

Rodolfo Zunino (S'90–M'90) received the Laurea degree in electronic engineering from Genoa University, Genoa, Italy, in 1985.

From 1986 to 1995, he was a Research Consultant with the Department of Biophysical and Electronic Engineering of Genoa University. He is currently with the same department as an Associate Professor in Industrial Electronics. His main scientific interests include electronic systems for neural networks, efficient models for data representation and learning, advanced techniques for multimedia data processing, and distributed-control methodologies.

Prof. Zunino is a Member of SMPTE.