PICTURE QUALITY EVALUATION
BASED ON PIXEL VALUE EMPIRICAL DISTRIBUTION

Irwan Prasetya Gunawan¹ and Mohammed Ghanbari²
¹PT. Telekomunikasi Indonesia, Divisi Regional V, Surabaya, Jawa Timur, Indonesia
²Dept. of Computing and Electronic Systems, University of Essex, Colchester, UK
Email: irwan@ieee.org, ghan@esx.ac.uk

ABSTRACT

In this paper, we present an investigation of a reduced-reference image quality model for blocking artefact measurement based on histogram (empirical pixel distribution) of pixels in a picture. Quality is derived from a comparison between histogram taken from the original and the decoded image, using a couple of histogram dissimilarity measures such as Kolmogorov-Smirnoff and χ² tests. Our results show that the model presented in this paper is promising. The performance of the proposed quality metrics in terms of prediction accuracy is far better than the traditional, de-facto standard Peak Signal-to-Noise Ratio (PSNR).

Keywords : Blockiness, Objective Measurement, Reduced-Reference, Image Quality Assessment.

10 INTRODUCTION

Recently, there has been a growing interest in the use of pictorial-based communication systems. Its success depends on, among other things, the quality of its services. This is obviously leading to the awareness of the perceived quality of digital pictures (image/video). In today’s highly mixed and heterogeneous environment, digital image/video may have been subjected to various processes and delivered through heterogeneous infrastructures, any of which can introduce degradation that may impair the quality of their visual representation. Hence it is important to have reliable methods of image/video quality assessment. To service providers, picture quality metric indicates the level of services that can be offered and delivered to their customers. Conversely, the same metric also provides the customers with reassurances that they get what they have paid for. In addition, it is also desired that the perceived quality can be automatically predicted in an objective manner since subjective evaluation [11] is time-consuming, laborious, and expensive [2].

Regardless of how objective models were designed, many existing quality models share a common goal to mimic human eyes’ perception. To that end, many researchers have come up with quality models that try to incorporate the human visual system (HVS) by several different methods [3]: either by modifying the traditional Mean Squared Error (MSE)-based metrics [4], assigning certain weighting factors in the frequency domain [5], mimicking the HVS through non-linearities and filters [6], or a combination of the aforementioned methods. However, despite the vast amount of literatures on HVS-based models in the literatures, a number of researchers argue that these models still have some disadvantages. This is usually attributed to the dependency of HVS-based models to the factors such as viewing conditions (most notably the viewing distance) [7], compression methods and image processing operations [8].

In practise, not all applications require the use of sophisticated models of the HVS [9]. Instead, an a priori knowledge of the type of distortions introduced in coding, for example, may be sufficient. The criteria to determine the types of the distortion which are important to human viewers is discussed by [10]. Such distortion metrics still have the options to incorporate certain aspects of HVS, though. The trade off between versatility and efficiency of the model will determine the complexity of the distortion-based metrics. Furthermore, of all the distortions known to exist in a compressed image/video sequence described by [11] only some of them maybe sufficient to be analysed. For example, blocking artifacts [12, 13] is the most prominent distortions found in digital images compressed by Discrete Cosine Transform (DCT)-based encoding still used in many today’s consumer digital photography equipment or in the digital television distribution systems.

In this paper, we present a picture quality assessment method based on blockiness metrics by means of descriptive statistical characteristics. An objective metric was developed from a histogram (empirical pixel distribution) which was extracted from a picture. The visual quality of a test image is estimated by comparing its histogram to the histogram of the original image. We propose our method based on the observation that pixel
distributions of digital images compressed at different rates exhibit some patterns that can be related and compared to the distribution of its original counterpart. It will be shown in the subsequent sections of this paper that simple quantity such as descriptive statistics of an image can be exploited to estimate the ‘quality’ of the reconstructed image by employing some of the well known statistical tools.

11 REVIEW

11.1 Objective quality measurement

Objective quality assessment methodologies can be categorised into three different classes according to the extent of the measurement: 1) full-reference (FR) in which the original video is available; 2) reduced-reference (RR) where only a subset or certain features of the original video is available; and 3) no-reference (NR) which relies only on the test video without having any prior knowledge of the original one. Different flavors of quality assessment methods exist because naturally each of them lends itself to diverse applications. Among these objective methods, the widely used traditional metrics for video/image are the PSNR (peak signal-to-noise ratio) or MSE (mean squared error).

PSNR/MSE are full-reference methodologies which are easy to compute. MSE for an image of size $N \times M$ can be defined as

$$MSE = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (I^{(r)}(x,y) - I^{(d)}(x,y))^2$$

where $I^{(r)}(x,y)$ and $I^{(d)}(x,y)$ are luminance values of the original and the decoded pixel, respectively. A variation of MSE by taking the logarithmic of its value is known as Peak Signal-to-Noise Ratio (PSNR), and defined as

$$PSNR(dB) = -10 \log \left( \frac{MSE}{\max \{I(x,y)\}^2} \right) = -10 \log \left( \frac{MSE}{255^2} \right)$$

where $\max \{\ldots\}$ is a function that returns the highest (maximum) value of its input arguments. It is assumed that the maximum value of $I(x,y)$ is $2^8 - 1 = 255$ since the depth of the pixel is 8 bits per pixel.

PSNR/MSE are renowned largely because of their practicality. Unfortunately, they do not always correlate well with the quality perceived by human users [14]. To mitigate such problems, many researchers have developed various methods to modify the PSNR/MSE metrics by incorporating the HVS such that their quality predictions can be improved. The inclusion of HVS, however, has to be done at the expense of increased complexity, which may not be desirable for certain applications. Hence, research on objective methodology usually aims at replacing the MSE or PSNR without resorting to complicated methods.

11.2 Pixel value distribution analysis

Empirical pixel distribution (histogram) may be chosen as an analysis tool because of its simplicity. For a digital image, frequency of occurrence of a particular pixel value can be graphically illustrated by means of an histogram, which is a practical means for the approximation of the probability density function [15] of image/scene. Instead of summarising the complex statistical behaviour of pixel data in a single number (such as mean, median, standard deviation, etc), more facts can be learnt by analysing the probability distribution function or the cumulative distribution function.

We can utilise the histogram to serve quality assessment purposes by comparing two different histograms analytically. This idea is illustrated in Figure 1. In this figure, starting from the top-left corner picture in Figure 1(a) in a horizontal zigzag direction, we have an original picture followed by three JPEG-compressed pictures at different quality factors resulting in various amount of blockiness due to quantization (low, medium, and high blockiness corresponding to quality factor of 77, 55 and 15, respectively). The associated descriptive histogram is given in Figure 1(b); these graphs suggest that there is a relationship between the amount of blockiness in the picture with the empirical distribution of its pixels. We can see that in the picture with severe blockiness the difference between the histograms are prominent; blockiness in the picture is reflected in the spikiness of the histogram.
12 PROPOSED METHOD

12.1 Dissimilarity measures between empirical pixel distribution

An inherent idea in the quality assessment context is to make use of ‘differences’ or ‘dissimilarity’ to predict quality, based on some distance calculation. For example, in this paper we use the dissimilarity between two different empirical distributions as quality indicator. The idea is to use the available existing quantitative statistical test methods to see how similar (or how different) the two empirical distributions are and then infer the quality of the associated images from this.

One of the tests that we found useful is the well-known Kolmogorov-Smirnov (KS) two sample test (see for example, [16]), the result of which can be regarded as the maximum distance between two cumulative distributions; stated mathematically, it is an $L_{\infty}$ norm. Another test that we could also use is the $\chi^2$-test [17], which can be considered as an $L_2$ norm. The $\chi^2$-test is applicable to histograms whereas the KS is defined only to cumulative distributions [18]. Both of these tests are described briefly in the following paragraphs.

For a luminance image $I(x,y)$ of size $M \times N$, we usually have $L_{\text{min}} \leq I(x,y) \leq L_{\text{max}}$ where $L_{\text{min}} = 0$ and $L_{\text{max}} = 255$. We can count the number of occurrences of pixels within certain interval of gray-level values. This interval is designated as $\Delta B$. Mostly, $\Delta B = 1$ which means that in total we have 256 bins. This total number of bin, the bin size, is designated as $B$. From this assumption, we can construct a discrete distribution function $h_I(k)$ of an image as

$$h_I(k) = \begin{cases} \frac{n_k}{MN} & I_{\text{min}} < k < I_{\text{max}}, \\ 0 & \text{elsewhere}. \end{cases} \quad (3)$$

where

$$n_k = \text{card} \left[ \{ I(x,y) | I(x,y) \in [I_{\text{min}} + k\Delta B, I_{\text{min}} + (k+1)\Delta B] \} \right] \quad (4)$$

is the pixel count, i.e. the number of pixel in the digital image with luminance in the interval $[I_{\text{min}} + k\Delta B, I_{\text{min}} + (k+1)\Delta B]$ such that $\sum n_k = MN$, $k \in [0, B - 1]$ and $\text{card} \{ \cdot \}$ is the cardinality of a set. We can see that the distribution function satisfies the necessary condition of $\sum h_I(k) = 1$. In essence, a histogram practically groups the intensity of pixels into separate bins (certain interval).
A test can be performed on the histogram $h_I(k)$ using $\chi^2$-test. It is a statistical method for determining if two distributions were drawn from different distributions. The $\chi^2$-test is given by

$$\chi^2(I, I) = \sum_k \frac{(h_I(k) - h_I(k))^2}{h_I(k) + h_I(k)}.$$ 

The cumulative distribution profile, $H_I(k)$, is given by

$$H_I(k) = \sum_{\ell=1}^k h_I(\ell).$$

The computation of KS test on cumulative distribution profile follows [16]. It computes the maximum difference between the cumulative distribution functions $H_I(k)$ and $H_Î(k)$. This is simply expressed as

$$t = \max I |H_I(k) - H_Î(k)|$$

where $I$ and $Î$ denote the original and the processed (compressed) images, respectively. This method can also be illustrated graphically by plotting these cumulative distributions on the same graph and finding out the largest gap (= difference value) between them. The test statistic, $t$, is a distance measure between two samples; an agreement between these samples is indicated by small value of the KS test statistic. We may expect that when two samples are in close agreement between one another, their discrepancy is also small; i.e., when the processed/degraded image is statistically close to the original, the quality of the former is comparable to the latter.

We observed that when a picture is compressed coarsely (bit rate is low), its distribution shows a significant departure from the original’s. The statistical test value should be high for such pictures. On the other hand, picture with moderate/low compression (as in the case when the target bit rate is high) exhibits distribution similarity with the original. Accordingly, the statistical test value must be low. This is illustrated in Figure 2 for the KS test results. Similar trend is also shown by the $\chi^2$ test. This pattern/behaviour is consistent for natural images, and this is the reason why statistical tests based on $\chi^2$ test or KS two-sample statistical test can be used to take advantage of this fact.

If we denote the quality as $Q$, then we may have an inversely proportional relationship between $Q$ and $t$ or $\chi^2$; i.e.

$$Q_{KS} = \frac{1}{t}$$

$$Q_{\chi^2} = \frac{1}{\chi^2(I, I)}$$

for the KS and $\chi^2$ tests, respectively.

### 12.2 Wavelet-based multiresolution analysis

Wavelet transforms are used to represent signals simultaneously in space and frequency [19]. They can model the early stages of visual systems and have found many uses in vision research (for example, to account for multichannel characteristics of the HVS). The non-parametric tests in the previous sections can also be applied to the coefficients of wavelet decomposition on images. We can use 1- and 2-level wavelet decomposition to the images, and apply the test to each of the resulting sub-bands. The test scheme is still the same as expressed by Eq. (7) or (5), with a little modification by using the appropriate wavelet coefficient distribution:

$$t_{\psi} = \max I \left| H_{\psi(Î)}(k) - H_{\psi(I)}(k) \right|$$

and

$$\chi^2_{\psi}(\psi(I), \psi(Î)) = \sum_k \frac{(h_{\psi(Î)}(k) - h_{\psi(I)}(k))^2}{h_{\psi(Î)}(k) + h_{\psi(I)}(k)}$$

where $\psi \in \{\text{chosen subbands}\}$, and it depends on the context in which decomposition level was performed; for example, in one-stage decomposition depicted in Figure 3, the chosen sub-bands might be one of $LL, HL, LH$, and $HH$ or any combinations thereof, whereas in two-stage decomposition we also have to include those sub-bands coming from the decomposition of $LL$ in the first stage. Accordingly, the quality indexes in the equations (8) and (9) may have to be re-defined as

$$Q_{KS, \psi} = \sum_{\psi} \frac{1}{t_{\psi}}$$

and

$$Q_{\chi^2, \psi} = \sum_{\psi} \frac{1}{\chi^2_{\psi}(\psi(I), \psi(Î))}$$

for the KS and $\chi^2$ tests, respectively.

A variation of the above method is possible; for example, instead of using the coefficient of wavelet transform, we could apply the tests to the
sub-sampled version of the coefficients. In the LL band (in 1-level decomposition), a local average is applied to non-overlapped blocks of certain size (e.g., $4 \times 4$ or $8 \times 8$). Similar sub-sampling operations can also be applied to other sub-bands. In the HH band, however, it may be better to compute the standard deviation of the sub-sampled block since the HH band is very likely to have histogram with positive kurtosis in which standard deviation value would be more meaningful than its average. The rationale of this is because the HH band is the result of high-pass filtering process in the decomposition, and it has been reported that high-pass filtering will produce histograms with positive kurtosis [20]. From here, as we will describe in the experiments, we try several different approaches to pool the test results as an objective quality indicator.

**12.3 Proposed quality method**

Our proposed method is depicted in Figure 3. First, wavelet decomposition were performed to the images. In this paper, 1- and 2-level decompositions, using biorthogonal3.7 wavelet [19], were used. Histograms are then extracted from the coefficients of a wavelet decomposition. For comparison purposes, we also used histograms of the images without decomposition. Sub-sampling by using local averages was computed on the LL band (or to the image itself in the case of the scheme without decomposition), whereas standard deviations were calculated on the other bands (HL, LH, HH for 1-level decomposition). Several sub-sampling block sizes were used in the experiments: $1 \times 1$ (similar to a scheme without sub-sampling), $4 \times 4$, and $8 \times 8$. To construct the pdf/histogram and the empirical cumulative distribution, various bin sizes were chosen (256, 128, 64). Histograms from the original image were compared to the one from the decoded image. We used the KS test and $\chi^2$-test we have described in the previous section for the comparison.

![Figure 3. Proposed method: As an illustration, wavelet decomposition is performed at 1-level decomposition.](image)

**13 PERFORMANCE EVALUATION**

**13.1 Experiments**

Simulation tests were conducted based on the proposed method described in the previous section. The test materials were chosen from a database of images quality assessment [1] which consists of 24 original images that are compressed at five different JPEG quality factor (different rates), ranging from the low quality at quality factor of 15 to the fairly high quality at quality factor of 79, resulting in 120 images in total. JPEG images were chosen because our method was based on our observation on the spikiness (or local peaks) of the histograms which were prominent in pictures contaminated by blocking artifacts. No trainings were required during the experiments. The output of the quality models were the quality index we have defined in the equations (8), (9), (12), and (13). The performance of the output of these objective evaluation methods is compared to the Mean Opinion Score (MOS) value of subjective evaluation in the image database library by calculating their Pearson correlation ($r_p$) and Spearman correlation ($r_s$) [21]. Good performance will be shown by a good agreement between objective measurement results and MOS value.

**13.2 Test on histograms of image without decomposition**

First, we were interested at the results of the tests on histograms/cumulative histograms of the degraded images without wavelet decomposition. Tests were conducted according to equations (7) and (5) for the KS and $\chi^2$ tests, respectively. Different values of bin size were selected; i.e. $B$ [64,128,256]. As a benchmark, we also calculated the PSNR values in Eq.(2) of the degraded images with respect to the original from the same test materials. The correlation performances of the KS Test, $\chi^2$-test, and PSNR were then evaluated by comparing the output of the model (equations (12), (13), and (2), respectively) against the subjective data.

The results were summarised as the bar charts in Figure 4. The correlations of the PSNR for this dataset were quite low; i.e. $r_p(PSNR) = 0.48$ and $r_s(PSNR) = 0.47$. These values were drawn as the horizontal line in each chart. The two proposed methods were better than the PSNR; however, the predictive power of the scheme using the $\chi^2$-tests was far superior than the one using the KS test. The results also revealed that the performances of the model with the $\chi^2$-test were strongest when $B = 256$ were used, and significantly lower when $B = 64$. 
However, larger bin size is not preferable from the RR point of view because it implies larger RR data rate. A compromise between the predictive power and the bin size were achieved when $B = 128$; Figure 4 shows that the correlations of the proposed $\chi^2$ based model with $B = 128$ were not that far away from the same model using $B = 256$.

We also applied the method to the sub-sampled version of the image; i.e., the histograms were computed to the non-overlapped block average values of the image, not the image itself. The results, however, did not show any significant improvements over the one for the full-resolution images. As an example, the model with $8 \times 8$ non-overlapped block average and $B = 256$ can only achieve $r_p = 0.77$.

This time, since block processing the image in the previous results has shown some gains, albeit small, we also apply the method to the sub-sampled version of the coefficients using $4 \times 4$ block average.

In the test with 1-level decomposition, we performed the analysis to each of the LL, HL, LH, and HH bands separately. We also tested the method to some combinations of these sub-bands; e.g., such as $\psi^i \in \{LL,HH,HL,HH\}$, $\psi^j \in \{HL,LH\}$, $\psi^s \in \{HL,LH,HH\}$, or $\psi^d \in \{LL,HH\}$. Similar procedure was also applied to the test with 2-level decomposition. Only LL band was sub sampled using non-overlapped block average computation; the others were sub sampled using their standard deviation values.

Throughout the experiment we have found that in general the performance of the model using KS tests were lower than that using $\chi^2$-tests. In the tests involving 1-level decomposition, only tests on individual subband LL, $\psi^i$, and $\psi^j$, have shown good results; i.e., we needed to consider either one of the approximation coefficients of the image (LL), the coefficients in all subband (LL,HL,HH,HH), or the combination of the approximation and diagonal details of the image (LL,HH). Similarly, in 2-level decomposition good results have been shown either by the tests on the approximation coefficients (LL$_0$), the tests on all subbands (from LL$_0$, HH$_0$), or the combination of the approximation and diagonal details of 2-level decomposition (LL$_0$,HH$_0$, HH$_0$). As an example, Figure 5(a) illustrates the results of tests on 1-level decomposition for different subbands and bin sizes. It shows that, regardless of the bin size of the histograms, tests involving all subbands (64-(All), 128-(All), 256-(All)) performed better. However, using all subbands may not be desirable for some applications since it means that the reduced-reference data is quadrupled in size. In Figure 5(a) comparable performances are shown by the tests using $\{LL,HH\}$ subbands for $B \in [64,128]$. This may be a better alternative than the one that uses all subbands although it still requires double the size of the reduced-reference data. Note also the benefit of applying 1-level wavelet decomposition to the analysis; in Figure 5(a) where we have applied 1-level decomposition, smaller bin sizes can compete with the larger ones. This contradicts our result in the analysis without wavelet decomposition where better performance can only be achieved when larger bin size was used.

The experiments using 2-level decomposition, however, indicate slightly different results; see Figure 5(b). It shows that going further into higher level decomposition did not always result in performance gain. Notable decrease in the

**Figure 4:** Results on histograms/cumulative histograms of images without wavelet decompositions scheme

**13.3 Test on histograms of image with decomposition**

The previous results show that the proposed models without wavelet decomposition have managed to outperform the PSNR. The performance margin, however, was not really convincing since the highest $r_p$ that we had was still below 0.8 correlation. To see whether any further improvement can be achieved, we also analysed and compared the performance of the model with/without wavelet decomposition. Three different bin sizes were tested; $B \in \{64,128,256\}$. 

![Image](86x290 to 294x553)

![Image](344x627 to 350x634)
correlations were shown by the $\chi^2$-based analysis using $B = 64$ and $B = 256$; the analysis with $B = 128$ was more stable. The results of the KS test method also demonstrated stronger results than the ones in 1-level decomposition. Unfortunately, even the best results of KS test on 2-level decomposition could not outperform the $\chi^2$-test. Last but not least, it is worth noting that the experiments using 2-level decomposition have seen the superiority of the analysis with $B = 128$, in particular on the $LL_2$ subband. Even though the analysis on the $\{LL, HH, HH_1\}$ subbands gave stronger performance, it required three times as much of the reduced-reference data. Since the $LL_2$ subband is the approximation coefficients of the analysis, it suggested that more benefit might be gained by focusing the analysis on the coarser scale of the image. In Figure 5, the $\chi^2$-test method on the approximation coefficients of the image ($LL$ in 1-level and $LL_2$ in 2-level, both using $B = 128$) managed to achieve $r_p$ slightly below 0.8 correlations.

![Figure 5](image)

(a) Level 1 Decomposition

(b) Level 2 Decomposition

Figure 5: Results of KS and $\chi^2$-tests on histograms of wavelet coefficients (various sub-bands) of decomposition level 1 and 2.

![Figure 6](image)

Figure 6: Results of $\chi^2$-test on histograms of wavelet coefficients ($LL$ sub-band only) of various decomposition levels.

Interesting results were found in the $\chi^2$-tests on the $LL$ subband for different levels of decomposition; it is illustrated in Figure 6. This figure suggests that better performance of the analysis without wavelet decomposition can only be achieved at higher bin size (such as $B = 256$). On the other hand, smaller bin sizes were more suitable for analyses with wavelet decomposition, in particular at level 1 decomposition. At level 1 decomposition and $B = 128$, we have $r_p(Q(LL)) \chi^2, 128) = 0.79$ and $rs(Q(LL)) \chi^2, 128) = 0.79$ whereas $B = 64$ shows a slight decrease, $r_p(Q(LL)) \chi^2, 64) = 0.78$ and $rs(Q(LL)) \chi^2, 64) = 0.79$.

![Figure 7](image)

Figure 7: Results of $\chi^2$-tests on histograms of wavelet coefficients (various sub-bands) at decomposition level 1 with sub-sample using $4 \times 4$ block processing on the subband.

The influence of using non-overlapped block processing method to the subbands is illustrated in Figure 7. In this figure, $BLKSIZE$ is the size of the non-overlapped block ($BLKSIZE = 4$ means that block of $4 \times 4$ was used during the tests). The analysis with $BLKSIZE = 1$ is none other than the method without any block processing. Either $LL$ or a combination of $\{LL, HH\}$ subbands were given as samples in the figure; tests on $LL$ subbands computed the histograms of the sub-sampled $LL$ subbands using the average value of every $4 \times 4$
coefficients, whereas tests on HH subbands used its standard deviation. Figure 7 shows that performance of the models was better when BLKSIZE = 4; no gain in performance was acquired when smaller or bigger block sizes were used. Apart from BLKSIZE = 4, bigger block size perform better when B was also made large (such as B = 256). The best results, however, were achieved when B = 128 and BLKSIZE = 4; i.e., $r_p(Q(LL) χ^2, 128) = 0.88$ for the LL subband.

The performance differences between KS test and $χ^2$-test when non-overlapped block processing was involved are depicted in Figure 8. This figure is very similar to Figure 5(a), except that Figure 8 shows more improvement of the $χ^2$-test method due to block processing scheme over the one without block processing. Since Figure 7 suggested that more gain could be achieved in performance when BLKSIZE = 4, in Figure 8 we have used BLKSIZE = 4 as well. Similar to Figure 5(a), Figure 8 also suggests that better performance could be gained when moderate bin size was chosen, B = 128. In Figure 8 the analysis on LL band with B = 128 virtually gave no performance differences with the analysis either on all bands or on subbands (LL, HH). This figure gives us another hint that it might be sufficient just to analyse the LL subband to get good agreement with the subjective data.

Figure 8: Results of KS and $χ^2$-tests on histograms of wavelet coefficients (various sub-bands) of decomposition level 1 with sub-sample using 4 × 4 block processing on the subband.

Figure 9 illustrates the scatter plot of the proposed method ($Q(ψ) χ^2, ψ = LL, B = 128$, LL sub-sampled using 4 × 4 non-overlapped block average). It shows a non-linear behaviour of the proposed model. Following [21], this non-linearity can be fitted using non-linear function. In Figure 9, we use a 3rd order polynomial function such as given in Eq.(14):

$$\text{DMOS}_p = b_0 + b_1 \cdot Q(\psi)^{LL} χ^2 + b_2 \cdot (Q(\psi)^{LL} χ^2)^2 + b_3 \cdot (Q(\psi)^{LL} χ^2)^3$$

$$Q(\psi)^{LL} χ^2$$

to transform the immediate output of the model, $Q(LL) χ^2$, into the final prediction value, DMOS. The polynomial fit is shown in the figure. When we take into account the polynomial fit, the Pearson correlation value of the model increases to $r_p(\text{DMOS}_p, 128) = 0.94$ whereas the Spearman correlation, which is already high, does not change much.

To compare the predictive power of $Q(ψ) χ^2$ and $Q(ψ) KS$, Figure 10 is also included here. This figure is the scatter plot of the method using KS test ($Q(ψ) KS, ψ = LL, B = 128$, with LL sub-sampled using 4 × 4 non-overlapped block average). By comparing Figure 9 and Figure 10, we can see why the model with $χ^2$-test was better than KS test; the KS test produces more outliers than that of $χ^2$-test. In Figure 10, $r_p(Q(LL) KS, 128) = 0.8$ and $r_s(Q(LL) KS, 128) = 0.81$ – considerably weaker than those of $χ^2$-test.

Taken together all these results suggest that the histogram-based method benefited from computation on the sub-sampled version of LL subband in 1-level wavelet decomposition, with recommended BLKSIZE = 4 and bin size B = 128. The use of the LL subband from the wavelet-based multiresolution analysis is justified since much of the image energy is concentrated in this band. The introduction of sub-sampling operation to the LL band, by computing a local average of 4 × 4 in the subband, could further increase the performance of the proposed method. The method seems very similar to the 8 × 8 sub-sampled version of the image without wavelet decomposition (Section 4.2), however the results shown by these two methods prove otherwise. The sub-sampling computation to the approximation coefficient (in the wavelet domain) enhanced the differences between histograms of the degraded image and the original one; the method with wavelet decomposition benefited from this and, as a result, achieved better performance.

Figure 9: Scatter plot of $Q(LL) χ^2$ on sub-sampled LL band, 1-
distribution comparison based on a simple, well-known statistical test.

ACKNOWLEDGMENT

The authors would like to thank University of Essex and BT CTO, Ipswich, United Kingdom for the financial support for the research presented in this paper. Thanks also go to Prof. Yuukou Horita of Toyama University Japan for the image quality database used in this paper.

REFERENCES


