AN EFFICIENT PERCEPTUAL QUALITY INDEX METRICS

D.Neelima[1], Shaika Sultana[2]

[1] *Assistant Professor, Dept of CSE, Jawaharlal Nehru Technological University, Kakinada, India.*
[2] *M.Tech student, Dept of Information Technology, Jawaharlal Nehru Technological University, Kakinada, India.*

Abstract

In Human Visual System (HVS) video quality assessment is more important. We have studied HVS, we observed non systematic and impairing of videos in that. So we propose a novel approach that is called Perceptual quality index based on video quality assessment(VQS).In this paper visual performance equation for the foveal and extra-foveal vision, perceptible noise detection, instantaneous error summation metrics are introduced. For this metrics cortical magnification theory, spatial-temporal just noticeable difference model, perceptual quality index are taken to introduce these metrics.

Index Terms: *Human visual system, perceptual quality index, temporal noise, video quality assessment.*

----------***----------

1. INTRODUCTION

Video quality is a characteristic of a video passed through a video transmission/processing system, a formal or informal measure of perceived video degradation (typically, compared to the original video). Video processing systems may introduce some amounts of distortion or artifacts in the video signal, so video quality evaluation is an important problem. Objective video evaluation techniques are mathematical models that approximate results of subjective quality assessment, but are based on criteria and metrics that can be measured objectively and automatically evaluated by a computer program. Objective methods are classified based on the availability of the original video signal, which is considered to be of high quality (generally not compressed). Therefore, they can be classified as Full Reference Methods (FR), Reduced Reference Methods (RR) and No-Reference Methods (NR). FR metrics compute the quality difference by comparing every pixel in each image of the distorted video to its corresponding pixel in the original video. RR metrics extract some features of both videos and compare them to give a quality score. They are used when all the original video is not available, *e.g.* in a transmission with a limited bandwidth. NR metrics try to assess the quality of a distorted video without any reference to the original video. These metrics are usually used when the video coding method is known.

Digital video data, stored in video databases and distributed through communication networks, is subject to various kinds of distortions during acquisition, compression, processing, transmission, and reproduction. For example, lossy video compression techniques, which are almost always used to reduce the bandwidth needed to store or transmit video data, may degrade the quality during the quantization process. For another instance, the digital video bit streams delivered over error-prone channels, such as wireless channels, may be received imperfectly due to the impairment occurred during transmission. Package-switched communication networks, such as the Internet, can cause loss or severe delay of received data packages, depending on the network conditions and the quality of services. All these transmission errors may result in distortions in the received video data. It is therefore imperative for a video service system to be able to realize and quantify the video quality degradations that occur in the system, so that it can maintain, control and possibly enhance the quality of the video data. An effective image and video quality metric is crucial for this purpose.

The most reliable way of assessing the quality of an image or video is subjective evaluation, because human beings are the ultimate receivers in most applications. The mean opinion score (MOS), which is a subjective quality measurement obtained from a number of human observers, has been regarded for many years as the most reliable form of quality measurement. However, the MOS method is too inconvenient, slow and expensive for most applications. The goal of objective image and video quality assessment research is to design quality metrics that can predict perceived image and video quality automatically. Generally speaking, an objective image and video quality metric can be employed in three ways:

- It can be used to *monitor* image quality for quality control systems. For example, an image and video acquisition system can use the quality metric to monitor and automatically adjust itself to obtain the best quality

image and video data. A network video server can examine the quality of the digital video transmitted on the network and control video streaming.

- It can be employed to *benchmark* image and video processing systems and algorithms. If multiple video processing systems are available for a specific task, then a quality metric can help in determining which one of them provides the best quality results.
- It can be embedded into an image and video processing system to *optimize* the algorithms and the parameter settings. For instance, in a visual communication system, a quality metric can help optimal design of the prefiltering and bit assignment algorithms at the encoder and the optimal reconstruction, error concealment and postfiltering algorithms at the decoder.

Objective image and video quality metrics can be classified according to the availability of the original image and video signal, which is considered to be distortion-free or perfect quality, and may be used as a reference to compare a distorted image or video signal against. Most of the proposed objective quality metrics in the literature assume that the undistorted reference signal is fully available. Although "image and video quality" is frequently used for historical reasons, the more precise term for this type of metric would be image and video *similarity* or *fidelity* measurement, or full-reference (FR) image and video quality assessment. It is worth noting that in many practical video service applications, the reference images or video sequences are often not accessible. Therefore, it is highly desirable to develop measurement approaches that can evaluate image and video quality blindly. Blind or no-reference (NR) image and video quality assessment turns out to be a very difficult task, although human observers usually can effectively and reliably assess the quality of distorted image or video without using any reference. There exists a third type of image quality assessment method, in which the original image or video signal is not fully available. Instead, certain features are extracted from the original signal and transmitted to the quality assessment system as side information to help evaluate the quality of the distorted image or video.

This is referred to as **reduced-reference (RR)** image and video quality assessment. Currently, the most widely used FR objective image and video distortion/quality metrics are mean squared error (MSE) and peak signal-to-noise ratio (PSNR), which are defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \quad (1)$$

$$\mathrm{PSNR} = 10\log_{10}\frac{L^2}{\mathrm{MSE}} \quad (2)$$

Where N is the number of pixels in the image or video signal, and x_i and y_i are the i-th pixels in the original and the distorted signals, respectively. L is the dynamic range of the pixel values. For an 8bits/pixel monotonic signal, L is equal to 255. MSE and PSNR are widely used because they are simple to calculate, have clear physical meanings, and are mathematically easy to deal with for optimization purposes (MSE is differentiable, for example). However, they have been widely criticized as well for not correlating well with perceived quality measurement. In the last three to four decades, a great deal of effort has been made to develop objective image and video quality assessment methods (mostly for FR quality assessment), which incorporate perceptual quality measures by considering human visual system (HVS) characteristics. Some of the developed models are commercially available.

Perceptual image distortion

The model consists of four stages: (1) front-end linear filtering, (2) squaring, (3) normalization, and lastly (4) detection. The first stage of the model decomposes the image locally into its spatial frequency and orientation components. The coefficients of the linear filtering are then squared to yield local energy measures. Because the human visual system is differentially sensitive to local image frequency composition, the third stage normalizes the squared coefficients accordingly. Both the reference and distorted images are subjected to the first three stages; the final detection stage then determines the amount of distortion visible in the distorted image.

Linear Transform

Many researchers have suggested a variety of linear transforms which resemble the orientation and spatial frequency tuning of cortical receptor fields [71, 70, 72] or psychophysically determined visual sensors [68]. These linear transforms often have the following characteristics: (1) octave spacing and frequency bandwidths and (2) narrow orientation selectivity. In addition to these considerations, we are also concerned about the computational efficiency of the transform (and its inverse). In our previous work, we used a quadrature mirror filter suite on a hexagonally-sampled image [74]. The transform is orthogonal and thus is compact in its representation and efficiently computed. Unfortunately, being orthogonal, the basis functions describing a local region of an image severely alias one another. 1 Moreover, as noted previously, the orientation bandwidth of the hex-QMF's are a little too broad. In this paper, we adopt the steerable pyramid transform introduced by Simoncelli et al [73]. The transform decomposes the image locally into several spatial frequency levels within which each level is further divided into a set of

orientation bands. Figure 1 shows an analysis/synthesis representation of the transform. The basic functions for each level of the pyramid have octave bandwidths and are separated from those of neighboring levels by an octave as well. In our implementation, we divide every level into six orientation bands with bandwidths of approximately thirty degrees. The orientation decomposition at each level is steerable [75], i.e. the response of a filter tuned to any orientation can be obtained through a linear combination of the responses of the six basis filters computed at the same location. This property is important as it implies that the orientation decomposition is locally rotationally-invariant. The pyramid is also designed to minimize the amount of aliasing within each sub band. Thus, the steerable pyramid, unlike the hex-QMF transform, is over complete and non-orthogonal. Even so, the transform is self-inverting which allows the inverse to be efficiently computed despite its non-orthogonality.

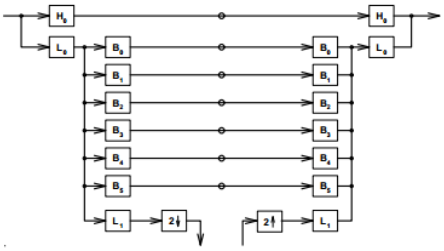


Figure 1: Analysis/synthesis representation of the steerable pyramid transform implemented. H0 is a high-pass filter; the Li 's represent low-pass filters and the Bi's represent orientation selective filters.

Squaring and Normalization

The front-end linear transform yields a set of coefficient values for every region in the image. These coefficients are next squared to obtain energy measures of the local orientation and spatial frequency components. Since the front-end transform is linear, a coefficient's magnitude increases linearly with the contrast of the input image. Furthermore, these linear coefficients are equally sensitive (or insensitive) to perturbations of the input regardless of image contrast. Squaring introduces a simple contrast-dependence on sensitivity. However, squaring alone does not account for masking effects. Furthermore, the magnitude of the response of each sensor can potentially be very large. On the other hand, the dynamic range of the mechanisms in the visual system is limited. Normalization is required to predict masking effects and to restrict the range of response magnitudes of our hypothetical visual sensors. The normalization scheme is divisive and is determined by two parameters: an overall scaling constant, k, and a saturation constant, σ^2. Let A be a coefficient of the frontend linear transform. The squared and normalized output, R , is computed as follows:

$$R^{\theta} = k \frac{(A^{\theta})^2}{\sum_{\phi} (A^{\phi})^2 + \sigma^2} \quad (3)$$

Where ϕ ranges over all sensors tuned to different orientations. In our implementation, - 2 f0; 30; 60; 90; 120; 150g. We treat each spatial frequency level of the pyramid separately and conduct this pooling only over sensors tuned to different orientations. Hence, the normalized output of a sensor tuned to orientation θ is computed by dividing its original squared response $(A^{\theta})^2$, by the sum of the squared responses of a pool of sensors over all orientations in the same region of the image. Since this summation, $\sum_{\phi} (A^{\phi})^2$, includes the term A^{θ}, that appears in the numerator (i.e., each sensor suppresses itself), as long as σ is nonzero, the normalized sensor response will always be a value between 0 and k, saturating at high contrasts. Each of the normalized sensors has a limited dynamic range as shown in Figure 2. In other words, each sensor is able to discriminate contrast differences only over a narrow range of contrasts. This range is determined by the scaling and saturation constants, k and σ^2, respectively. Hence, several contrast normalization mechanisms, each having different k_i's and σ_i^2's, are required to discriminate contrast changes over the full range of contrasts. In the current implementation of the model, we have four different contrast discrimination bands (that is, four different choices and ki). In summary, the front-end linear transform yields a set of coefficients which measure the different orientation and spatial frequency components in each local region of the image. With squaring and multiple normalizations, the number of measurements for each local region is increased fourfold. However, these local image measurements now analyze the image into its orientation, spatial frequency and contrast components. Furthermore, masking effects over orientation are captured by the pooling step in the normalization.

Detection

The detection mechanism determines locally if a distortion is visible. Let R_{ref} be a vector of normalized sensor responses from a local region in the reference image. Let R_{dist} be the vector of normalized responses from the corresponding region in the distorted image. The detection mechanism adopted by the model is the simple squared-error norm (i.e., the vector distance between R_{ref} and R_{dist}):

$$\triangle \mathcal{R} = ||\mathcal{R}_{ref} - \mathcal{R}_{dist}||^2 \quad (4)$$

One might include all of the normalized sensor responses (all spatial positions, spatial frequencies, orientations, and contrast discrimination bands) in the vectors, R0 and R1, and compute a single number representing the overall detectability

of differences between the two images. We find it more informative, however, to implement the detection mechanism independently for each local patch (or block) of the images. The vector distance detection mechanism can be justified in terms of an ideal observer model. The vector of normalized sensor responses from each image correspond to the mean responses of noisy sensors. For the purposes of the model, we assume the noise to be additive, independent, identically-distributed, zero-mean Gaussian noise. Furthermore, the standard deviation of the noise is independent of the mean response. With these assumptions and an ideal observer model, the vector distance detection mechanism gives the likelihood that the ideal observer would detect the distortion. For example, assuming a standard deviation of one for the noise, the observer is able to detect the distortion 76% of the time when the squared difference is exactly one. In signal detection theory, this corresponds to a d0 of one. In our model, we assume detection at this efficiency. Hence, $\triangle \mathcal{R}$ in equation (2) is equal to one at threshold.

2. RELATED WORK

2.1 The Human Visual System

Figure 2 schematically shows the early stages of the HVS. It is not clearly understood how the human brain extracts higher-level cognitive information from the visual stimulus in the later stages of vision, but the components of the HVS depicted in Figure 41.1 are fairly well understood and accepted by the vision science community.

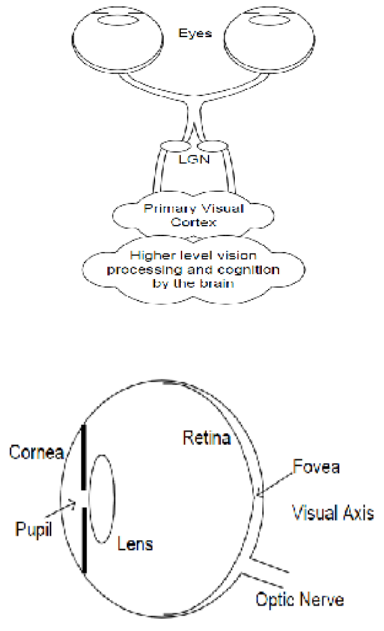


Figure 2 Schematic diagram of the human visual system.

2.2 Anatomy of the HVS

The visual stimulus in the form of light coming from objects in the environment is focussed by the optical components of the eye onto the retina, a membrane at the back of the eyes that contains several layers of neurons, including photoreceptor cells. The optics consists of the cornea, the pupil (the aperture that controls the amount of light entering the eye), the lens and the fluids that fill the eye. The optical system focuses the visual stimulus onto the retina, but in doing so blurs the image due to the inherent limitations and imperfections. The blur is low-pass, typically modelled as a linear space-invariant system characterized by a point spread function (PSF). Photoreceptor cells in the retina sample the image that is projected onto it.

There are two types of photoreceptor cells in the retina: the cone cells and the rod cells. The cones are responsible for vision in normal light conditions, while the rods are responsible for vision in very low light conditions, and hence are generally ignored in the modelling. There are three different types of cones, corresponding to three different light wavelengths to which they are most sensitive. The L-cones, M-cones and S-cones (corresponding to the Long, Medium and Short wavelengths at which their respective sensitivities peak) split the image projected onto the retina into three visual streams. These visual streams can be thought of as the Red, Green and Blue color components of the visual stimulus, though the approximation is crude. The signals from the photoreceptors pass through several layers of interconnecting neurons in the retina before being carried off to the brain by the optic nerve.

The photoreceptor cells are non-uniformly distributed over the surface of the retina. The point on the retina that lies on the visual axis is called the fovea (Figure 41.1), and it has the highest density of cone cells. This density falls off rapidly with distance from the fovea. The distribution of the ganglion cells, the neurons that carry the electrical signal from the eye to the brain through the optic nerve, is also highly non-uniform, and drops off even faster than the density of the cone receptors. The net effect is that the HVS cannot perceive the entire visual stimulus at uniform resolution.

The visual streams originating from the eye are reorganized in the optical chiasm and the lateral geniculate nucleus (LGN) in the brain, before being relayed to the primary visual cortex. The neurons in the visual cortex are known to be tuned to various aspects of the incoming streams, such as spatial and temporal frequencies, orientations, and directions of motion.

Typically, only the spatial frequency and orientation selectivity is modelled by quality assessment metrics. The

neurons in the cortex have receptive fields that are well approximated by two-dimensional Gabor functions. The ensemble of these neurons is effectively modelled as an octave-band Gabor filter bank [74,70], where the spatial frequency spectrum (in polar representation) is sampled at octave intervals in the radial frequency dimension and uniform intervals in the orientation dimension [71]. Another aspect of the neurons in the visual cortex is their saturating response to stimulus contrast, where the output of a neuron saturates as the input contrast increases.

Many aspects of the neurons in the primary visual cortex are not modelled for quality assessment applications. The visual streams generated in the cortex are carried off into other parts of the brain for further processing, such as motion sensing and cognition. The functionality of the higher layers of the HVS is currently an active research topic in vision science.

2.3 Psychophysical HVS Features

2.3.1 Foveal and Peripheral Vision

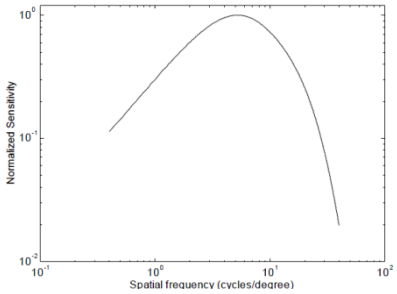
As stated above, the densities of the cone cells and the ganglion cells in the retina are not uniform, peaking at the fovea and decreasing rapidly with distance from the fovea. A natural result is that whenever a human observer fixates at a point in his environment, the region around the fixation point is resolved with the highest spatial resolution, while the resolution decreases with distance from fixation point. The high-resolution vision due to fixation by the observer onto a region is called *foveal* vision, while the progressively lower resolution vision is called *peripheral* vision. Most image quality assessment models work with foveal vision.

2.3.2 Light Adaptation

The HVS operates over a wide range of light intensity values, spanning several orders of magnitude from a moonlit night to a bright sunny day. It copes with such a large range by a phenomenon known as light adaptation, which operates by controlling the amount of light entering the eye through the pupil, as well as adaptation mechanisms in the retinal cells that adjust the gain of post-receptor neurons in the retina. The result is that the retina encodes the contrast of the visual stimulus instead of coding absolute light intensities. The phenomenon that maintains the contrast sensitivity of the HVS over a wide range of background light intensity is known as Weber's Law.

2.3.3 Contrast Sensitivity Functions

The contrast sensitivity function (CSF) models the variation in the sensitivity of the HVS to different spatial and temporal frequencies that are present in the visual stimulus. This variation may be explained by the characteristics of the receptive fields of the ganglion cells and the cells in the LGN, or as internal noise characteristics of the HVS neurons. Consequently, some models of the HVS choose to implement CSF as a filtering operation, while others implement CSF through weighting factors for subbands after a frequency decomposition. The CSF varies with distance from the fovea as well, but for foveal vision, the spatial CSF is typically modelled as a space-invariant band-pass function (Figure 41.2). While the CSF is slightly band-pass in nature, most quality assessment algorithms implement a low-pass version. This makes the quality assessment metrics more robust to changes in the viewing distance.



2.3.4 Masking and Facilitation

Masking and facilitation are important aspects of the HVS in modelling the interactions between different image components present at the same spatial location. Masking/facilitation refers to the fact that the presence of one image component (called the mask) will decrease/increase the visibility of another image component (called the test signal). The mask generally reduces the visibility of the test signal in comparison with the case that the mask is absent. However, the mask may sometimes facilitate detection as well. Usually, the masking effect is the strongest when the mask and the test signal have similar frequency content and orientations.

2.3.5 Pooling

Pooling refers to the task of arriving at a single measurement of quality, or a decision regarding the visibility of the artifacts, from the outputs of the visual streams. It is not quite understood as to how the HVS performs pooling. It is quite obvious that pooling involves cognition, where a perceptible distortion may be more annoying in some areas of the scene (such as human faces) than at others. However, most quality assessment metrics use Minkowski pooling to pool the error signal from the different frequency and orientation selective streams, as well as across spatial coordinates, to arrive at a fidelity measurement.

3. PROPOSED PERCEPTUAL QUALITY INDEX

By incorporating the above-mentioned HVS properties, we propose a systematic framework to simulate the subjective video quality evaluation, as illustrated in Fig. 1. First, attention guides the fixations, and retinal images are generated accordingly. Waveform errors over certain thresholds (influenced by background content) are detected by the HVS. Then, the supra-threshold errors are quantified in different channels, reflecting several perspectives of visual quality of a test video. The quantification process is affected by visual error sensitivity that varies due to masking effects of the background. Visual impairments in each channel at a certain instant are collected and measured as a degradation intensity recorded in the working memory, named as the instantaneous quality degradation. In the visual summation of local errors, weak distortions are inhibited by other stronger impairments. Instantaneous quality degradation in a channel accumulates over time, and turns into a quality feature. Finally, all quality features are fused into an overall quality measurement for the test video, which is usually conveyed by an opinion score.

Based on this perceptual framework, a PQI metric is developed and described specifically in the following subsections. Note that the HVS is far more complicated than this model. For example, human brain resembles a collection of highly specialized parallel-processing machines with high-bandwidth interconnection [25], which indicates the error quantification channels may not work independently. Besides, there may also exist feedback loops in the visual information processing, such as the reactivation of working memory elements. These sophisticated mechanisms are not considered in this framework in view of their complexity.

3.1 Error Detection

In PQI, a spatial-temporal JND model [45] is employed to determine the visibility of waveform distortions. Spatial JND thresholds (*JNDSL*) related to the luminance masking effect were measured by subjective experiments at a viewing distance of six times of the image height (6H) [45]. Viewing distance is recommended to be reduced as the display size increases [2]. Since subjective evaluations nowadays tend to use larger screens, the viewing distance is reduced to typically 4H-6H. Distortions invisible at 6H could be still perceived at 4H. Thus, we repeated the experiment in [45] at 4H to develop a more restrict JND model [55]. The obtained spatial JND thresholds are generally lower than those measured at 6H, especially when the background luminance is low (luminance level < 64). Spatial JND thresholds related to contrast masking (*JNDSC*) are modeled as a function of the background edge strength. The spatial JND profile (for a pixel k) is then determined as the maximum of the two sub-profiles [45], [55]

$$JND_k^S = \max\left(JND_k^{SL},\ JND_k^{SC}\right). \quad (5)$$

The spatial-temporal JND threshold ($JNDS{-}T$) is modeled as the product of the spatial JND value and a non-linear scale factor determined by inter-frame luminance difference (ild)

$$JND_k^{S-T} = f(ild_k) \cdot JND_k^S \quad (6)$$

where the experimental f (ild_k) curve drawn in [45] can be approximated by an exponential function in [55]. Distortions over the JND thresholds are considered as perceptible noise which will be quantified by two means, as to be discussed in the following two subsections.

3.2 Instantaneous Error Summation

Error perception is assumed to be based on a bottom-up model, and the quality degradation at an instant is obtained by the pooling of the local spatial and temporal noises as measured in the previous two subsections. Since the working memory is capacity-limited, only about three to four objects can be remembered at a time [43]. In an impaired video clip, plenty of visual distortions appear simultaneously. It is likely that some weak distortions are inhibited by other more salient distortions, as discussed in Section II-D. Some distorted pixels are spatially close to each other and together present a large-size impairment, while some others are distributed sparsely, inducing small and isolated distortions. In this paper, we reasonably assume an impairment of a small size (e.g., subtending a visual angle less than a threshold θ) is less salient than a large-size counterpart and finally inhibited in the error summation. By suppressing the small-size impairments, the error pooling is expected to be more accurate. Thus, spatial and temporal noise intensity over a single frame n (i.e., SNI_n and TNI_n) are calculated by

$$SNI_n = \sum_{i=1}^{I} s_{i,n}^S \cdot SNI_{i,n}^E,\ s_{i,n}^S = \begin{cases} 1, & \text{if } \theta_{i,n}^S > \theta \\ 0, & \text{otherwise} \end{cases} \quad \textbf{(7)}$$

$$TNI_n = \sum_{i=1}^{I} s_{i,n}^T \cdot TNI_{i,n}^E,\ s_{i,n}^T = \begin{cases} 1, & \text{if } \theta_{i,n}^T > \theta \\ 0, & \text{otherwise} \end{cases} \quad \textbf{(8)}$$

where I denotes the number of blocks in the frame. $s^S_{i,n}$ (or $s^T_{i,n}$) is the inhibition factor based on the size of local spatial (or temporal) noise. For a distorted block, its neighboring $N_\theta \times N_\theta$ region (N_θ pixels subtending a visual angle θ) will be checked. If less than half of the pixels in the region are impaired, the pixel belongs to a small-size distortion; otherwise, it is considered as a part of a large distortion. Empirically, we set θ as a small value 0.1°.

3.3 Quality Degradation Accumulation

In a viewing test, subjects do not rate one frame after another. Owing to the limitation of information processing in our brain, it is more likely that a subject integrates several frame

intervals as a "moment" and rates the moment. Then, quality degradation of these moments will be accumulated in the subject's mind which is finally translated into a quality rating. This moment-based assumption is in accordance with the visual persistence effect [58] that a transient stimulus visually retains for a short period of time (or a moment) beyond the physical termination of the stimulus. Thus, a distorted frame of severe noise may initiate a bad experience that lasts a (short) moment after the frame (i.e., the following N_T frames within T sec). It seems that distortions in a bad-quality frame are perceptually propagated to the next several frames within the described moment. Here, a bad-quality frame N_0 is defined as a frame that has stronger distortions in terms of SNI than the two temporally adjacent frames (i.e., satisfying $SNI_{N_0} > SNI_{N_0-1}$ and $SNI_{N_0} > SNI_{N_0+1}$). To model this effect, the SNI of the N_T frames which have better quality than N_0 (i.e., satisfying $SNI_{N_0+k} < SNI_{N_0}$, $k = 1, 2, ..., N_T$) are modified to be the same as the SNI of the bad-quality frame N_0, that is

$$SNI_{N_0+k} = SNI_{N_0}, k \in \{1, 2, ..., N_T\} \text{ if } SNI_{N_0} > SNI_{N_0-1} \,\&\, SNI_{N_0} > SNI_{N_0+1} \,\&\, SNI_{N_0+k} < SNI_{N_0}. \quad (9)$$

However, if there exists an even worse frame N_1 (i.e., $SNI_{N_1} > SNI_{N_0}$) during the time T, the worse frame refreshes the bad moment and the distortion intensity is further updated to SNI_{N_1}. Then, the bad moment continues for the next N_T frames with distortion intensity of SNI_{N_1}. As visual persistence generally lasts for 0.1 to 0.5 s [58]–[60], we simply consider a moment as $T = 0.25$ s. This pooling algorithm presents an asymmetric tracking on good-quality and bad quality frames (as mentioned in Section II-F). The pooling of the temporal noise uses the same algorithm by simply replacing the SNI with the TNI. Besides, we also consider the recency effect in subjective viewing which may result from the time-dependent activation decay of the working memory elements that record the visual impairments. We normalize the decay function in (2) with $A(0) = 1$, that is

$$A(t) = 1 - \lambda \ln(t+1), \quad t \geq 0, \quad \lambda > 0 \quad (10)$$

And weight it on the SNI and TNI of each frame, that is

$$SNI_n = SNI_n \cdot A(t_n) \quad (11)$$

$$TNI_n = TNI_n \cdot A(t_n) \quad (12)$$

Where t_n means the interval between the frame n and the end of the sequence. The rate of decay λ is determined according to the experiments in [49], in which the same coding distortions are located in the last or the first 10 s of 30 s videos, resulting in different subjective quality degradation score S_1 and S_2 ($S_1 > S_2$). Assuming: 1) the coding distortions are uniform during the 10 s with a constant visual intensity VI, and 2) the subjective quality degradation scores are proportional (with a ratio η) to the total impairment intensity, we have

$$S_1 = \eta \cdot \int_0^{10} VI \cdot A(t)\, dt = \eta \cdot VI \cdot \int_0^{10} 1 - \lambda \ln(t+1)\, dt \quad (11)$$

$$S_2 = \eta \cdot \int_{20}^{30} VI \cdot A(t)\, dt = \eta \cdot VI \cdot \int_{20}^{30} 1 - \lambda \ln(t+1)\, dt. \quad (12)$$

Dividing (28) by (29) and solving the integral, we obtain

$$\frac{[t(1+\lambda) - \lambda(t+1)\ln(t+1)]|_0^{10}}{[t(1+\lambda) - \lambda(t+1)\ln(t+1)]|_{20}^{30}} = \frac{10 - 16.37\lambda}{10 - 32.52\lambda} = \frac{S_1}{S_2} = k. \quad (13)$$

Based on the four pairs of (S_1,S_2) in [49], we have four k values (1.14, 1.45, 1.11, and 2.00), and accordingly derive four λ values (0.068, 0.146, 0.056, 0.201). The λ in (25) is determined as the average of the four values, i.e., $\lambda = 0.12$, which corresponds to $A(10) \approx 0.71$. Then, we normalize the spatial and temporal noise intensities of a test video with respect to each pixel, and obtain the mean spatial noise intensity (MSNI) and mean temporal noise intensity (MTNI)

$$MSNI = \frac{1}{(Fr-1)\cdot P}\sum_{n=2}^{Fr} SNI_n \quad (14)$$

$$MTNI = \frac{1}{(Fr-1)\cdot P}\sum_{n=2}^{Fr} TNI_n \quad (15)$$

where Fr is the number of frames in the video sequence and P is the number of pixels in a frame. Here, the first frame is excluded in the evaluation due to the lack of motion information for the first frame. However, the first frame can be treated as a scene cut following a gray image (as a gray image is usually presented between two video sequences in subjective viewing), and it has been reported that artifacts can be greatly concealed at a scene cut [61]. In this sense, the exclusion of the first frame in quality evaluation would have little influence on the overall quality performance.

3.4 Perceptual Quality Index

Human vision exhibits plenty of nonlinear features [62]. Basically, the Weber's law indicates that the perceived signal intensity is a logarithmic transform of the physical signal intensity. To exploit this non-linear feature, we convert MSNI of the test video into perceived spatial noise intensity (PSNI) using a logarithmic function

$$PSNI = \log_{10} MSNI. \quad (16)$$

Because MSNI must be in the range of [0, 255] for 8-bit videos, PSNI varies in the range of $(-\infty, \log_{10} 255^2]$. Spatial quality index (SQI) is normalized in the range of [0, 1] from PSNI by a division function

$$SQI = 1 - \frac{-1}{\log_{10} MSNI - \log_{10} 255^2 - 1} \quad (16)$$

where

$$\lim_{MSNI \to 0^+} SQI = 1 - \lim_{MSNI \to 0^+} \frac{-1}{\log_{10} MSNI - \log_{10} 255^2 - 1} = 1 \quad (17)$$

and SQI is equal to 0 when $MSNI = 255^2$. Likewise, we define temporal quality index (TQI) as

$$TQI = 1 - \frac{-1}{\log_{10} MTNI - \log_{10} 255^2 - 1}. \quad (18)$$

Since spatial and temporal quality represent two distinct aspects of the overall video quality, we treat them as two equal and complementary bases in the perceptual quality space, and thus define the overall PQI as the Minkowski summation of SQI and TQI, that is

$$PQI = \sqrt{\frac{1}{2}\left(SQI^2 + TQI^2\right)} \quad (19)$$

in which ½ is used for normalizing PQI in the range of [0, 1].

3.5 Error Quantification Channel : Temporal Noise

Some objective waveform distortions change over time, resulting in flickering effects as well as fidelity degradation. These temporally varying spatial distortions are termed as temporal noise. The spatial noise measurement is not capable to discern the temporal noise or to measure the flickering effect, which needs to be quantified separately by a temporal noise measurement discussed in this subsection. Temporal changes can be detected by both the first-order and second order vision. In this paper, we track temporal noise with a first-order (luminance variation) detector on the spatial distortions. Thus, the spatial distortion (SD) is defined in a way toward the average luminance (first-order) impairments with consideration of the error visibility thresholds as follows:

$$SD_{i,n} = \begin{cases} \max(T_{i,n} - R_{i,n} - JND_{i,n}^{S-T},\ 0), & \text{if } T_{i,n} - R_{i,n} \geq 0 \\ \min(T_{i,n} - R_{i,n} + JND_{i,n}^{S-T},\ 0), & \text{otherwise} \end{cases} \quad (16)$$

where $T_{i,n}$ and $R_{i,n}$ are the average luminance of the ith block in the nth frame of the test and the reference sequences, respectively. $JNDs_{-T\ i,n}$ is the average of the JND profile of the block (i, n). SD is an amplitude descriptor, while the energy of the spatial distortion (ESD) is defined as

$$ESD_{i,n} = \min\left(\left(SD_{i,n}\right)^2,\ ST_{i,n}^2\right) \cdot B_{i,n} \quad (17)$$

where $B_{i,n}$ is the number of pixels in the block. $ST_{i,n}$ denotes the average saturation threshold for the block, and is uniformly determined as 30 according to [46]. Based on the temporal variations of the spatial distortions, temporal noise is categorized in three stages (an example is illustrated in Fig. 2). 1) Increasing stage (IS), when the spatial distortion increases beyond the corresponding level range and reaches a higher level (type 1), or its sign changes (type 2). The level range is (−CT, CT), where CT denotes a change threshold. Since human peak sensitivity to flicker is around 0.008 to 0.02 in the photopic condition [38], by assuming the average luminance of a video is γ (γ = 128 for 8-bit videos), CT (in digital level) for each block (i, n) is uniformly determined as

$$CT_{i,n} = \gamma \times \frac{0.008 + 0.02}{2}. \quad (18)$$

IS (type 1)

$$SD_{i,n} - SD_{i+mv_{i,n},n-1} \geq CT_{i,n}. \quad (19)$$

For simplicity, $SD_{i,n}$ in (15) and (18) are treated as positive. Similar expressions can be found for negative $SD_{i,n}$. IS (type 2)

$$Sign(SD_{i,n}) \cdot Sign(SD_{i+mv_{i,n},n-1}) < 0 \quad (20)$$

3) Static stage (SS), when the spatial distortion fluctuates slightly over time but remains at the same level. In this case, the spatial distortion is perceptually static. SS

$$-CT_{i,n} < SD_{i,n} - SD_{i+mv_{i,n},n-1} < CT_{i,n}. \quad (21)$$

The perceptual intensity of the temporal noise is measured based on the tracked temporal distortion variations. In a static stage, the temporal noise intensity (TNI) is determined to be zero, since the static spatial distortion does not cause flickering effects. In an increasing stage (type 1) or a declining stage, the spatial distortion varies (with the same sign). The change of the spatial quality between the pervious and the current instants evokes a flicker, revealing the existence of distortions to subjects. Thus, we measure the TNI at the IS (type 1) or DS with the visual intensity of the distortion variation, i.e., using the difference between the energy of the spatial distortion [ESD as defined in (13)] at the current instant and that at the previous instant. In an increasing stage (type 2), the sign of the spatial distortion changes (but the amplitude may decrease). This temporal variation could be visually stronger than any of the two cases when the spatial distortion turns from the previous amplitude (e.g., a positive) to zero or from zero to the current amplitude (e.g., a negative), as shown in Fig. 2. The TNI at such an instant is measured as the summation of the ESDs at the two instants to account for the visually stronger variation. As a summary, the TNI of a tracked block is defined as

$$TNI_{i,n} = \begin{cases} 0, & if\ (i,n) \in \text{SS} \\ ESD_{i+mv_{i,n},\,n-1} - ESD_{i,n}, & if\ (i,n) \in \text{DS} \\ ESD_{i,n} - ESD_{i+mv_{i,n},\,n-1}, & if\ (i,n) \in \text{IS(type 1)} \\ ESD_{i,n} + ESD_{i+mv_{i,n},\,n-1}, & if\ (i,n) \in \text{IS (type 2)}. \end{cases} \quad (22)$$

Similar to the quantification of the spatial noise, an error sensitivity function is applied to adapt the TNI for different background conditions, that is

$$TNI_{i,n}^E = e_{i,n}^T \cdot c_{i,n}^T \cdot m_{i,n}^T \cdot TNI_{i,n} \quad (23)$$

Where $e^Ti_{,n}$, $c^T_{i,n}$, and $m^T{}_{i,n}$ denote three weighting factors for temporal noise related to the error eccentricity, background contrast, and retinal motion speed, respectively. We assume the error sensitivity factors related to the contrast and motion masking effects are similar for both spatial noise and temporal noise, and apply the same masking models to the temporal noise, i.e., $c^T_{i,n} = c^Si_{,n}$ and $m^Ti_{,n} = m^S_{i,n}$. In addition, since visual performance to the first-order luminance variations generally decreases much more slowly than acuity toward the periphery, we double the $E2$ value in eTi,n (i.e., $E2$ = 5° for temporal noise) to compensate this difference.

CONCLUSION

In this paper a novel approach Perceptual Quality index in video quality assessment framework is efficient than the novel approaches based on HVS. In this method measures waveform distortions from both spatial and temporal perspectives and also combines the HVS properties. PQI approach is more systematic and efficient compared to visual models in the other objective metrics. PQI consistently presents high correlation with the subjective evaluations on the two VQA databases. Compared with some state-of-the-art VQA algorithms.

REFERENCES

[1] H. R. Wu and K. R. Rao, *Digital Video Image Quality and Perceptual Coding*. Boca Raton, FL: CRC Press, 2006, chs. 4, 9, 14.
[2] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R Rec. BT.500-11, International Telecommunication Union, Geneva, Switzerland, 2002.
[3] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.
[4] A. P. Hekstra, J. G. Beerends, D. Ledermann, F. E. de Caluwe, S.
Kohler, R. H. Koenen, S. Rihs, M. Ehrsam, and D. Schlauss, "PVQM: A perceptual video quality measure," *Signal Process. Image Commun.*, vol. 17, no. 10, pp. 781–798, Nov. 2002.
[5] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An objective video quality assessment system based on human perception," in *Proc. 4th SPIE Hum. Vis. Visual Process. Digit. Display*, Feb. 1993, pp. 15–26.
[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[7] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[8] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
[9] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.
[10] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 253–265, Apr. 2009.
[11] M. Barkowsky, J. Bialkowski, B. Eskofier, R. Bitto, and A. Kaup, "Temporal trajectory aware video quality measure," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 266–279, Apr. 2009.
[12] Y. Zhao and L. Yu, "Evaluating video quality with temporal noise," in *Proc. IEEE ICME*, Jul. 2010, pp. 708–712.
[13] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 121–132, Feb. 2004.
[14] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, vol. 24, no. 12, pp. B61–B69, Dec. 2007.
[15] H. Marmolin, "Subjective MSE measures," *IEEE Trans. Syst. Man Cybern.*, vol. 16, no. 3, pp. 486–489, May 1986.
[16] K. T. Tan, M. Ghanbari, and D. E. Pearson, "An objective measurement tool for MPEG video quality," *Signal Process. Image Commun.*, vol. 70, no. 3, pp. 279–294, Nov. 1998.
[17] M. A. Masry and S. S. Hemami, "A metric for continuous quality evaluation of compressed video with severe distortions," *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 133–146, Feb. 2004.
[18] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, no. 10, pp. 1385– 1422, Oct. 1993.
[19] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 16, pp. 1427–1441, Jun. 2010.
[20] S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Process.*, vol. 78, no. 2, pp. 231–252, Oct. 1999.
[21] A. Duchowski, *Eye Tracking Methodology: Theory and Practice*, 2nd ed. London, U.K.: Springer, 2007, chs. 2–3.
[22] V. Virsu, R. Näsänen, and K. Osmoviita, "Cortical magnification and peripheral vision," *J. Opt. Soc. Am. A*, vol. 4, pp. 1568–1578, Aug. 1987.
[23] S. L. Sally and R. Gurnsey, "Foveal and extra-foveal orientation discrimination," *Exp. Brain Res.*, vol. 183, no. 3, pp. 351–360, Jul. 2007.
[24] P. Mäkelä, J. Rovamo, and D. Whitaker, "The effects of eccentricity and stimulus magnification on simultaneous performance in position and movement acuity tasks," *Vis. Res.*, vol. 37, pp. 1261–1270, Apr. 1997.
[25] C. Ware, *Information Visualization: Perception for Design*, 2nd ed. San Francisco, CA: Morgan Kaufmann, 2004, chs. 1, 2, 11.
[26] O. Le Meur, P. Le Callet, and D. Barba, "Predicting visual fixations on video based on low-level visual features," *Vis. Res.*, vol. 47, no. 19, pp. 2483–2498, Sep. 2007.

[27] C. Rashbass, "The relationship between saccadic and smooth tracking eye movements," *J. Physiol.*, vol. 159, no. 2, pp. 326–338, 1961.
[28] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001.
[29] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vis. Res.*, vol. 42, no. 1, pp. 107–123, Jan. 2002.
[30] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, "Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1928–1942, Nov. 2005.
[31] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.
[32] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric," *Signal Process. Image Commun.*, vol. 25, no. 7, pp. 547–558, Aug. 2010.
[33] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," *J. Vis.*, vol. 8, no. 7, pp. 1–20, 2008.
[34] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 802–817, May 2006.
[35] B. W. Tatler, "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions," *J. Vis.*, vol. 7, no. 14, pp. 1–17, 2007.
[36] V. Manahilov, W. A. Simpson, and J. Calvert, "Why is second-order vision less efficient than first-order vision?" *Vis. Res.*, vol. 45, no. 21, pp. 2759–2772, Oct. 2005.
[37] A. J. Schofield and M. A. Georgeson, "The temporal properties of first and second-order vision," *Vis. Res.*, vol. 40, pp. 2475–2487, Aug. 2000.
38] D. H. Kelly, "Visual responses to time-dependent stimuli: I. Amplitude sensitivity measurements," *J. Opt. Soc. Am.*, vol. 51, no. 4, pp. 422–429, 1961.
[39] B. G. Breitmeyer, "Visual masking: Past accomplishments, present status, future developments," *Adv. Cognitive Psychol.*, vol. 3, nos. 1– 2, pp. 9–20, 2007.
[40] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Am.*, vol. 70, no. 12, pp. 1458–1471, Dec. 1980.
[41] N. Li, S. Desmet, A. Deknuydt, and L. V. Eycken, "Motion adaptive quantization in transform coding for exploiting motion masking effect," in *Proc. SPIE Visual Commun. Image Process.*, Nov. 1992, pp. 1116– 1123.
[42] J. H. D. M. Westerink and K. Teunissen, "Perceived sharpness in
complex moving images," *Displays*, vol. 16, no. 2, pp. 89–97, 1995.
[43] R. Marois and J. Ivanoff, "Capacity limits of information processing in the brain," *Trends Cognitive Sci.*, vol. 9, pp. 296–305, Jun. 2005.
[44] G. Houghton and S. P. Tipper, "Inhibitory mechanisms of neural and cognitive control: Applications to selective attention and sequential action," *Brain Cognition*, vol. 30, no. 1, pp. 20–43, Feb. 1996.
[45] C. H. Chou and C. W. Chen, "A perceptually optimized 3-D subband codec for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 143–156, Apr. 1996.
[46] V. Kayargadde and J. B. Martens, "An objective measure for perceived noise," *Signal Process.*, vol. 49, no. 3, pp. 187–206, Mar. 1996.
[47] S. Winkler, "Quality metric design: A closer look," in *Proc. 5th SPIE Hum. Vis. Electron. Imaging*, Jun. 2000, pp. 37–44.
[48] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proc. IEEE ICIP*, Oct. 2006, pp. 2945–2948.
[49] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D. Pearson, "Measurement of scene-dependent quality variations in digitally coded television pictures," *IEE Proc. Vis. Image Signal Process.*, vol. 142, no. 3, pp. 149–154, Jun. 1995.
[50] D. S. Hands and S. E. Avons, "Recency and duration neglect in
subjective assessment of television picture quality," *Appl. Cognitive
Psychol.*, vol. 15, no. 6, pp. 639–657, 2001.
[51] J. R. Anderson and M. Matessa, "A production system theory of serial memory," *Psychol. Rev.*, vol. 104, no. 4, pp. 728–748, 1997.
[52] A. E. Burgess and H. Ghandeharian, "Visual signal detection: II. Signallocation identification," *J. Opt. Soc. Am. A*, vol. 1, no. 8, pp. 906–910, Aug. 1984.
[53] D. Walther and C. Koch, "Modeling attention to salient Proto-objects," *Neural Netw.*, vol. 19, no. 9, pp. 1395–1407, Nov. 2006.
[54] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Do video coding impairments disturb the visual attention deployment?" *Signal Process. Image Commun.*, vol. 25, no. 8, pp. 597–609, Sep. 2010.
[55] Z. Chen and C. Guillemot, "Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 806–819, Jun. 2010.

[56] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. 5th SPIE Hum. Vis. Visual Process. Digit. Display*, May 1994, pp. 127–139.
[57] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *J. Opt. Soc. Am. A*, vol. 14, no. 9, pp. 2379–2391, Sep. 1997.
[58] R. W. Bowen, J. Pola, and L. Matin, "Visual persistence: Effects of flash, luminance, duration and energy," *Vis. Res.*, vol. 14, no. 4, pp. 295–303, Apr. 1974.
[59] G. E. Meyer and W. M. Maguire, "Spatial frequency and the mediation of short-term visual storage," *Science*, vol. 198, pp. 524–525, Nov. 1977.
[60] J. J. Mezrich, "The duration of visual persistence," *Vis. Res.*, vol. 24, no. 6, pp. 631–632, 1984.
[61] W. J. Tam, L. B. Stelmach, L. Wang, D. Lauzon, and P. Gray, "Visual
masking at video scene cuts," in *Proc. 6th SPIE Hum. Vis. Visual*
Process. Digit. Display, Apr. 1995, pp. 111–119.
[62] G. Buchsbaum, "An analytical derivation of visual nonlinearity," *IEEE*
Trans. Biomed. Eng., vol. 27, no. 5, pp. 237–242, May 1980.
[63] F. de Simone, M. Tagliasacchi, M. Naccari, S. Tubaro, and T. Ebrahimi, "H.264/AVC video database for the evaluation of quality metrics," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 2430–2433.
[64] (2000). *Final Report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment* [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv−phaseI
[65] Z. Wang, E. P. Simoncelli, A. C. Bovik, and M. Matthews, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals Syst. Comput.*, Nov. 2003, pp. 1398–1402.
[66] H. R. Sheikh and A. C. Bovik, "A visual information fidelity approach to video quality assessment," in *Proc. 1st Int. Conf. Video Process. Quality Metrics Consumer Electron.*, Jan. 2005, pp. 23–25.
[67] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, pp. 121–130.
[68] N Graham. Visual Pattern Analyzers. Oxford University Press, 1989.
[69] D J Heeger. Normalization of cell responses in cat striate cortex. Visual Neuroscience, 9:181{198, 1992a.
[70] J P Jones and L A Palmer. An evaluation of the two-dimensional Gabor lterative model of simple receptive fields in cat striate cortex. Journal of Neurophysiology, 58:1233{1258, 1987c.
[71] S Marcelja. Mathematical description of the response of simple cortical cells. Journal of the Optical Society of America A, 70:1297{1300, 1980.
[72] R A Young. The Gaussian derivative model for spatial vision: I. Retinal mechanisms. Spatial Vision, 2:273{ 293, 1987.
[73] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger. Shiftable multi-scale transforms. IEEE Transactions on Information Theory, Special Issue on Wavelets, 38:587{607, 1992.
[74] P Teo and D J Heeger. Perceptual image distortion. In Proceedings of SPIE, volume 2179, pages 127{141, San Jose, CA, Feb 1994.
[75] W T Freeman and E H Adelson. The design and use of steerable filters. IEEE Pattern Analysis and Machine Intel ligence, 13:891{906, 1991.

BIOGRAPHIES

Mrs. D. Neelima is B.Tech(CSE), M.Tech(CSE) from JNTU Hyderabad and currently pursuing P.hd in JNTU Kakinada , Andhra Pradesh, India. She is working as Assistant professor in Computer Science & Engineering department in Jawaharlal Nehru Technological University Kakinada , Andhra Pradesh, India. She has 4 years of experience in teaching Computer Science and Engineering related subjects . She is a research scholar and her area of interest and research includes Video Image Processing. She has published several Research papers out of which 2 are international Journals and 2 are international conferences. She has guided more than 60 students of Bachelor degree, 20 Students of Master degree in Computer Science and Engineering in their major projects.

I am **Shaika Sultana** doing M.Tech in Jawaharlal Nehru Technological University, Kakinada ,A.P.,and interesting research area is Video Image Processing.