SIViP (2013) 7:391–401
DOI 10.1007/s11760-013-0442-5

REVIEW

Survey of information theory in visual quality assessment

Rajiv Soundararajan · Alan C. Bovik

Received: 29 November 2011 / Revised: 17 April 2012 / Accepted: 15 June 2012 / Published online: 17 March 2013

Abstract We survey information theoretic approaches to solve a variety of visual quality assessment (QA) problems. These approaches are generally built on natural scene statistical models and lead to practical automatic QA algorithms delivering excellent performance in terms of correlation with human judgments of quality. We study all three categories of image QA models: full reference (FR), reduced reference (RR) and no reference (NR) image QA, as well as FR video QA and information weighting strategies for FR image and video QA. We demonstrate the application of information theory in each of these problems. Each of these problems presents its own challenges in the design of information theoretic QA indices leading to different algorithms under different scenarios. In the algorithms, we survey, FR image and video QA algorithms are based on mutual information or conditional Kolmogorov complexities; RR image QA algorithms either use relative entropy or entropic differences, while the NR QA algorithm applies Rényi entropy, and the weighting strategies rely on mutual information. We also discuss various open research questions, particularly in the realm of NR image QA and all classes of video QA.

Keywords Visual quality assessment · Natural scene statistics · Image information · Mutual information · Relative entropy · Pooling strategies

R. Soundararajan (✉) · A. C. Bovik
Department of ECE, The University of Texas at Austin, Austin, TX 78712, USA
e-mail: rajivs@utexas.edu

A. C. Bovik
e-mail: bovik@ece.utexas.edu

1 Introduction

The advent of the internet and mobile telephony has led to an explosion of highly visual applications and services. These applications currently range from such basic Internet services as video streaming, video conferencing and video on demand to sophisticated computational photography capabilities on mobile handsets and tablets. In all of these applications, the human is the ultimate consumer of the visual content and being able to assess and act upon the quality of experience of the human can help provide better services. Visual quality assessment (QA) concerns the design of automatic algorithms that can predict the quality of a visual signal, such as a still image, video or three-dimensional presentation, in a manner that agrees well with human judgments of quality. The quality index output by a QA algorithm can then be used to achieve a better perceptually optimized design, or to guide suitable control actions, such as rate control, to enhance the user's visual experience.

Over the years, various researchers have attempted to develop algorithms for visual QA. Although many of the earlier algorithms were statistical, they were not necessarily model based or information theoretic and measured empirical statistics between the reference and distorted visual signals. The mean squared error was computed in a transformed domain in [1], while the idea of a Bayesian ideal observer was used to evaluate the probability of perceiving a distortion at a given pixel location in [2]. A comprehensive survey of various distance measures between the images in the pixel domain is contained in [3]. Figures of merit for different targeted image processing applications such as classification and estimation are presented in [4–6]. None of these algorithms have been tested comprehensively against subjective quality predictions, and their applicability for a wide range of distortions is unknown.

There exist two broad classes of methods that successfully solve the problem of visual QA. One class of methods is a human visual system (HVS)-based paradigm, where the degree of perceived distortion is predicted after processing the reference and distorted signals. The processing resembles the filtering and processing that occurs along the early human visual pathways. For example, multiscale multiorientation transforms, divisive normalization or adaptive gain control, motion tuning, Minkowski pooling [7,8], and so on are all related to models of human visual processing of visual stimuli. Complementary to HVS methods, are natural scene statistics (NSS)-based methods, whereby measured deviations of certain statistics of a distorted signal from the expected statistics under an NSS model are exploited to measure quality. The hypothesis beneath this paradigm is that the deviations from the NSS are relevant for QA in that they capture a loss of visual "naturalness." While these two paradigms are complementary to each other, in many ways, they are also dual, and NSS-based methods can be viewed as "matching" HVS models and vice-versa.

Recent information theoretic approaches to visual QA come under the class of NSS model–based methods and attempt to quantify the distances between reference and distorted signals using information theoretic quantities. Many algorithms that are based on this approach attempt to quantify the amount of loss of visual information that occurs in the distorted visual signal with respect to a presumably distortionless reference. The idea motivating the design of these algorithms is that natural images[1] occupy a very small subspace of all possible images that could be represented by matrices of numbers. In this regard, measuring an information theoretic distance from a distorted image to the space of natural images could correlate strongly with a loss of naturalness, that, in turn affects subjective impressions of visual quality. Of course, algorithms developed using such principles need to have demonstrable perceptual relevance and to generate quality indices that correlate well with human judgment of quality. This makes the problem of designing information theoretic visual QA algorithms a challenging problem.

Here, we broadly overview information theoretic approaches to a subset of problems in QA, including full reference (FR), reduced reference (RR) and no reference (NR) still image QA, video QA and information weighting methods in image and video QA. While information theoretic quantities can be directly used to predict quality of the distorted image in local patches [9,10], they may also be indirectly used to weight quality evaluations at different locations in "quality maps" [11].

[1] We use the term "natural images" as it is ordinarily used by vision scientists, meaning images (or videos, or 3D) taken by ordinary optical cameras sensitive to visible light.

The classification of QA models into three different categories depending on the availability of reference information (FR, RR and NR QA) has important practical implications. The constraint of the problem (with the constraint being the amount of reference information that can be made available) implies that only certain information theoretic quantities can be computed under the constraints. For example, it is not possible to compute the mutual information between a reference signal and a corresponding distorted signal in NR QA, since such a computation requires the joint distribution, which cannot be obtained when the reference is not available. Thus, every problem requires its own adaptation of what information theoretic quantities can be computed and are perceptually relevant and meaningful for QA.

Information theoretic FR image QA (IQA) indices [9,10, 12] that have been developed to date generally use measures of the mutual information between reference and distorted images to quantify losses of visual information arising from distortion. While different approaches yield different statistical models and algorithms, in many of these algorithms, local mutual information computations form the core of the analysis. Recently, there has also been work on using the notion of conditional Kolmogorov complexity to obtain FR IQA algorithms [13]. Note that both mutual information and conditional Kolmogorov complexity require full knowledge of both reference and distorted images.

If only partial reference information is available, as under RR IQA models, one possible solution is to parametrize the reference distribution, then compute a measure of dissimilarity (such as Kullback–Liebler (KL) divergence) [14] between the empirical distribution of the distorted image and of the parametrized reference image. It is possible to evaluate the divergence using just the parameters of the reference distribution and the available distorted image. Improved versions of this model can be found in [15]. The other solution is to compute the local information in the reference and send the weighted sum of this information from the reference [16]. Sending weighted sums from the reference leads to a reduction in the amount of reference information sent, albeit at the cost of reduced performance of the algorithm. The model in [17] leads to an NR IQA algorithm that measures the amount of randomness in the orientations (anisotropy) through Rényi entropy calculations.

The underlying information theoretic distortion measures in FR video QA (VQA) algorithms are similar to those in IQA, with the difference being that VQA algorithms operate on spatio-temporal data [18,19]. Although there exist other ideas such as motion tuning [19,20] that can improve performance, these ideas concern the selection of suitable information for QA and do not fundamentally measure any other distances. The role played by different information measures under different constraints on the problem (with regard to the availability of reference information) is still an open research question that can only be answered after good RR and NR VQA algorithms have been developed.

We also discuss information theoretic weighting strategies for image and video QA [11,21]. Weighting strategies are typically employed during the spatial pooling of local quality scores at different spatial locations in an image or a frame of a video. The relative importance of different locations can be assigned based on the amount of associated information. While we discuss only spatial weighting strategies here, the question of whether temporal weighting based on information theoretic principles is perceptually relevant is an interesting future research direction.

The rest of the paper is organized as follows. In Sect. 2, we survey information theoretic IQA algorithms including the statistical models underlying these algorithms. Section 3 covers information theoretic VQA algorithms. In Sect. 4, we describe information weighting strategies for both IQA and VQA and conclude the paper in Sect. 5.

2 Image quality assessment

We begin with an in-depth study of IQA models that use information theoretic measures to assess image quality. We explain the underlying statistical models used in the IQA models; then, we describe FR, RR and NR IQA algorithms that naturally arise from these models. In FR IQA, both a reference and a distorted image are available for quality computation, while in NR IQA no reference information is available. RR IQA is a class of QA algorithms where partial reference information can be made available. The design of RR IQA algorithms concerns the questions of how much and what reference information needs to be made available.

2.1 Statistical models

The use of information theory in image quality assessment necessitates good statistical models which can be used to compute relevant information theoretic quantities such as entropy, relative entropy and mutual information. We overview two popularly used natural scene statistical models that have resulted in a plethora of IQA algorithms under various constraints on the availability of reference information. The wavelet coefficients of natural images tend to have heavy-tailed distributions and are modeled well by both Gaussian scale mixture models as well as by generalized Gaussian distribution models. It is noteworthy that both these models have been successfully used in other image processing applications such as denoising, restoration, compression, retrieval, and so on. In general, while it may not be clear whether one model is better than another, the choice of the model depends on the target application and on reliable estimation of the parameters of the corresponding distributions.

2.1.1 Gaussian scale mixture (GSM) models

Gaussian scale mixtures are an effective model of the wavelet coefficients of natural images [22]. In particular, we model the wavelet coefficients in a given subband of a multiscale multiorientation decomposition of an image. The wavelet coefficients in the subband are partitioned into non-overlapping blocks, and each block of coefficients is modeled as a Gaussian scale mixture vector. Let $\bar{C}$ denote a block of wavelet coefficients. Then, $\bar{C}$ is distributed as

$$\bar{C} = S\bar{U}, \tag{1}$$

where S and $\bar{U}$ are independent and $\bar{U} \sim \mathcal{N}(0, \mathbf{K}_U)$. We refer to S as the premultiplier random variable, which modulates the covariance matrix $\mathbf{K}_U$ for every block. In order to simplify the estimation and computation of quality indices, it is often assumed that different blocks in the subband are independent and that the Gaussian vector $\bar{U}$ is independently and identically distributed for all the blocks. S is a spatially varying continuous random variable that helps better model local variances in the wavelet coefficients.

2.1.2 Generalized Gaussian distribution

A given subband in the wavelet decomposition of an image can also be modeled as obeying a generalized Gaussian distribution. Unlike the Gaussian scale mixture model, this is a global model that assigns a single distribution to every wavelet coefficient in the given subband. Let C denote a wavelet coefficient. The probability density function of a generalized Gaussian distribution with mean zero, shape parameter ν and scale parameter σ is given by

$$f(c; \nu, \sigma) = \frac{\nu}{2\sigma\Gamma(1/\nu)} e^{-|c/\sigma|^{\nu}}, \tag{2}$$

where $\Gamma(\cdot)$ is the gamma function defined as $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t$ for all $z \geq 0$.

Next, we survey and explain various full reference, reduced reference and no reference IQA algorithms that are based on information theoretic approaches. These algorithms employ either of the two statistical models discussed above.

2.2 Full reference IQA

Information theoretic full reference IQ indices can be designed either based on the empirical joint histograms between reference and distorted images or based on a model of the joint distribution between the reference and distorted images. The second method involves estimating the model parameters given both the reference and distorted image, from which the quality index is computed as the amount of information shared between the pristine and distorted images. A key aspect of information theoretic full reference IQA algorithms is the use of the joint distribution between the reference and distorted signals on account of the availability of the reference. As we shall discuss later, in the case of reduced reference and no reference models, this is not possible and we resort to other ways of computing information theoretic measures.

The work in [12] is an example of the first of the above paradigms, where empirical joint distributions of reference and distorted images are used to compute mutual information. In [12], the empirical mutual information is calculated between the reference and distorted after each image has been band pass filtered and then thresholded using the contrast sensitivity function. The quality index is expressed as weighted combinations of the ratio of mutual information between the reference and the distorted to the mutual information of the reference with itself in each subband. The weights are chosen corresponding to the ratio of the energy in a given subband to the total amount of energy in the image.

In [9], the Information Fidelity Criterion (IFC) is developed as the amount of information shared between the reference and distorted wavelet coefficients. In particular, the conditional mutual information between local reference and distorted image wavelet coefficients is calculated, where the conditioning is on the premultiplier of the GSM model being its local maximum likelihood estimate. Further, the reference and the distorted images are related by a linear model involving a scaling/attenuation parameter with independent additive white Gaussian noise. The performance (in terms of correlation with human perception) of this index against human subjectivity is as good as other successful full reference quality indices such as the structural similarity index [23].

There is also related work on IQA algorithms based on the idea of measuring the similarity between two images using Kolmogorov complexity [13]. The normalized information density (NID) is a distance measure between two objects motivated by the conditional Kolmogorov complexity of one object relative to another related object [24]. This distance measure is applicable in multiple scenarios including retrieval, registration and so on. The Kolmogorov complexity of a sequence/object is defined as the length of the shortest program that is required to output the given sequence/object. Let $K(x|y)$ denote the conditional Kolmogorov complexity of x relative to y, where x and y are two related objects. $K(x|y)$ is the length of the shortest program required to output x from y. The authors of [13] define a new quantity known as the normalized conditional compression distance (NCCD) between two objects by using the idea of a conditional image compressor C_T. Let $\{T_i\}_{i=1}^n$ denote a set of transformations of the image to convert it to another image. These transformations could include contrast/luminance changes, Fourier spectrum power scaling and affine transforms such as rotation and translation and so on. Let C_i^p denote the compressor of the parameters used in transformation T_i. The conditional image compressor C_T is defined as

$$C_T(x|y) = \min_i\{C(y - T_i(x)) + C_i^p(p(T_i, x)) + \log_2 N\},$$

where C denotes a practical image compressor (such as a lossless image compressor) and $p(T_i, x)$ denotes the vector of all parameters used in the transformation $T_i(x)$. Then, the NCCD is defined as

$$\text{NCCD}(x, y) = \frac{\max\{C_T(x|y), C_T(y|x)\}}{\max\{C(x), C(y)\}}.$$

While there are a few examples where NCCD gives a better evaluation of the quality of the image when compared to the structural similarity index (SSIM) [23], as compared to human judgments, NCCD appears to be particularly useful for evaluating the quality of non-natural or synthetic images. Many of the most successful IQA algorithms exploit the regularity of natural scene statistics and could potentially fail to successfully evaluate the quality of non-natural images, where the notion of NCCD could be useful.

We now discuss in detail, one of the most successful full reference IQA algorithms based on computing the amount of visual information shared between wavelet coefficients of the reference and distorted images.

2.2.1 Visual information fidelity

Before describing the visual information fidelity (VIF) index [10], we describe the source, distortion and QA models used in the design of the index. Let $\bar{C}_{mkr}$ and $\bar{C}_{mkd}$ denote vectors of wavelet coefficients of length N in Block m and Subband k of the wavelet decomposition of the reference and distorted images, respectively, where $m \in \{1, 2, \ldots, M_k\}$ and $k \in \{1, 2, \ldots, K\}$. The reference vector is modeled as a GSM vector in (1), and represented as $\bar{C}_{mkr} = S_{mk}\bar{U}_{mk}$, where $\bar{U}_{mk} \sim \mathcal{N}(0, \mathbf{K}_{U_k})$. Note that all of the Gaussian vectors corresponding to different non-overlapping blocks in the subband are assumed independent and identically distributed with the same covariance matrix $\mathbf{K}_{U_k}$ for every block of wavelet coefficients in a subband. The vector of wavelet coefficients in the distorted image is modeled using a linear channel model between a block of wavelet coefficients in the reference and distorted signals. More specifically,

$$\bar{C}_{mkd} = g_{mk}\bar{C}_{mkr} + \bar{Z}_{mk}, \tag{3}$$

where $\bar{Z}_{mk}$ is distributed as $\bar{Z}_{mk} \sim \mathcal{N}(0, \sigma_Z^2\mathbf{I}_{N\times N})$ and is independent of $\bar{C}_{mkr}$, while $g_{mk} \in \mathbb{R}$. Despite the use of a simplistic channel model of the processes that transform

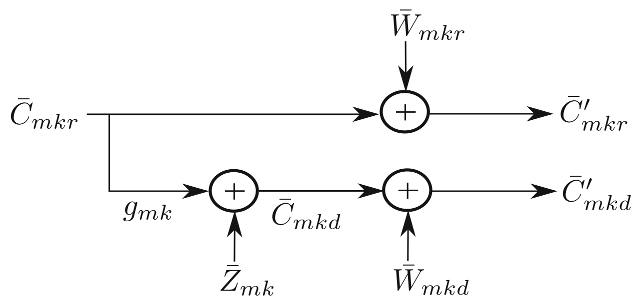

Fig. 1 QA Model for VIF Index

the reference signal into the distorted one, this still delivers very good performance of the VIF QA indices. Further, the noise vectors are assumed independent across different blocks in every subband. In order to model imperfect reception of the signals by the perceptual apparatus, the reference and distorted coefficients are passed through an additive white Gaussian noise channel corresponding to neural noise. In particular, the noisy wavelet coefficients can be expressed as

$$\bar{C}'_{mkr} = \bar{C}_{mkr} + W_{mkr} \qquad \bar{C}'_{mkd} = \bar{C}_{mkd} + W_{mkd}, \qquad (4)$$

where $\bar{W}_{mkr} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_{N\times N})$ and $\bar{W}_{mkd} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_{N\times N})$. The distortion channel and neural noise model are depicted in Fig. 1.

Let s_{mk} denote the maximum likelihood estimate of S_{mk} given the observed wavelet coefficients in the subband. Then, the Visual Information Fidelity (VIF) index is given by

$$\text{VIF} = \frac{\sum_{k=1}^{K} \sum_{m=1}^{M_k} I(C_{mkr}; C'_{mkd} | S_{mk} = s_{mk})}{\sum_{k=1}^{K} \sum_{m=1}^{M_k} I(C_{mkr}; C'_{mkr} | S_{mk} = s_{mk})}. \qquad (5)$$

It is worth mentioning the perceptual relevance of various aspects of the VIF index, especially those that are obtained as a result of adopting the information theoretic approach in its design. For a more detailed description of these aspects, we refer the reader to [9,10]. Note that the mutual information is computed between conditional random variables, where the conditioning is on the realization of the premultiplier random variable. Conditioning on the premultiplier random variable corresponds to divisive normalization models in the early stages of human visual processing. One of the objectives of both divisive normalization as well as conditioning on premultipliers is to reduce correlations between neighboring wavelet coefficients. Further, mutual information expressions under the GSM model (when conditioned on the premultiplier) are evaluated as the logarithm of one plus the "local signal to noise ratio." The application of this logarithm has parallels with findings in vision literature, wherein the ability to perceive differences in distortions decreases with the severity of the distortion [25]. The application of $\log(1+\text{SNR})$ implies that the function saturates below to 0 for small SNR. This property enables the algorithm to respond with reduced sensitivity to differences in distortions for small values of SNR.

The VIF index is one of the best performing quality indices on the LIVE Image Quality Assessment Database [26], with its performance similar to that of the multiscale (MS) structural similarity index (SSIM) [27]. It is shown in [28] that under the GSM model for natural images, there exists a monotonic relation between the mutual information terms in the VIF index and the structure term in the MS-SSIM index. This explains the observation that both these quality indices exhibit similar performances although developed independently of each other.

2.3 Reduced reference IQA

The problem of reduced reference IQA is more constrained than the FR IQA problem, since only partial information about the reference is available for quality computation. The design of RR IQA algorithms involves the question of how much and what information needs to be supplied from the reference in order to be able to evaluate the quality of the distorted image. Further restricting ourselves to solutions based on information theoretic approaches raises the specific question of what information theoretic measures can actually be computed in such scenarios. As mentioned earlier, the availability of the reference enables the computation of the joint distribution and consequently the mutual information between the reference and the distorted signals. This is not possible in the case of reduced reference. Current techniques either compute the relative entropy between the probability distributions of the reference and the distorted [14,15] or the differences of the entropies of the reference and distorted [16]. We discuss each of these techniques in greater detail in the following.

2.3.1 KL divergence based methods

The algorithms in [14,15] are based on computing the KL divergence between the distributions of the wavelet coefficients in a subband of the reference and the distorted images. In [14], the generalized Gaussian distribution is used to model wavelet coefficients in a subband of the reference image. The parameters of the generalized Gaussian distribution along with the KL divergence between the fitted generalized Gaussian distribution and the empirical distribution of the wavelet coefficients in a subband of the reference are transmitted. The quality index involves computing the KL divergence between the empirical distribution of the distorted image and the reference distribution parametrized using the generalized Gaussian distribution.

Let $k \in \{1, 2, \ldots, K\}$ denote the subband of the wavelet decomposition. Let $p_k(c)$ denote the generalized Gaussian distribution used to fit the wavelet coefficients in subband k of the reference image using parameters (ν_k, σ_k), and let $q_k(c)$ denote the empirical distribution. The information sent from the reference includes $\{(\nu_k, \sigma_k)\}_{k=1}^{K}$ and $\{d(p_k||q_k)\}_{k=1}^{K}$, where $d(p||q)$ denotes the KL divergence between the distributions $p(c)$ and $q(c)$. The KL divergence between $p(c)$ and $q(c)$ is

$$d(p||q) = \int p(c) \log \frac{p(c)}{q(c)}.$$

Now, let the empirical distribution of the wavelet coefficients of the distorted image in subband k be $r_k(c)$. We calculate the distance

$$\hat{d}(q_k||r_k) = d(p_k||r_k) - d(p_k||q_k).$$

The resulting quality index is calculated as [14]

$$D = \log_2\left(1 + \frac{1}{D_0}\sum_{k=1}^{K}|\hat{d}(q_k||r_k)|\right),$$

where D_0 is a constant scaling factor.

The algorithm just mentioned is global and does not account for the distortion of local structures in an image. An improved version of this algorithm can be found in [15], where a divisive normalization transform step is introduced prior to fitting distributions and computing KL divergences. This leads to an improvement in the performance of the algorithm since the concept of divisive normalization has parallels with the signal processing that happens in the human visual pathways and achieves further decorrelation in neighboring wavelet coefficients (neural responses tuned to different spatial locations and frequencies) [8]. Interestingly, in this work, the divisive normalization transform (DNT) is performed by using a Gaussian scale mixture model of blocks of wavelet coefficients, where the block includes neighboring coefficients in the same subband and those in neighboring orientations and scales. The maximum likelihood estimate of the premultiplier in the GSM model in (1) is used to divisively normalize the wavelet coefficients. The normalized coefficients are modeled globally using a Gaussian distribution per subband. As in the previous example, the parameters of the Gaussian distribution are sent along with the error in fitting a Gaussian distribution to the normalized coefficients. A distance is calculated between the empirical distributions of the DNT wavelet coefficients of the distorted image and the reference image, using the KL divergence between the parametrized reference distribution and the distorted image and the error in fitting the Gaussian distribution to the DNT coefficients of the reference image. In addition to the KL divergence, the absolute values of the differences between the standard deviations, kurtosis and skewness of the distributions of the normalized coefficients are also calculated. We refer the reader to [15] for further details on this algorithm.

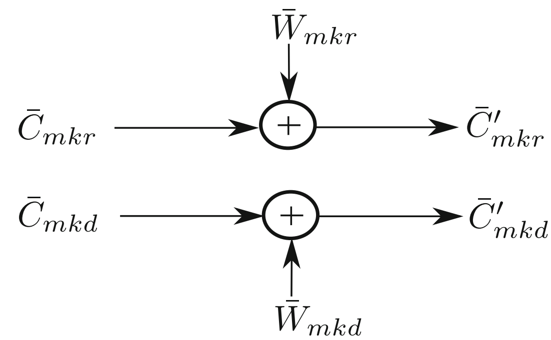

Fig. 2 QA Model for RRED Indices

2.3.2 RRED indices

The second approach to computing information theoretic measures in reduced reference scenarios is to compute visual information differences between the reference and distorted images. This can be done by computing entropic differences between the wavelet coefficients of the reference and distorted images. The framework used in [10] to compute mutual information between the reference and the distorted signals is adapted in this algorithm to compute entropic differences. Let $\bar{C}_{mkr}$ denote a block of wavelet coefficients in subband k, with $k \in \{1, 2, \ldots, K\}$ and $m \in \{1, 2, \ldots, M_k\}$, where M_k denotes the total number of blocks in subband k after partitioning the subband of wavelet coefficients into non-overlapping blocks. Unlike [10], where a linear additive noisy model was used to represent the joint distribution between the wavelet coefficients of the reference and the distorted images, here, both the reference and the distorted images are fitted with GSM models as in (1).

The quality index essentially involves computing the differences of scaled conditional entropies of the neural noisy reference and distorted wavelet coefficients. Let $\bar{C}'_{mkr}$ and $\bar{C}'_{mkd}$ denote neural noisy wavelet coefficients of the reference and distorted images, respectively. These are represented as

$$\bar{C}'_{mkr} = \bar{C}_{mkr} + \bar{W}_{mkr} \qquad \bar{C}'_{mkd} = \bar{C}_{mkd} + \bar{W}_{mkd},$$

and pictorially depicted in Fig. 2.

The RRED index in subband k using M_k scalars from the reference is given by

$$\begin{aligned}\mathrm{RRED}_k^{M_k} &= \frac{1}{M_k}\sum_{m=1}^{M_k}|\gamma_{mkr}h(C'_{mkr}|S_{mkr} = s_{mkr}) \\ &\quad -\gamma_{mkd}h(C'_{mkd}|S_{mkd} = s_{mkd})|,\end{aligned}$$

where s_{mkr} and s_{mkd} are maximum likelihood estimates of S_{mkr} and S_{mkd}, respectively, and $\gamma_{mkr} = \log_2(1+s^2_{mkr})$ and $\gamma_{mkd} = \log_2(1+s^2_{mkd})$. For further details on the estimation of the parameters of the distributions leading to computation of the RRED indices, the reader is referred to [16]. The RRED index in subband k using a single scalar from the reference is given by

$$\mathrm{RRED}^1_k = \frac{1}{M_k}\left|\sum_{m=1}^{M_k} \gamma_{mkr} h(C'_{mkr}|S_{mkr} = s_{mkr}) - \gamma_{mkd} h(C'_{mkd}|S_{mkd} = s_{mkd})\right|.$$

The authors of [16] develop a family of algorithms that systematically utilize differing amounts of side information required from the reference for quality computation. Depending on the application, the user could pick a desired algorithm from this family.

2.4 No reference IQA

No reference IQA is a challenging perceptual research problem that concerns the estimation of the quality of a distorted image without using any information from the corresponding reference image. Over the years, researchers have developed training-based methods (motivated by machine learning applications) to learn the way humans predict quality, leading to algorithms that can automatically predict the quality of the distorted image in a manner that agrees well with human subjective judgments of quality. While algorithms in this class were first developed only for specific types of distortions, recently, there has been significant progress on designing algorithms for arbitrary distortions and that can predict quality in a manner that correlates quite well with subjective judgments of quality [29–32]. Although these methods are successful, they rely on the availability of distorted images during the training phase. In particular, these algorithms can predict the quality of a distorted image in a manner that agrees well with human perception, provided that the algorithm has already been trained on images (different from the test image) corrupted by the distortion types assumed to afflict the given image being assessed.

The features used in the machine learning paradigm for quality assessment are statistical and exploit the regularity of natural scene statistics. The use of information theoretic features is one avenue for designing information theoretic blind IQA algorithms, building on the machine learning paradigm. On the other hand, [17] designs a blind IQA algorithm based on measuring the amount of anisotropy (or the amount of randomness in the orientation of local structures) through the Rényi entropy. While the performance of this algorithm is not particularly impressive [31], we briefly review it since it represents a concrete instance of using information theoretic approaches to design a truly blind IQA algorithm.

The algorithm attempts to compute randomness in local image orientation by computing the Rényi entropy. Consider a pseudo-Wigner distribution of the distorted image. The pseudo-Wigner distribution is a spatial–spatial frequency representation of the image that is complementary to the use of wavelets in signal processing. Let $P(n,k)$ denote the pseudo-Wigner distribution at spatial index n and spatial frequency index k. Note that according to this representation, the spatial indices of the image are represented as a single dimensional array of indices. Since the pseudo-Wigner distributions are not necessarily normalized, we define

$$Q(n,k) = \frac{1}{C_n} P(n,k)P^*(n,k),$$

where $P^*(n,k)$ is the complex conjugate of $P(n,k)$ and C_n is chosen such that such that

$$\sum_k Q(n,k) = 1.$$

Next, treat $Q(n,k)$ as a probability distribution that sums to 1 over k and define the Rényi entropy of order α as

$$R_\alpha(n) = \frac{1}{1-\alpha}\log_2\left(\sum_k Q^\alpha(n,k)\right).$$

The Rényi entropy associated with spatial index n is computed along different orientations using the pseudo-Wigner distributions along different orientations. For further details on how the pseudo-Wigner distributions may be computed for different directions, the reader is referred to [17]. Let $\theta \in \{1, 2, \ldots, \Theta\}$ denote the orientation indexes along which the pseudo-Wigner distributions are evaluated. Define the average Rényi entropy in orientation θ as

$$\bar{R}(\theta) = \frac{1}{N}\sum_{n=1}^{N} R^\theta_3(n),$$

where $R^\theta_3(n)$ denotes the Rényi entropy of order 3 in orientation θ at location index n. The quality index is defined as the standard deviation of the Rényi entropy along different directions. Letting the mean

$$\mu = \frac{1}{\Theta}\sum_{\theta=1}^{\Theta} \bar{R}_3(\theta),$$

the quality index is then given by

$$\sigma = \sqrt{\frac{1}{\Theta}\sum_{\theta=1}^{\Theta}\left(\bar{R}_3(\theta) - \mu\right)^2}.$$

The design of information theoretic blind QA indices is in its early stages and clearly more research is needed in this direction. The purely blind approach in [17] relies on the changes in the statistics of local structures (through anisotropy) in the distorted image alone to estimate quality. This method is particularly interesting since it does not require a database of pristine images for quality prediction. It would be interesting to find out deviations of what other notions of local structures in the distorted images can help estimate the quality. The choice of the information theoretic quantity used in such an analysis (for example, Rényi entropy in case of anisotropy) would depend on the choice of features being analyzed. The work in [33] concerning the information theoretic analysis of interscale and intrascale dependencies between wavelet coefficients of images appears to be particularly relevant in this context.

3 Video quality assessment

We now discuss two video quality assessment (VQA) algorithms [18,19] that are based on information theoretic principles. They represent early examples of how algorithms may be designed using information theoretic principles. Both of the algorithms are FR VQA algorithms and extend the framework used in VIF [10] to videos in two different ways. While [18] is based on modeling spatio-temporal blocks of the spatio-temporal derivatives of the frames, [19] models a three-dimensional block of spatio-temporal Gabor decomposition coefficients as GSM vectors. Further details on these algorithms are presented in the following.

The algorithm in [18] closely resembles the algorithm in [10] with the difference being in the definition of the corresponding blocks in the videos between which the mutual information is calculated. In particular, the mutual information between the spatio-temporal blocks of spatio-temporal derivatives are calculated using GSM models in (1) for the block of derivatives. Further, the derivatives are obtained along all the 3 dimensions in spatio-temporal space in the YUV color domain leading to 9 subbands. Let $k \in \{1, 2, \ldots, K\}$ index the subbands and $m \in \{1, 2, \ldots, M\}$ index all the spatio-temporal non-overlapping blocks in the video. Let $\bar{C}_{mkr}$ and $\bar{C}_{mkd}$ denote blocks of the spatio-temporal derivatives in subband k and block m in the reference and distorted videos, respectively. These are related by a linear model as in (3). Further, we consider the neural noisy versions of C_{mkr} and C_{mkd}, denoted by C'_{mkr} and C'_{mkd}, respectively, following (4). The video VIF index is then defined in the same manner as in (5).

By contrast, the method explained in [19] builds on the information fidelity criterion (IFC) introduced in [9] for images, to design information theoretic VQA algorithms. A given video is decomposed using K spatio-temporal Gabor filters, and blocks of Gabor coefficients surrounding a given point in the spatio-temporal space are modeled as GSM vectors. Let $\bar{C}_{mkr}$ and $\bar{C}_{mkd}$ denote blocks of Gabor coefficients with $m \in \{1, 2, \ldots, M\}$ indexing all the spatio-temporal blocks in the video and $k \in \{1, 2, \ldots, K\}$ indexing the Gabor filters used in the decomposition. At every point in the spatio-temporal space, the mutual information is only calculated between blocks corresponding to a subset of the filters used to obtain the Gabor filter responses. Let $\mathcal{K}(m) \subset \{1, 2, \ldots, K\}$ denote the subset of filters chosen based on the optical flow vectors at these points as follows. Only those filters are chosen that have a significant overlap with the local orientation plane containing the frequency spectrum as estimated by using the optical flow vectors. For further details on determining this plane and other details on measuring the overlap, we refer the reader to [19]. The quality index is then computed as

$$\mathrm{V-IFC} = \frac{\sum_{m=1}^{M}\sum_{k\in\mathcal{K}(m)} I(\bar{C}_{mkr}; \bar{C}_{mkd}|S_{mkr} = s_{mkr})}{\sum_{m=1}^{M}|\mathcal{K}(m)|}.$$

While both these algorithms perform quite well on the VQEG dataset [34], the performance of [18] on the LIVE Video Quality Assessment Database [35] is not among the best. The MOVIE index [20], which is a more general version of these algorithms, is currently the best performing index on the LIVE Video Quality Assessment Database. Certain aspects of the MOVIE index are essentially equivalent to both [18] and [19] owing to the relation between mean squared error and IFC/VIF described in [28].

4 Information content weighting in image and video quality assessment

A key ingredient of perceptual visual QA algorithms is the method of pooling of local quality scores obtained at localized regions/blocks of the image to obtain global predictions of quality. In particular, many successful FR visual QA algorithms employ perceptually relevant pooling strategies, where pooling refers to the aggregation of local quality scores into a single number as a measure of the quality of the visual signal. While IQA only concerns spatial pooling of scores, VQA needs to resolve questions of both spatial and temporal pooling of quality scores, making it a more challenging problem. In the literature, multiple ways of dealing with pooling have been explored. The methods in [36] and [20] are statistical methods of pooling quality scores. While the former deals with percentile methods, the latter adopts the notion of coefficient of variation. In this section, we survey information theoretic approaches to pooling quality scores. The premise of this approach is to weight the quality evaluations at locations containing rich information more than the other locations. We present one IQA [11] and one VQA algorithm [21] to illustrate how local quality evaluations may be pooled using information theoretic quantities.

4.1 Information weighting for IQA

In [11], the local quality scores between corresponding blocks at a given scale and position are weighted using the total amount of perceptual information content in both the reference and the distorted images. Let $\bar{C}_{mkr}$ and $\bar{C}_{mkd}$ denote blocks of transform coefficients in Block $m \in \{1, 2, \ldots, M_k\}$ and Subband $k \in \{1, 2, \ldots, K\}$ of reference and distorted images respectively. The number of subbands chosen needs to match the number of those in the QA algorithm that is weighted. If the underlying QA algorithm is not obtained by using K subbands, then the information weights thus obtained must be appropriately related to the underlying QA algorithm. For further details on this issue, the reader may refer to [11]. $\bar{C}'_{mkr}$ and $\bar{C}'_{mkd}$ denote neural noisy versions of $\bar{C}_{mkr}$ and $\bar{C}_{mkd}$, respectively, where

$$\bar{C}'_{mkr} = \bar{C}_{mkr} + \bar{W}_{mkr} \qquad \bar{C}'_{mkd} = \bar{C}'_{mkd} + \bar{W}_{mkd},$$

with $\bar{W}_{mkr} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I})$ and $\bar{W}_{mkd} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I})$. Further, $\bar{C}_{mkr}$ and $\bar{C}_{mkd}$ are modeled as GSM vectors as in (1).

The total amount of perceptual information content in the reference and the distorted images is expressed as

$$\begin{aligned}\gamma_{mk} &= I(\bar{C}_{mkr}; \bar{C}'_{mkr} | S_{mkr} = s_{mkr}) \\ &\quad + I(\bar{C}_{mkd}; \bar{C}'_{mkd} | S_{mkr} = s_{mkr}) \\ &\quad - I(\bar{C}'_{mkr}; \bar{C}'_{mkd} | S_{mkr} = s_{mkr}).\end{aligned} \tag{6}$$

Using these weights, any IQA algorithm that yields local quality scores at different locations (blocks) can be weighted. Suppose Q_{mk} represents a local quality score corresponding to Block m and Subband k. Then, the weighted quality score is given by

$$Q = \frac{\sum_{k=1}^{K} \sum_{m=1}^{M_k} \gamma_{mk} Q_{mk}}{\sum_{k=1}^{K} \sum_{m=1}^{M_k} \gamma_{mk}}. \tag{7}$$

The authors of [11] show that this information theoretic weighting strategy can be used to improve the performance of many existing IQA algorithms. In particular, the application of information content weighting to peak signal to noise ratio (PSNR) and SSIM improves their performance on multiple image quality assessment databases as shown in [11].

The VIF index [10] can also be cast in the context of information content weighting strategies as follows [11]. By analyzing the VIF index, (5) may be interpreted according to (7) as

$$\mathrm{VIF} = \frac{\sum_{k=1}^{K} \sum_{m=1}^{M_k} \gamma_{mk} \mathrm{VIF}_{mk}}{\sum_{k=1}^{K} \sum_{m=1}^{M_k} \gamma_{mk}},$$

where

$$\begin{aligned}\mathrm{VIF}_{mk} &= \frac{I(\bar{C}_{mkr}; \bar{C}'_{mkd} | S_{mkr} = s_{mkr})}{I(\bar{C}_{mkr}; \bar{C}'_{mkr} | S_{mkr} = s_{mkr})} \\ \gamma_{mk} &= I(\bar{C}_{mkr}; \bar{C}'_{mkr} | S_{mkr} = s_{mkr}).\end{aligned} \tag{8}$$

This implies that in the VIF index, local quality evaluations obtained as VIF_{mk} are weighted according to γ_{mk}. However, observe that γ_{mk} used in (8) is different from the γ_{mk} defined in general in (6) for information content weighting in QA algorithms.

4.2 Information weighting for VQA

We discuss another information theoretically motivated weighting strategy for the problem of VQA [21]. The algorithm is obtained by weighting the local structural similarity index (SSIM) values [23] according to the amount of perceived motion information at that location. The assumption underlying the algorithm is that quality evaluations at locations with high motion information need to be weighted more since they have a larger impact on the quality of the video. Note that this paradigm only concerns spatial weighting of quality scores. A simple averaging over all the frames is used for temporal pooling.

There are three notions of motion associated with the algorithm. While global motion $\bar{v}_g$ refers to the background motion, $\bar{v}_a$ refers to the absolute motion at a given location. The relative motion is defined as $\bar{v}_r = \bar{v}_a - \bar{v}_g$. As in [21], we use the notation $v = \|\bar{v}\|$. Note that we refer to optical flow vectors by motion here, although there could be other notions of motion such as frame differences, for example. Based on observations in [37], a power law distribution is assumed for the prior on relative motion and is given by

$$p(v_r) = \frac{\tau}{v_r^{\alpha}},$$

where α and τ are constants. Note that this is not a valid probability distribution and can only be used away from $v_r = 0$. The amount of information associated with the relative motion at a given location can be quantified as

$$I = -\log p(v_r) = -\log \tau + \alpha \log v_r.$$

Further, the perceived global motion v'_g is modeled as a log normal distribution [37] and is related to the global motion v_g by

$$p(v'_g | v_g) = \frac{1}{\sqrt{2\pi}\, v'_g \sigma} e^{\frac{-(\log v'_g - \log v_g)^2}{2\sigma^2}},$$

where the width parameter σ parametrizes the deviation of v'_g from v_g. σ is modeled as

$$\sigma = \frac{\lambda}{c^{\gamma}},$$

where λ and γ are positive constants and c is the local contrast. Consequently, different locations have different deviations from the global motion. The amount of perceptual uncertainty is expressed as the entropy of the likelihood function $p(v'_g|v_g)$ and is given by

$$\begin{aligned} U &= -\int_{-\infty}^{\infty} p(v'_g|v_g)\log p(v'_g|v_g)\mathrm{d}v'_g \\ &= \frac{1}{2} + \frac{1}{2}\log(2\pi\sigma^2) + \log v_g \\ &= \log v_g - \gamma\log c + \delta, \end{aligned}$$

where $\delta = \frac{1}{2} + \frac{1}{2}\log(2\pi) + \log\lambda$ is a constant. The weighting at a given spatio-temporal location is given by

$$w = I - U = (\alpha\log v_r + \beta) - (\log v_g - \gamma\log c + \delta).$$

Finally, suppose $q(x, y, t)$ is the quality evaluation at a given spatio-temporal index in the video. Then, the weighted VQA index is given by

$$Q = \frac{\sum_t\sum_x\sum_y w(x, y, t)q(x, y, t)}{\sum_t\sum_x\sum_y w(x, y, t)}.$$

Indeed, it is verified in [35] that the performance of weighting the SSIM index based on perceptual uncertainty of motion does improve its performance over the algorithm that does not perform such a weighting. Although such an algorithm does not perform as well as the best performing QA indices on the LIVE Video Quality Assessment Database, the results indicate that information theoretic weighting of spatial QA indices using motion information can be useful in video QA.

5 Conclusion

Information theoretic methods have found an important role in image and video quality assessment. We have summarized key contributions in the design of information theoretic methods for FR, RR and NR IQA. Although these algorithms are often motivated by a natural scene statistics paradigm, we showed how some aspects of their design are dual with respect to IQA algorithms based on perceptual approaches. Further, these algorithms achieve state of the art performances within the class of FR and RR IQA algorithms. We also looked at extensions of information theoretic FR IQA algorithms to VQA. Finally, we surveyed information weighting approaches for any QA algorithm toward designing methods for pooling the local quality evaluations into a single QA score. Pooling based on the amount of inherent information at different locations is a principled, yet perceptually relevant strategy and helps improve the performance of both image and video QA algorithms.

There are multiple open research problems for future directions. As we mentioned earlier, the use of information theoretic features in learning/training-based methods for NR image/video QA is a possible direction. The other direction is to extend the methods in [17] and attempt purely information theoretic NR QA model design based on the measurement of local structures. However, in such an approach, the key research question is in figuring out what data needs to be analyzed information theoretically rather than the question of what information theoretic quantities to use. There is considerable scope for improvement in the design of information theoretic VQA models that could better analyze temporal video information. While both [18] and [19] look at spatio-temporal data, it might be interesting to consider the spatial and temporal data independently in view of the accepted model of independent processing of these in the visual cortex [20]. Based on the work in [28], which shows the equivalence of some of the information theoretic QA indices to other successful image QA indices under some assumptions, it might be possible to adapt some of the key ideas in successful visual QA indices into information theoretic–based approaches. Overall, we hope that we have been able to deliver a meaningful sample of how information theoretic methods are playing an important role in the creation of visual QA algorithms exhibiting excellent performance. However, there remain many open research problems. For example, the use of information theoretic methods is still nascent in NR IQA and in all classes of VQA, and the level of success of these methods can probably be significantly improved.

References

1. Mannos, J.L., Sakrison, D.J.: The effects of visual fidelity criterion on the encoding of images. IEEE Trans. Inf. Theory **4**, 525–536 (1974)
2. Heeger, D.J., Teo, P.C.: A model of perceptual image fidelity. In: Proceedings of the IEEE International Conference on Image Processing, pp. 343–345 (1995)
3. Eskicioglu, A.M., Fisher, P.S.: Image quality measures and their performance. IEEE Trans. Commun. **43**, 2959–2965 (1995)
4. Barrett, H.H.: Objective assessment of image quality: effects of quantum noise and object variability. J. Opt. Soc. Am. A **7**(7), 1266–1278 (1990)
5. Barrett, H.H., Denny, J.L., Wagner, R.F., Myers, K.J.: Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance. J. Opt. Soc. Am. A **12**, 834–852 (1995)

6. Barrett, H.H., Abbey, C.K., Clarkson, E.: Objective assessment of image quality III: ROC metrics, ideal observers, and likelihood-generating. J. Opt. Soc. Am. A **15**(12), 1520–1535 (1998)
7. Wang, Z., Bovik, A.C.: Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Processing Magazine **26**(1), 98–117 (2009)
8. Wang, Z., Bovik, A.C.: Reduced- and no-reference image quality assessment. The natural scene statistic model approach. Special Issue on Multimedia Quality Assessment, IEEE Signal Processing Magazine (2011)
9. Sheikh, H.R., Bovik, A.C., de Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Trans. Image Process. **14**(12), 2117–2128 (2005)
10. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. **15**(2), 430–444 (2006)
11. Wang, Z., Li, Q.: Information content weighting for perceptual image quality assessment. IEEE Trans. Image Process. **20**(5), 1185–1198 (2011)
12. Haijun, Z., Huayi, W.: New paradigm for compressed image quality metric: exporing band similarity with CSF and mutual information. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium (2005)
13. Nikwand, N., Wang, Z.: Generic image similarity based on Kolmogorov complexity. In: Proceedings of IEEE International Conference on Image Processing, Hong Kong, China (2010)
14. Wang, Z., Wu, G., Sheikh, H.R., Simoncelli, E.P., Yang, E.H., Bovik, A.C.: Quality-aware images. IEEE Trans. Image Process. **15**(5), 1680–1689 (Jun. 2006)
15. Li, Q., Wang, Z.: Reduced-reference image quality assessment using divisive normalization-based image representation. IEEE J. Sel. Top. Signal Process. Special issue on Visual Media Quality Assessment **3**, 202–211 (2009)
16. Soundararajan, R., Bovik, A.C.: RRED indices: reduced reference entropic differencing for image quality assessment. IEEE Trans. Image Process. **21**(2), 517–526 (2012)
17. Gabarda, S., Cristobal, G.: Blind image quality assessment through anisotropy. J. Opt. Soc. Am. A **24**, B42–B51 (2007)
18. Sheikh H.R., Bovik, A.C.: A visual information fidelity approach to video quality assessment. In: Proceedings of 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ (2005)
19. Seshadrinathan, K., Bovik, A.C.: An information theoretic video quality metric based on motion models. In: Proceedings of 3rd International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Arizona (2007)
20. Seshadrinathan, K., Bovik, A.C.: Motion-tuned spatio-temporal quality assessment of natural videos. IEEE Trans. Image Process. **19**, 335–350 (2010)
21. Wang, Z., Li, Q.: Video quality assessment using a statistical model of human visual speed perception. J. Opt. Soc. Am. A (Optics, Image Science and Vision) **24**(12), B61–B69 (2007)
22. Portilla, J., Strela, V., Wainwright, M.J., Simoncelli, E.P.: Image denoising using scale mixture of Gaussians in the wavelet domain. IEEE Trans. Image Process. **12**(11), 1338–1351 (2003)
23. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error measurement to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
24. Li, M., Chen, X., Li, X., Ma, B., Vitányi, P.M.B.: The similarity metric. IEEE Trans. Inf. Theory **50**, 3250–3264 (2004)
25. Watson, A.B., Kreslake, L.: Measurement of visual impairment scales for digital video. In: Proceedings of SPIE 4299 (2011)
26. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: Live image quality assessment database release 2. Available at http://live.ece.utexas.edu/research/quality
27. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: Proceedings of 37th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA (2003)
28. Seshadrinathan, K., Bovik A.C.: Unifying analysis of full reference image quality assessment. In: Proceedings of IEEE International Conference on Image Processing, San Diego, CA (2008)
29. Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. IEEE Signal Process. Lett. **17**(5), 587–599 (2010)
30. Saad, M.A., Bovik, A.C., Charrier, C.: A DCT statistics based blind image quality index. IEEE Signal Process. Lett. **17**(6), 583–586 (2010)
31. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: from scene statistics to perceptual quality. IEEE Trans. Image Process. **20**(12), 3350–3364 (2011)
32. Saad, M.A., Bovik, A.C.: Blind image quality assessment: a natural scene statistics approach in the DCT domain. IEEE Trans. Image Process. **21**, 3339–3352 (2012)
33. Liu, J., Moulin, P.: Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. IEEE Trans. Image Process. **10**, 1647–1658 (2001)
34. Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment. Available at http://www.its.bldrdoc.gov/vqeg/projects/frtv_phase1
35. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: Study of subjective and objective quality assessment of video. IEEE Trans. Image Process. **19**, 1427–1441 (2010)
36. Moorthy, A.K., Bovik, A.C.: Visual importance pooling for image quality assessment. IEEE J. Sel. Top. Signal Process. **3**, 193–201 (2009)
37. Stocker, A.A., Simoncelli, E.P.: Noise characteristics and prior expectations in human visual speed perception. Nat. Neurosci. **9**(4), 578–585 (2006)