Visual Distortion Assessment With Emphasis on Spatially Transitional Regions

EePing Ong, Weisi Lin, Zhongkang Lu, Susu Yao, and Minoru Etoh

***Abstract*—It is known that the human visual system (HVS) does not pay equal attention to each error and even region in judging picture quality. In this paper, we combine a perceptual model with an integrated detection of the spatially transitional regions in visual distortion evaluation to better match the HVS perception to visual quality. For decompressed images or video, the spatially transitional regions are the regions where major perceptually disturbing artefacts caused by edge impairments (mainly due to blurring and locations where the edge information is not adequately represented) and the presence of false edges (mainly due to blockiness and the presence of strong rippling effects of ringing) usually occur. Such regions are efficiently detected based on a single two-dimensional spatial high-pass filter in our work. Good correlation between the proposed method and the human perception has been demonstrated with the full set of 50-Hz video quality expert group test data.**

***Index Terms*—Blockiness, blurring, edge impairment, false edges, high-pass spatial filter, ringing, visual perceptual distortion.**

I. Introduction

BESIDES online and offline visual quality evaluation, how distortion is gauged also plays a determinative role in shaping most algorithms for image and video manipulations, such as enhancement, reconstruction, data hiding, compression, and joint source/channel coding. Visual quality control within an encoder or distortion assessment for a decoded signal is particularly of interest due to the widespread applications of H.26x/MPEG-x compression and coding. Since human eyes are the end receiver of most decoded images and video, it is desirable to develop visual quality metrics that correlate better with human eyes' perception than the conventional pixel-wise error [e.g., mean square error (MSE) and peak SNR (PSNR)] measures.

A number of approaches have been tried to model the temporal, spatial, and masking characteristics of human vision [1]–[4], to evaluate common coding artefacts [5]–[7], and to combine these two paradigms [8], [9].

Perceptual models based upon human vision characteristics can be constructed [1]–[4]. In the metric proposed in [1] and [2], the color-transformed original and decoded sequences are subjected to blocking and the discrete cosine transform (DCT), and the resultant DCT coefficients are then converted to the local contrast, which is defined as the ratio of the ac amplitude to the temporally low-pass filtered dc amplitude. A temporal recursive discrete second-order IIR filtering operation follows to implement the temporal part of the contrast sensitivity function (CSF). The results are then converted to measures of visibility by dividing each coefficient by its respective visual spatial threshold. The difference of two sequences is subjected to a contrast masking operation, and finally the masked difference will be pooled over various dimensions to illustrate perceptual error. With the same paradigm, Winkler's metric [3], [4] consists of color conversion, temporal filters, spatial subband filters, contrast control, and pooling for various channels, as an attempt to emulate the spatio-temporal mechanisms in the human visual system. The difference between original and decoded video is evaluated to give an estimate of visual distortion of the decoded signal. The metric's parameters were determined by fitting the metric's output to the experimental data on human eyes [10]–[14].

DCT-based coding (as in H.261/263, MPEG-1/2/4, JPEG, and the emerging H.264) introduces specific types of artefacts [27], [5], [6], [8], [9], such as, blockiness, ringing, and blurring. The metrics presented in [5] and [6] evaluate blocking artefacts as the distortion measure. The metric in [7] measured five types of error (i.e., low-pass filtered error, Weber's law and CSF corrected error, blocking error, correlated error, and high-contrast transitional error), and used principal component analysis (PCA) to decide the compound effect on visual quality. In [9], switching is suggested between a perceptual model and a blockiness detector depending on the video under test. In [8], a modified version of the perceptual model proposed in [25] and [26] is applied to blockiness dominant regions.

Perceptual models are usually applied with equal weightings/emphasis to the whole frame of an image, and this still follows the methodology similar to the conventional MSE or PSNR measures. In fact, a more effective metric is possible with unequal weightings/emphasis toward visual data available, because:

1) the human vision system does not process and perceive all of the information equally [15], [10], [11], [12];
2) excessive perceptually insignificant data (due to masking effects [1], [14]) may disturb the measurement and therefore hinder the effectiveness of such metrics.

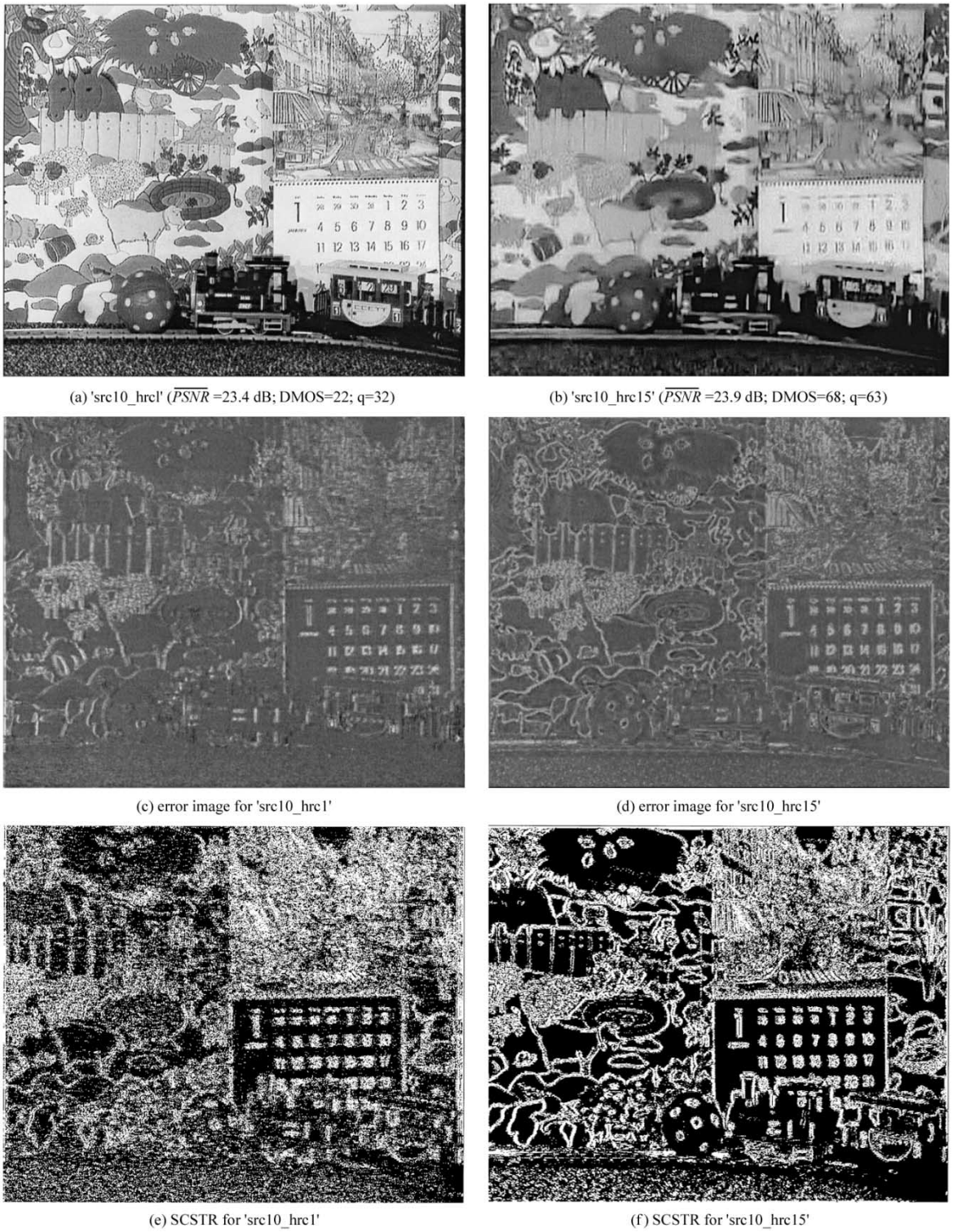
Appropriate manipulation of weightings/emphasis of data processed from input images/video is an interesting issue for visual distortion assessment. The aforementioned application of a perceptual model to certain regions in an image (such as in [8]) suggests a possibility along that direction. Edge information is found to be of primary importance in visual perception [16], and this is also evident from the fact that a sketch image conveys the most visual structuring information in the scene. Damage in object boundaries tends to introduce more notable quality degradation than that in textured regions. Moreover, edgelets, together with the common coding artefacts such

Manuscript received September 13, 2002; revised November 11, 2003. This paper was recommended by Associate Editor A. Kot.

E. Ong, W. Lin, Z. Lu, and S. Yao are with the Institute For Infocomm Research, Singapore 119613 (e-mail: epong@i2r.a-star.edu.sg).

M. Etoh was with the Signal Processing Laboratory, NTT DoCoMo, Inc., Kanagawa, 239-8536 Japan. He is now with DoCoMo Communications Laboratories USA, San Jose, CA 95110 USA.

Digital Object Identifier 10.1109/TCSVT.2004.825574



(a) 'src10_hrc1' ($\overline{PSNR}$ =23.4 dB; DMOS=22; q=32)

(b) 'src10_hrc15' ($\overline{PSNR}$ =23.9 dB; DMOS=68; q=63)

(c) error image for 'src10_hrc1'

(d) error image for 'src10_hrc15'

(e) SCSTR for 'src10_hrc1'

(f) SCSTR for 'src10_hrc15'

Fig. 1. Perceptual quality affected by edge impairment.

as blockiness and ringing, are associated with high spatial transition. Blockiness and ringing artefacts were detected for visual distortion metrics in [8] and [24], respectively. We believe that a good metric should consider multiple factors in a cost-effective manner. A scheme is hence proposed in this paper to detect artefacts such as those caused by the presence of false edges (e.g., blocking and ringing artefacts) and edge impairment (e.g., caused by blurring) in a single step via two-dimensional (2-D) high-pass spatial filter for the purpose of visual distortion gauge. Here, edge impairment refers to impairment of the edgels in decoded video, of which blurring is the major form of impairment. The scheme is able to work for situations where not all of the artefacts are present in a particular image or video (depending upon the coding scheme or allowed bit rate, for example). In the remainder of this paper, Section II presents the detection of changes in the spatially transitional regions for presence of

TABLE I
50-Hz VQEG SEQUENCES

Source No. (src)	1	2	3	4	5	6	7	8	9	10
sequence	Tree	Barcelona	Harp	Moving graphic	Canoa Valsesia	F1 Car	Fries	Horizontal scrolling 2	Rugby	Mobile & calendar

TABLE II
VQEG HYPOTHETICAL REFERENCE CONDITIONS

HRC #	Bitrate	resolution	Codec	coding quality
16	1.5 Mb/s	CIF	H.263	low
15	768 kb/s	CIF	H.263	low
14	2 Mb/s	3/4	mp@ml	low
13	2 Mb/s	3/4	sp@ml	low
12	4.5 Mb/s	full	mp@ml	low
11	3 Mb/s	full	mp@ml	low
10	4.5 Mb/s	full	mp@ml	low
9	3 Mb/s	full	mp@ml	low/high
8	4.5 Mb/s	full	mp@ml	low/high
7	6 Mb/s	full	mp@ml	high
6	8 Mb/s	full	mp@ml	high
5	8 & 4.5 Mb/s	full	mp@ml	high
4	19/PAL(NTSC)-19/PAL(NTSC)-12 Mb/s	full	422p@ml	high
3	50-50-…-50 Mb/s	full	422p@ml	high
2	19-19-12 Mb/s	full	422p@ml	high
1	n/a	full	n/a	high

false edges and edge impairment, while Section III gives a modified version of the perceptual model [3], [4] to be applied to the said regions. Section IV demonstrates the performance of the proposed scheme and presents the comparison with PSNR and the model in [3], [4], using the full set of Video Quality Expert Group (VQEG) [17] 50-Hz test sequences and their associated subjective test results. Section V concludes this paper.

II. DETECTION OF CHANGES IN SPATIALLY TRANSITIONAL REGIONS

Human eyes are highly sensitive to distortion of edges. In Fig. 1(a) and (b), a frame is shown for each of two decoded video sequences whose quality have been evaluated by human observers in the VQEG experiments [17] against the original sequence ("Mobile&Calendar"). The test combination 'src10_hrc15' (src# refers to the sequence as in Table I, while hrc# indicates the test condition (different bitrate and codec as in Table II) exhibits much more severe damage on edge than 'src10_hrc1' (it is obvious with the calendar, animals, fences, and other areas). This is also evident from Fig. 1(c) and (d) where more edge energy can be seen in the error image of 'src10_hrc15' than that of 'src10_hrc1.' The subjective visual quality is indicated with mean opinion scores (MOSs) [17] given by human observers. Difference MOS (DMOSs) [17] refers to the difference of MOSs between the reference video and the processed video and typically ranges from 0 (the highest quality) to 100 (the lowest quality). As expected, the DMOS for 'src10_hrc1' is significantly lower than that for 'src10_hrc15' (22 and 68, respectively), although the average PSNR for the two test combinations is similar (in fact, 'src10_hrc15' is even 0.5 dB higher). Thus, it is the concentration of error energy on edge that causes degradation of perceptual quality.

Blockiness, edge impairments (such as blurring), and ringing are annoying and may be respectively detected (e.g., [5]–[9] for blockiness and [24] for ringing). However, our interest here is to use a single detector for multiple artefacts for implementation efficiency. This is possible since blockiness, blurring, and ringing are associated with high-spatial-frequency bands. Spatial transitional regions can be detected using a single 2-D high-pass (HP) filter, which can be designed to include edge impairments (such as blurring) and the presence of false edges (such as blockiness and the strong rippling effects of ringing). The filter need not distinguish different orientations. The 2-D HP filter in [18] and [19] has been used in the current work, since the same filter band has been used for the spatial decomposition in the adopted perceptual model (as will be described Section III).

Suppose the results of the original and the corresponding distorted frames of video after the HP filtering are $\{s(x,y)\}$ and $\{s'(x,y)\}$, as shown in Fig. 4, respectively, where (x,y) denotes the position of a pixel. A bigger value for $s(x,y)$ or $s'(x,y)$ signals a higher extent of spatial transition in the vicinity. We are most interested in the changes in the spatial transitional regions

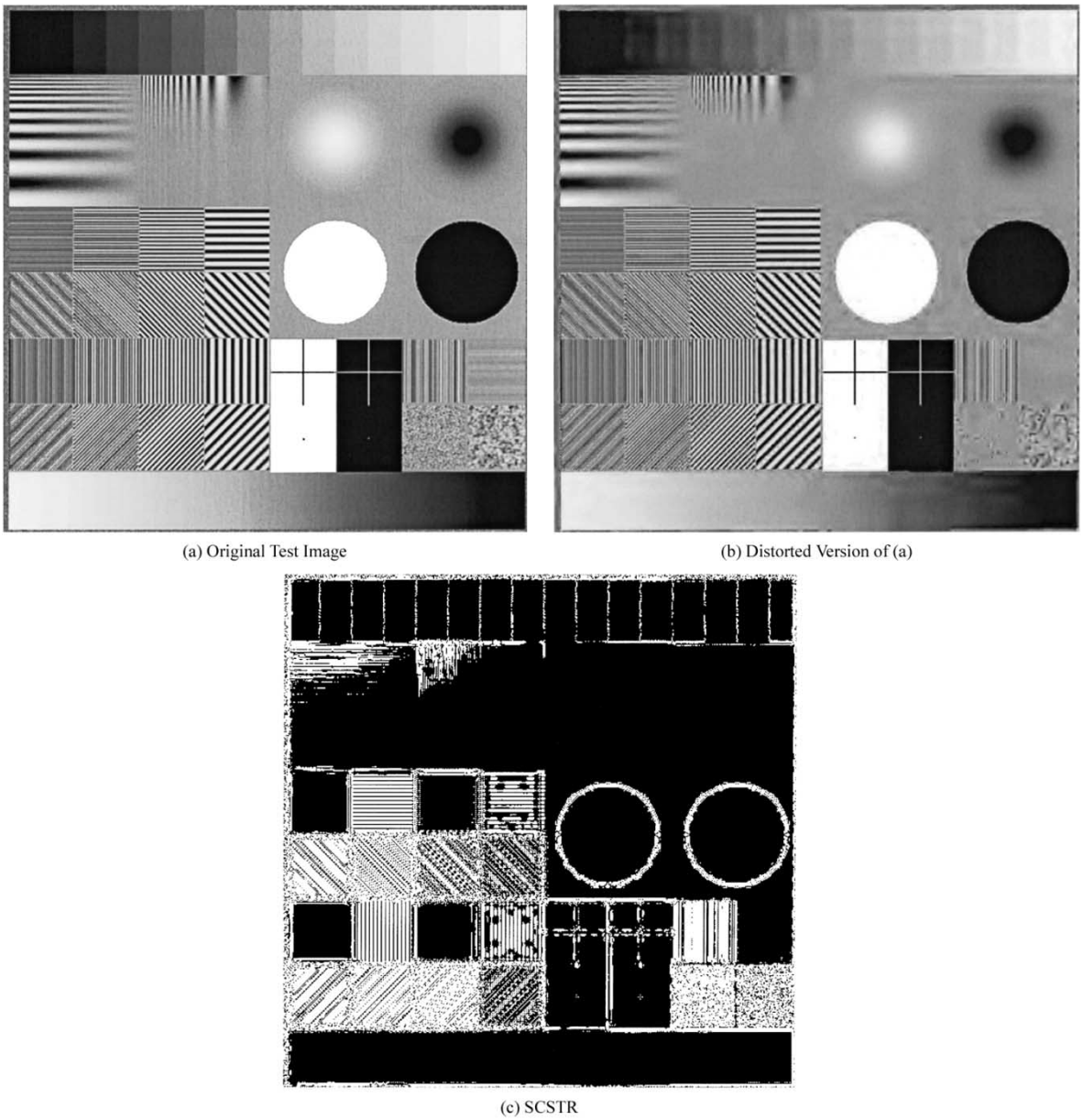


Fig. 2. Extraction of edge impairment and ringing effects.

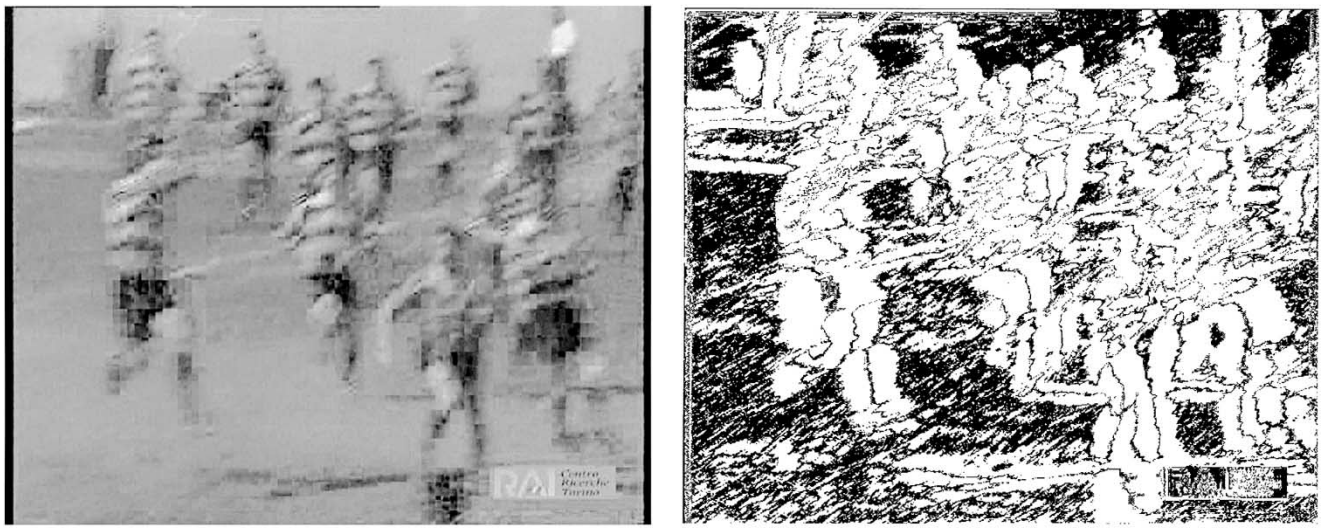
since they are caused mostly by damaged (and/or blurred) edges, ringing, and blockiness. Significant changes in the spatial transitional regions (SCSTRs) can be determined as follows:

$$p(x,y) = \begin{cases} 1, & \text{if } e(x,y) > e_1 \\ 0, & \text{if } e(x,y) < e_0 \\ (e(x,y) - e_0)/(e_1 - e_0), & \text{otherwise} \end{cases} \quad (1)$$

where $e(x,y) = |s(x,y) - s'(x,y)|$, e_1 is the upper threshold for the highband errors, and e_0 is the value where the highband error is regarded as insignificant. If e_0 is set to zero, changes are detected throughout the whole image. Absolute value is taken for $e(x,y)$ because we do not need to distinguish edge impairments from the presence of false edges (although, in theory, we can distinguish edge impairments from presence of false edges since a positive value of $s(x,y) - s'(x,y)$ indicates edge impairments (e.g., due to blurring) while a negative value of $s(x,y) - s'(x,y)$ indicates the presence of false edges (e.g., due to blockiness or ringing). A higher $p(x,y)$ means that an error in location (x,y) carries more weight in overall distortion score.

In Fig. 1(e) and (f), regions with edge impairment are located and relatively bigger $p(x,y)$ values are assigned to relevant areas for 'src10_hrc15' due to the severer edge impairment, and the experiments with the proposed metric in this paper show that this difference in $p(x,y)$ leads to the distortion assessment that is more consistent with the DMOS in these two cases (the distortion measurement q in these cases are 32 and 63, respectively). In Fig. 2(a) and (b) shows a standard test image and its distorted version, while (c) depicts the SCSTR that reflects edge impairment. Fig. 3(a) and (b) shows a decoded frame in 'src9_hrc9' ("Rugby") and the detected SCSTR indicating the occurrence of blockiness and edge impairment.

Appropriate detection of SCSTRs enables visual distortion assessment to be conducted in perceptually significant regions



(a) A decoded frame 'src9_hrc9' (b) SCSTR

Fig. 3. Detection of blocking effects and edge impairment.

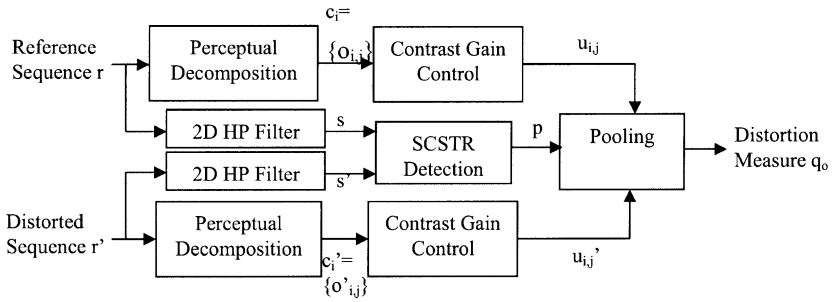


Fig. 4. Block diagram of the proposed perceptual distortion metric.

for both effectiveness and efficiency and provides a proper local significance indicator for distortion evaluation.

III. Perceptual Model

The perceptual model [3], [4] has been modified and then applied to areas in the image indicated by the nonzero values defined in (1). Fig. 4 depicts the block diagram of the proposed metric.

Perceptual decomposition is required to emulate the sensitivity for different frequencies and orientations in the human visual system. We adopt the similar temporal and spatial filtering as used in [3] and [4], and only the luminance component of the input video sequences has been used (as in [29] and [8]) for efficiency. The luminance component is used because it plays a much more important role in human visual perception than chrominance components do. A temporal FIR filter with a length of nine taps is designed with the following impulse response:

$$h(t) = e^{-(\frac{\ln(t/\tau)}{\sigma})^2} \tag{2}$$

to match the sustained mechanism (the transient effect is ignored) of human vision [20], where $\tau = 160$ ms and $\sigma = 0.2$.

The steerable pyramid transform introduced in [18] and [19] is utilized to perform spatial band-pass decomposition to approximate the characteristics of neurons in the human primary visual cortex. The original signal r and the distorted signal r' are decomposed into five frequency levels, resulting in $\{c_i\}$ and $\{c_i'\}, i = 1, 2, \ldots, 5$, respectively, where Level 1 corresponds to the highest resolution image and the increment of i by 1 means a 2:1 pixel down-sampling in both dimensions. Each frequency level is further divided for four orientations: $\{o_{i,j}\}$ and $\{o_{i,j'}\}, j = 1, 2, 3, 4$ for 0°, 45°, 90°, and 135°. Thus, there are a total of 20 subbands as labeled in Fig. 5. The spatial decom-

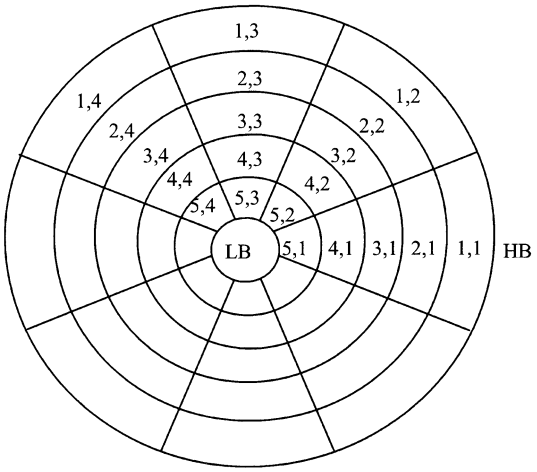


Fig. 5. Illustration of spatial decomposition (labeled with subband indices i, j).

TABLE III
MODEL PARAMETERS

e_0	e_1	b	d	w_1	w_2	w_3	w_4	w_5
6.5	3.5	2.0	0.6	371.8	16.0	881.5	1300.0	2100.0

position also produces a low-pass band (LB) and a high-pass band (HB) which is used for detection of the spatial transitional regions (as described in Section II). In [3] and [4], the weights of each level's output was determined using experimental data of human sensitivity [10]–[12], which were measured near the human visibility threshold under strict control of the environment (e.g., eye movement was compensated). There has been no evidence that when distortion is above the visibility threshold (the usual case in decoded images/video) the human perception behaves similarly as when distortion is near the visibility threshold. Besides, these data were derived from only pure sine wave stimuli. In this study, as in [8], subjective test data toward decoded video are to be used to optimize the weighting parameters, $\{w_i\}$, because such data are more meaningful with the normal viewing condition.

The contrast gain control in this work aims at emulating the masking between subbands with different orientations at the same level (because masking is much more significant with a similar spatial frequency), and therefore a modified model from that in [1] is used to derive the output of the process as

$$u_{i,j}(x,y) = \frac{d + o_{i,j}(x,y)^2}{b^2 + \sum_{\substack{l=1 \\ (l \neq j)}}^{4} o_{i,l}(x,y)^2} \tag{3a}$$

$$u'_{i,j}(x,y) = \frac{d + o'_{i,j}(x,y)^2}{b^2 + \sum_{\substack{l=1 \\ (l \neq j)}}^{4} o'_{i,l}(x,y)^2} \tag{3b}$$

where d is added for generality, and d and b are to be determined later by fitting the metric's responses to subjective ratings data.

The goal of the pooling stage is to emulate the human vision process in which the information represented in the various channels within the primary visual cortex is integrated in the brain areas. Structured pooling guided by the SCSTR is performed over the simplest Minkowski distance of all subbands to yield the metric's output as follows:

$$q_o = \frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{5}\frac{1}{n_i^t}\sum_{j=1}^{4}\sum_{x}\sum_{y}\left|p_i^t(x,y) \times \left(u_{i,j}^t(x,y) - u_{i,j'}^t(x,y)\right)\right| \tag{4}$$

where t indicates the frame under evaluation and N is total number of frames, $p_1^t\ (x,y)$ is calculated with highest resolution using (1) first and $p_i^t(x,y)\ (i > 1)$ can be calculated by down-sampling by 2 in each dimension consecutively, and n_i^t is the number of pixels in the ith level of the tth frame.

IV. PERFORMANCE AND COMPARISON OF THE PROPOSED METRIC

Performance is measured by comparing the metric output q_o with the DMOS between the original and distorted sequences. To facilitate monotonicity of prediction and a common analysis space of comparison, q_o is fitted via a four-parameter cubic polynomial [21] to the corresponding DMOS as

$$q = a_o + a_1(q_o) + a_2\left(q_o^2\right) + a_3\left(q_o^3\right). \tag{5}$$

In [17], the output of a metric (e.g., the PSNR, Winkler's metric) was fitted to either this cubic polynomial function or a four-parameter logistic curve (also defined in [21]) depending on which one achieves the best fit.

The Pearson correlation, which measures the prediction accuracy, i.e., the ability of a metric to predict subjective ratings, is defined as

$$r_p = \frac{\sum_k (q_k - \bar{q})(\mathrm{DMOS}_k - \overline{\mathrm{DMOS}})}{\sqrt{\sum_k (q_k - \bar{q})^2}\sqrt{\sum_k (\mathrm{DMOS}_k - \overline{\mathrm{DMOS}})^2}} \tag{6}$$

where $\bar{q}$ and $\overline{\mathrm{DMOS}}$ are the means of q and DMOS, and k is the index for the video sequence under test.

Spearman rank-order correlation, which measures the prediction monotonicity, i.e., whether the increases/decreases in one variable are associated with increases/decreases in the other variable independent of the magnitude of the increase/decrease, is defined as

$$r_s = \frac{\sum_k (\chi_k - \bar{\chi})(\gamma_k - \bar{\gamma})}{\sqrt{\sum_k (\chi_k - \bar{\chi})^2}\sqrt{\sum_k (\gamma_k - \bar{\gamma})^2}} \tag{7}$$

where χ_k is the rank of q_k and γ_k is the rank of DMOS_k in the ordered data series, and $\bar{\chi}$ and $\bar{\gamma}$ are the respective midranks. In the ideal match between a metric's output and DMOS, $r_p = 1$ and $r_s = 1$.

In order to determine the process parameter set, $\Omega = [e_0, e_1, b, d, w_1, \ldots, w_5]$, five 50-Hz VQEG [17] test sequences (src2, src3, src7, src9, and src10) (each with 16 test conditions) and the associated DMOSs have been used. The 30th and 120th frames are used for each test combination. The process is formulated as

$$\hat{\Omega} = \arg\max_{\Omega} \quad r_p(\Omega). \tag{8}$$

The efficient Hooke and Jeeves' optimization method [22] has been used because it is computationally simple and the process is approximated by tuning a variable at a time (as the application in [23]). The resultant parameters are listed in Table III.

The proposed metric is tested against the DMOS with all ten 50-Hz VQEG video sequences [17] (src1–10 in Table I). The frame size of all sequences is 720×576 pixels. Each sequence is evaluated in all 16 test conditions (hrc1–16 in Table II) which define a wide spectrum of bitrates for several codecs.

Table IV shows the r_p and r_s comparison for all 160 test combinations for the proposed metric against PSNR and the result of Winkler's method [3], [4] and confirms the better performance of the proposed approach. The upper and lower bounds indicate the 95% confidence interval. The results for PSNR and Win-

TABLE IV
CORRELATION MEASURES WITH RESPECT TO SUBJECTIVE RATINGS FOR VQEG 50-HZ SEQUENCES

	r_p			r_s
	correlation	upper bound	lower bound	
PSNR	0.78*	0.84	0.72	0.81*
Winklers metric	0.70*	0.77	0.61	0.71*
Proposed metric	0.84	0.88	0.78	0.83

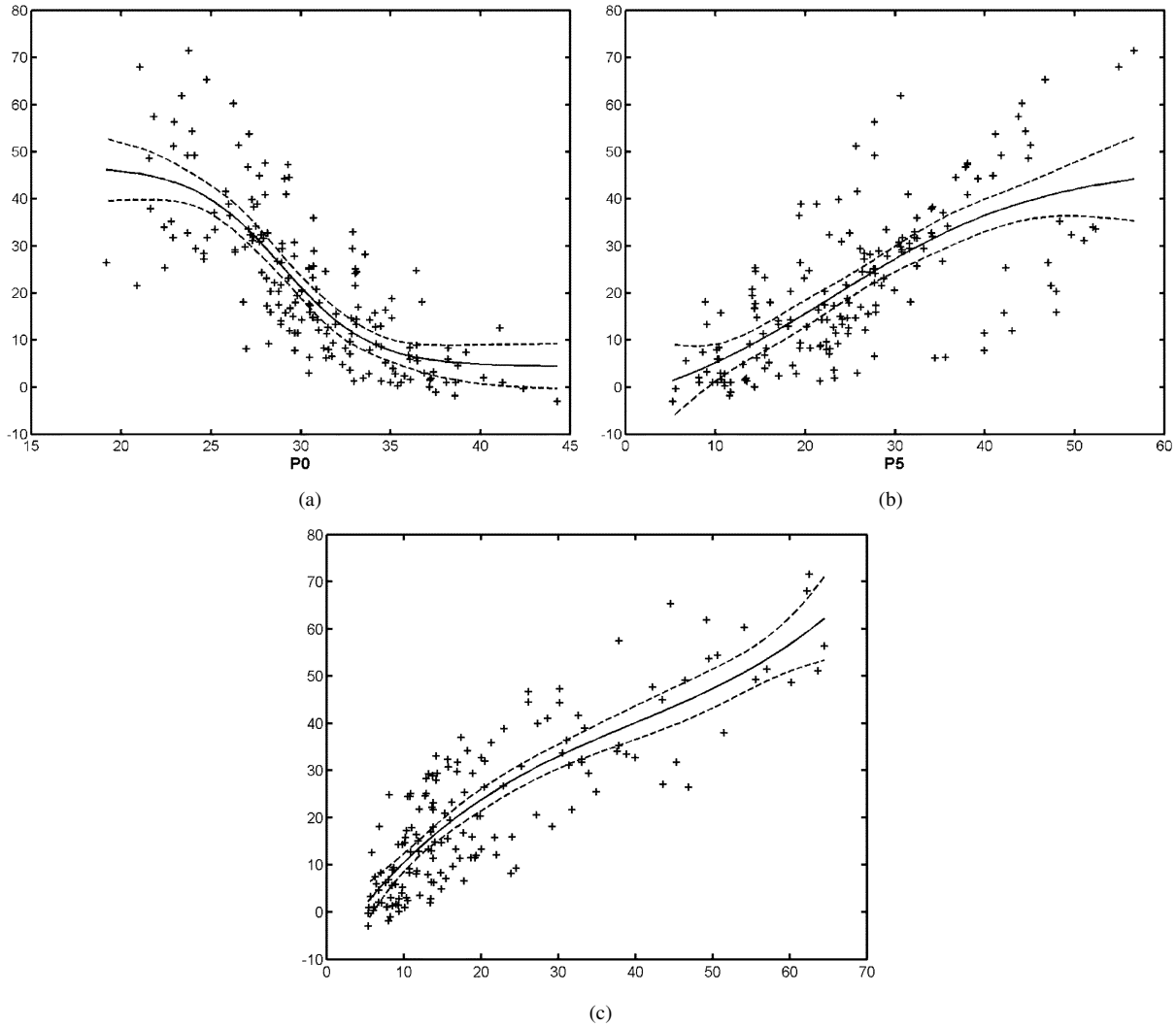


Fig. 6. (a) Scatterplot of DMOS versus PSNR. (b) Scatterplot of DMOS versus output of Winkler's metric. (c) Scatterplot of DMOS versus output of proposed metric.

kler's method are extracted from [17], while the results for the proposed scheme are calculated from the VQEG normalized TV-format sequences, which are also the sequences shown to human subjects for performing subjective ratings.

Fig. 6 shows the scatter plots of PSNR, Winkler's metric, and the proposed metric against DMOS. The proposed metric exhibits significantly less outliers than PSNR and Winkler's metric, and this is the reason for its better overall performance.

V. CONCLUSION

A new approach for evaluating perceptual distortion for visual signals has been proposed, based on the combination of a perceptual model and the integrated detection of the major perceptually disturbing artefacts (namely, edge impairment and the presence of false edges) in decoded images. Emphasis on perceptually critical regions is in line with the characteristics of the human visual perception and may also help to achieve the efficiency necessary for many practical applications. Regions with edge impairment (mainly blurring) and the presence of false edges (mainly blockiness and strong rippling effects of ringing) are found via a single spatial high-pass filter and form the basis for a perceptual model to be applied. Experiments with a full set of 50-Hz VQEG test data show that the proposed scheme assesses visual distortion with higher relevancy to the subjective test results.

The proposed scheme places emphasis on edge impairment and the presence of false edges as the major artifacts in the visual signal being evaluated. We know that these artifacts are the major ones affecting picture quality as far as the prevalent compression methods are concerned. Note that not all the aforementioned artefacts have to be present for evaluation. Although the scheme has been tested with block-coded video, it is expected to work for video without blockiness artifacts (e.g., for wavelet-coded video) since the detected transitional regions are then with ringing and blurring (still the major artifacts affecting picture quality there). After all, spatially transitional regions are where human eyes pay more attention in general when judging a picture for its quality.

REFERENCES

[1] A. B. Watson and J. A. Soloman, "A model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A*, vol. 14, no. 9, pp. 2379–2391, 1997.

[2] A. B. Watson, J. Hu, and J. F. McGowan III, "DVQ: A digital video quality metric based on human vision," *J. Electron. Imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[3] S. Winkler, *Vision Models and Quality Metrics for Image Processing Applications*, Lausanne, Switzerland: Ecole Polytecnique Federale De Lausanne (EPFL), Swiss Federal Inst. of Technol., Dec. 2000, thesis 2313.

[4] ——, "A perceptual distortion metric for digital color video," *Proc. SPIE*, vol. 3644, pp. 175–184, 1999.

[5] H. R. Wu and M. Yuen, "A generalize block-edge impairment metric for video coding," in *IEEE Signal Processing Lett.*, vol. 4, Nov. 1997, pp. 317–320.

[6] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in image based on human visual sensitivity," in *IEEE Trans. Image Processing*, vol. 4, Sept. 1995, pp. 713–724.

[7] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Trans. Commun.*, vol. 46, pp. 1215–1225, Sept. 1998.

[8] Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision-model-based impairment metric to evaluate blocking artifacts in digital video," *Proc. IEEE*, vol. 90, pp. 154–169, Jan. 2002.

[9] K. T. Tan and M. Ghanbari, "A multi-metric objective picture-quality measurement model for MPEG video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 1208–1213, Oct. 2000.

[10] D. H. Kelly, "Motion and vision I: Stabilized images of stationary gratings," *J. Opt. Soc. Amer.*, vol. 69, no. 9, pp. 1266–1274, 1979.

[11] ——, "Motion and vision II: Stabilized spatio-temporal threshold surface," *J. Opt. Soc. Amer.*, vol. 69, no. 10, pp. 1340–1349, 1979.

[12] ——, "Spatiotemporal variation of chromatic and achromatic contrast thresholds," *J. Opt. Soc. Amer.*, vol. 73, no. 6, pp. 742–750, 1983.

[13] J. M. Foley, "Human luminance pattern-vision mechanisms: Masking experiments require a new model," *J. Opt. Soc. Amer. A*, vol. 11, no. 6, pp. 1710–1719, 1994.

[14] E. Switkes, A. Bradley, and K. K. De Valois, "Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings," *J. Opt. Soc. Amer. A*, vol. 5, no. 7, pp. 1149–1162, 1988.

[15] W. James, *The Principles of Psychology*. Cambridge, MA: Harvard Univ. Press, 1980/1981.

[16] X. Ran and N. Farvardin, "A perceptually motivated three-component image model-Part I: Description of the model," *IEEE Trans. Image Processing*, vol. 4, pp. 401–415, Apr. 1995.

[17] Final Report from the Video Quality Expert Group on the Validation of Objective Models of Video Quality Assessment (2000, Mar.). [Online]. Available: http://www.vqeg.org

[18] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inform. Theory*, vol. 38, pp. 587–607, Mar. 1992.

[19] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 891–906, Sept. 1991.

[20] R. E. Fredericksen and R. F. Hess, "Temporal detection in human vision: Dependence on stimulus energy," *J. Opt. Soc. Amer. A*, vol. 14, pp. 2557–2569, Oct. 1997.

[21] Evaluation of New Methods for Objective Testing of Video Quality: Objective Test Plan (1998, Sept.). [Online]. Available: www.vqeg.org

[22] R. Hooke and T. A. Jeeves, "Direct-Search solution of numerical and statistical problems," *J. Assoc. Comput. Machinery*, vol. 8, pp. 212–229, 1961.

[23] W. Lin, "Task division for parallel implementation of object identification system based on alternating hypothesize-verify-extend strategy," *Concurrency: Practice and Experience*, vol. 9, no. 9, pp. 859–876, 1997.

[24] Z. Yu, H. R. Wu, and T. Chen, "A perceptual measure of ringing artifact for hybrid MC/DPCM/DCT coded video," in *Proc. IASTED Int. Conf.*, Las Vegas, NV, Nov. 2000, pp. 94–99.

[25] P. Lindh and C. Lambrecht, "Efficient spatio-temporal decomposition for perceptual processing of video sequences," in *Proc. Int. Conf. Image Processing*, Lausanne, Switzerland, Sept. 1996, pp. 331–334.

[26] C. Lambrecht, "Perceptual Models and Architectures for Video Coding Applications," Ph.D. dissertation, Ecole Polytecnique Federale De Lausanne (EPFL), 1996.

[27] M. Yuen and H. R. Wu, "A survey of MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, no. 3, pp. 247–278, Nov. 1998.

[28] VQEG Subjective Test Plan Ver. 3 (1999, July). [Online]. Available: http://www.vqeg.org

[29] S. Winkler, "Quality metric design: A closer look," in *Proc. SPIE Human Vision and Electronic Imaging Conf.*, vol. 3959, San Jose, CA, Jan. 2000, pp. 37–44.