# Fourier Transform-Based Scalable Image Quality Measure

Manish Narwaria, Weisi Lin, *Senior Member, IEEE*, Ian Vince McLoughlin,
Sabu Emmanuel, and Liang-Tien Chia, *Member, IEEE*

*Abstract*—We present a new image quality assessment algorithm based on the phase and magnitude of the 2-D discrete Fourier transform. The basic idea is to compare the phase and magnitude of the reference and distorted images to compute the quality score. However, it is well known that the human visual system's sensitivity to different frequency components is not the same. We accommodate this fact via a simple yet effective strategy of non-uniform binning of the frequency components. This process also leads to reduced space representation of the image thereby enabling the reduced-reference (RR) prospects of the proposed scheme. We employ linear regression to integrate the effects of the changes in phase and magnitude. In this way, the required weights are determined via proper training and hence more convincing and effective. Last, using the fact that phase usually conveys more information than magnitude, we use only the phase for RR quality assessment. This provides the crucial advantage of further reduction in the required amount of reference image information. The proposed method is, therefore, further scalable for RR scenarios. We report extensive experimental results using a total of nine publicly available databases: seven image (with a total of 3832 distorted images with diverse distortions) and two video databases (totally 228 distorted videos). These show that the proposed method is overall better than several of the existing full-reference algorithms and two RR algorithms. Additionally, there is a graceful degradation in prediction performance as the amount of reference image information is reduced thereby confirming its scalability prospects. To enable comparisons and future study, a MATLAB implementation of the proposed algorithm is available at http://www.ntu.edu.sg/home/wslin/reduced_phase.rar.

*Index Terms*—Fourier phase and magnitude, image quality assessment (IQA), non-uniform frequency bins.

## I. INTRODUCTION

TODAY'S multimedia systems generate an enormous amount of visual multimedia content in the form of images and videos. The nature of transmission channels (e.g., lossy transmission network) and the constraints arising out of limited resources (for instance this prompts the need for compression) usually lead to loss of perceptual quality of images and videos. This in turn will lower the satisfaction and enjoyment level of the viewers for which these images and videos are meant. As a result, visual quality assessment

is an important part of today's multimedia and visual communication systems. Because, the opinion of human viewers is the ultimate benchmark of quality, subjective assessment remains the most reliable and accurate method of quality assessment, in the scenario where it is feasible to engage a sufficient number of subjects and on-line, in-loop and/or real-time operations are not required. The International Telecommunication Union Recommendation BT.500 [1] has formally defined subjective assessment as the most reliable way of visual quality assessment.

However, subjective evaluation is time-consuming, cumbersome, expensive, and tends to be non-repeatable, as a result, it cannot be easily and routinely performed for many scenarios, e.g., selecting the prediction mode in H.264 video coding algorithm. These drawbacks have prompted the development of objective quality methods, which rely on computational models for quality prediction. Even though the objective methods are usually less accurate, they alleviate a majority of the mentioned drawbacks associated with the subjective method. Consequently, objective methods are useful in many applications. For example, the well-known image quality assessment (IQA) algorithm structural similarity index measure (SSIM) [6] has been recently used as the optimization criterion for the H.264 video coding algorithm [7]. Other applications of visual quality assessment include signal acquisition, synthesis, enhancement, watermarking, compression, transmission, storage, retrieval, reconstruction, authentication, display, and printing.

Objective IQA algorithms can be classified into three categories based on the amount of information used for predicting quality: 1) full-reference (FR) metrics that use complete reference image information; 2) reduced-reference (RR) metrics that use only partial information from the reference image; and 3) no-reference (NR) metrics that do not use any reference image information. FR metrics are generally more accurate, while NR metrics although less accurate and usually distortion specific can be used when the reference image is not available. RR algorithms are essentially a tradeoff between these two because only partial information of the reference image is required. Mean squared error (MSE) or the peak signal-to-noise ratio (PSNR) is the simplest and most widely employed FR method. It is however criticized [8] for its inability to mimic human perception especially when the distortion is not additive in nature.[1]

A survey of published literature shows that over the past few years, significant research effort has been put into [2]–[5] developing objective image quality measures to mimic human

[1]For additive noise, PSNR can be a good indicator of the perceptual quality.

visual system's (HVS's) perception more closely. The first image and video quality metrics were developed by Mannos *et al.* [9] and Lukas *et al.* [10]. Later, the well-known HVS-based metrics are the visual differences predictor [11], the Sarnoff just noticeable difference (JND) metric [12], moving picture quality metric [13], and the perceptual distortion metric [14]. The well-known algorithm SSIM [6] was also proposed, and is based on the assumption that structural degradation can be equated to the loss of visual quality. Another IQA scheme known as the visual information fidelity (VIF) index [15] has also been developed. It equates perceptual quality to the amount of information regarding the reference image that can be extracted from the distorted image (images are modeled using Gaussian scale mixtures to measure the amount of image information). The visual signal-to-noise ratio (VSNR) proposed in [16] deals with both detectability of distortions (low-level property of vision) and structural degradation based on the global precedence (mid-level visual property). Several transform-based FR methods have also been proposed in the literature, including the ones based on singular value decomposition [17], [18], [66], frequency-domain transforms, such as DCT and wavelets [19], discrete orthogonal transforms [20], contourlet transform [21], etc. The basic idea of these methods is to compare the transformed image signal components (of reference and distorted images) because the transformation will usually help in better representation of the image signals. Another class of FR schemes employs image gradient ([55], [56] for instance) to quantify image quality, and is based on the idea that edges play an important role in perceiving image structure. The mentioned IQA schemes are some of the FR algorithms and the list is by no means exhaustive. At this point, we would like to refer the reader to [2]–[5] for more comprehensive reviews and details regarding existing IQA schemes.

We now briefly introduce the relevant works regarding the phase and magnitude. There have been several studies examining the role played by phase and magnitude in images. These have concluded that phase generally contains more image information and the so called phase dominance in images has been long established. The early study in [65] highlighted the importance of phase in image processing filters. The study in [33] concluded that while both phase and magnitude convey information regarding the signal, it is the phase information that provides more significant details. The authors in [25] have argued that edges can be detected more efficiently at points of maximum phase congruency. A recent work in [52] employs Fourier analysis for the task of ranking data, and it was concluded that the phase is much more important to matching the appearance of the data than the magnitude. In addition, psychophysical studies [26], [27] also provide evidence in favor of the importance of phase. These studies primarily examine the effect of phase and magnitude manipulations on the interpretability of images. The studies in [28] and [29] explored the relative importance of spectral amplitude and phase errors on reconstructed images in terms of the expected mean-square-error in the image. Reference [30] investigated the human visual sensitivity to phase perturbations (phase quantization and randomization), by examining the global image statistics (skewness and kurtosis).

Because the phase can capture perceptually important features (such as edges and contours) it has been used in many image processing applications, such as measuring image sharpness [31], image registration [32], [36], [61], palmprint recognition [37], visual saliency detection [38], and face recognition [39], [40], to list a few. There also exist a few works [41], [42] which have exploited phase for IQA. These methods have, however, achieved limited success as they tend to rely upon a direct-phase comparison between the reference and distorted images for quality computation, which has its drawbacks (this point is elaborated further in Section II-B).

Even though several studies have pointed out (some of them mentioned above) that phase plays a bigger role, the magnitude information cannot be completely ignored. This is obvious because both phase and magnitude are required for perfect image reconstruction. In this paper, we therefore propose a new IQA scheme by utilizing the phase and magnitude of the Fourier transform. The proposed method is different from existing works based on phase in the following ways.

1) We take into account the human sensitivity to different frequency components: in general, the HVS can tolerate more error in higher frequency components, while the distortion in lower frequency components has a larger impact on the visual quality. We achieve this via binning of the higher frequency components leading to reduced space.

2) We employ a regression-based method for combining the quality scores from phase and magnitude changes leading to more convincing integration of the two.

3) A thorough set of experimental results are presented, which provide evidence in favor of the proposed scheme. To that end, we have used a total of nine publicly available subjectively rated databases: seven image databases (with a total of 3832 distorted images having diverse distortions) and two video databases (totally 228 distorted videos).

4) We also explore and demonstrate the scalability of the proposed method by use of only phase information. This helps us in significantly reducing the amount of reference information needed.

The remainder of this paper is organized as follows. In Section II, we first discuss the role of phase and magnitude of the Fourier transform. We then present a new IQA measure with proper analysis and reasoning. Extensive experimental validation and related analysis are reported in Section III. Finally, Section IV gives the concluding remarks.

## II. PROPOSED METHOD FOR IMAGE QUALITY PREDICTION USING PHASE AND MAGNITUDE OF 2-D DISCRETE FOURIER TRANSFORM (DFT)

### A. Phase and Magnitude Characterization

As pointed out in the previous section, there seems to be a general consensus that phase plays a more prominent role in characterizing the signal properties. As a simple illustration of this, Oppenheim and Lim [22] demonstrated that exchanging the amplitude or phase spectrum of two images tends to produce a hybrid image more closely resembling the image
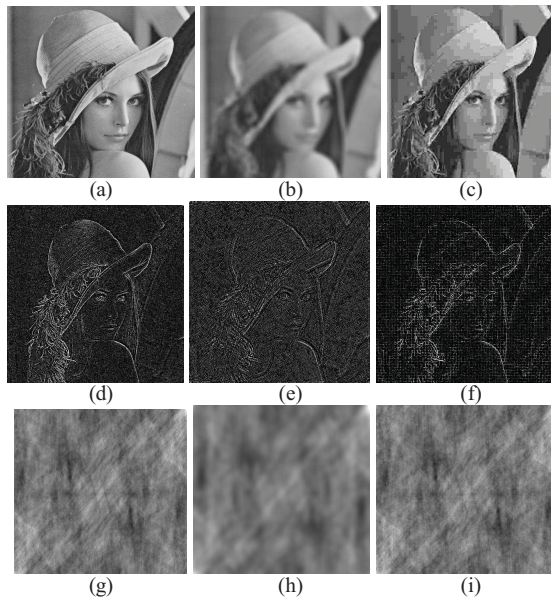
Fig. 1. (a) Original image. (b) Blurred image. (c) JPEG compressed image. (d) Image constructed using constant magnitude and phase of (a). (e) Image constructed using constant magnitude and phase of (b). (f) Image constructed using constant magnitude and phase of (c). (g) Image constructed using constant phase and magnitude of (a). (h) Constructed using constant phase and magnitude of (b). (i) Constructed using constant phase and magnitude of (c).

that contributed the phase spectrum. This means that phase conveys more crucial information than magnitude. The authors in [31] carried out experiments to investigate the visual impact of distortion in the phase and the magnitude of an image. First, the magnitude spectrum was distorted by adding a random shift $\alpha.S$, where $\alpha$ is a constant and $S$ is made of i.i.d. random variables uniformly distributed on $(-\pi, \pi)$. Then the image was reconstructed from the distorted magnitude and original phase. Next, the phase spectrum was distorted in a similar manner and the image was reconstructed using original magnitude and distorted phase. It was, found that the distortion in phase had a bigger impact on the visual appearance of image and hence its quality.

On the other hand, the effect of magnitude distortion was much less annoying and had a lower impact on the visual quality. Further statistical evidence has been presented in [23], where it is shown that random re-assignment of phase has a more severe effect on image quality as compared to random re-assignment of the magnitude. The image denoising method proposed in [24] also relies on preserving the perceptually important phase information in the signal. This is because the phase determines the locations of perceptually-significant features, and has a bigger contribution than magnitude in determining the image appearance.

It has been further pointed out [38], [62] that the amplitude spectrum specifies how much of each sinusoidal component is present, and the phase information specifies, where each of the sinusoidal components resides within the image. The authors in [38] used 1-D signals and demonstrated that when the waveform is a positive or negative pulse, its phase only reconstruction contains the largest spikes at the jump edge of the input pulse. This is because many varying sinusoidal

components locate there. On the other hand, when the input is a single sinusoidal component of constant frequency, there is no distinct spike in the reconstruction. Thus, phase of the signal carries information regarding edges and other salient parts. This is also true for a 2-D signal (image), and due to this phase has been used [38], [63] to obtain the image saliency map and also in edge detection [64]. We conducted experiments in which we reconstructed the image using constant magnitude and original phase (and vice-versa i.e., original magnitude and a constant phase). Fig. 1 shows three images and their reconstructed versions. The first row of Fig. 1 shows (a) original, (b) blurred, and (c) JPEG compressed images. The second row of Fig. 1 shows the images reconstructed from their respective phases but constant magnitude spectra, while the third row in Fig. 1 shows the images reconstructed using their respective magnitude but a constant phase. As can be seen, the images in Fig. 1(d)–(f) capture the most important features, such as edges and contours. One can also notice the damage that is caused to the image structure due to blurring and JPEG compression. Therefore, phase similarity (or difference) between reference and a distorted image is expected to give a reasonable estimate of structural degradation (provided that signal contents are properly discriminated as explained in the next section). On the other hand, the images in Fig. 1(g)–(i) convey less information although some changes can be noted due to the blurring and JPEG distortions.

### B. IQA by Non-Uniform Binning of the 2-D DFT Coefficients

Non-uniform binning of frequency coefficients was first explored in [60], which proposed a set of features for face recognition. In this paper, we provide analysis and justification for binning the frequency coefficients toward more accurate and efficient (in terms of RR prospects) visual quality assessment.

As mentioned, we propose the use of phase and magnitude to compute the image quality. A natural (and intuitive) way of determining the quality of distorted images (compared with the reference image) is to measure the similarity (or difference) between the phases and magnitude of the reference and distorted images. Conceptually, this would be similar to MSE (or PSNR), which directly computes the difference between the pixels of the reference and distorted images. However, like MSE, such an approach would be less effective because it neither accounts for any of the HVS' characteristics nor it distinguishes between the signal content. In particular, it fails to consider the unequal sensitivity of the HVS to distortions in different frequency components. It is known that natural images are characterized by a fair amount of redundancy. A common characteristic of most images is that the neighboring pixels are correlated. Exploiting this, there has been several image compression techniques that aim to reduce the number of bits needed to represent an image, by removing the spatial and spectral redundancies.

It can, therefore, be argued that for perceptual quality assessment, it would be more effective to focus on changes/distortion in perceptually important components. For

(a)       (b) DMOS = 0.031       (c) DMOS = 0.267       (d) DMOS = 0.509

(e)       (f) DMOS = 0.015       (g) DMOS = 0.056       (h) DMOS = 0.179

Fig. 2. Effect of distortion on image with relatively more smooth areas and more textured areas (these images are from CSIQ image database). First row: (a) original image "lady_liberty," (b) JPEG 2000 distorted image for the first distortion level, (c) JPEG 2000 distorted image for the second distortion level, and (d) JPEG 2000 distorted image for the third distortion level. Second row: (e) original image "foxy," (f) JPEG 2000 distorted image for the first distortion level, (g) JPEG 2000 distorted image for the second distortion level, and (h) JPEG 2000 distorted image for the third distortion level. The respective subjective scores in the form of difference mean opinion scores (DMOS) have been indicated below each distorted image.
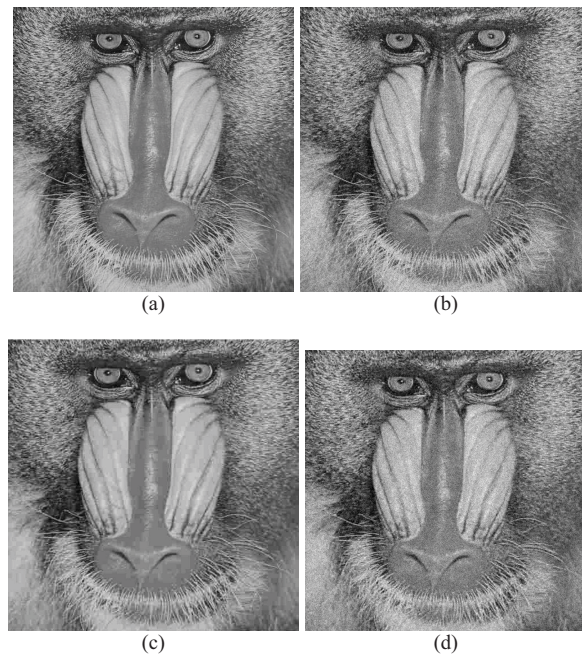


(a)         (b)

(c)         (d)

Fig. 3. Illustration of masking effect due to high texture. (a) Original "baboon" image. (b) Noisy image, SSIM = 0.8124, PSNR = 21.5090, $Q_{\text{Phase}}$ = 0.9495, and $Q_{\text{mag}}$ = 0.9776. (c) JPEG compressed image, SSIM = 0.7948, PSNR = 23.1508, $Q_{\text{Phase}}$ = 0.9511, and $Q_{\text{mag}}$ = 0.9801. (d) JPEG image, SSIM = 0.6785, PSNR = 18, 5705, $Q_{\text{Phase}}$ = 0.9292, and $Q_{\text{mag}}$ = 0.9629 in (c) distorted by the same amount of Gaussian noise that was used to obtain (b). Objective predictions from SSIM, PSNR, and proposed method are also indicated for distorted each image.

example, it is well known that textured regions (higher frequency) can usually tolerate more distortion (error) than smooth regions (lower frequency). This is because the distortion in textured regions is usually masked (texture masking). Masking effect refers to the reduction of the visibility of image distortion due to the presence of the original content in the reference image.

Stated differently, the JND in textured regions is higher than in the smoother areas of the image [43]. Psycho-visual exper-

iments have also shown that the HVS has reduced sensitivity for patterns with high spatial frequencies and therefore their distortion/perturbation is less annoying. This fact is used in JPEG compression, where the higher frequency signals are largely discarded via non-uniform quantization. As an illustration, we have shown four images in the first row of Fig. 2 [here image (a) is the original image, which is relatively smooth] and four images in the second row [here image (e) is the original image, which has more texture]. These images have been taken

Frequency components in each bin are
averaged out

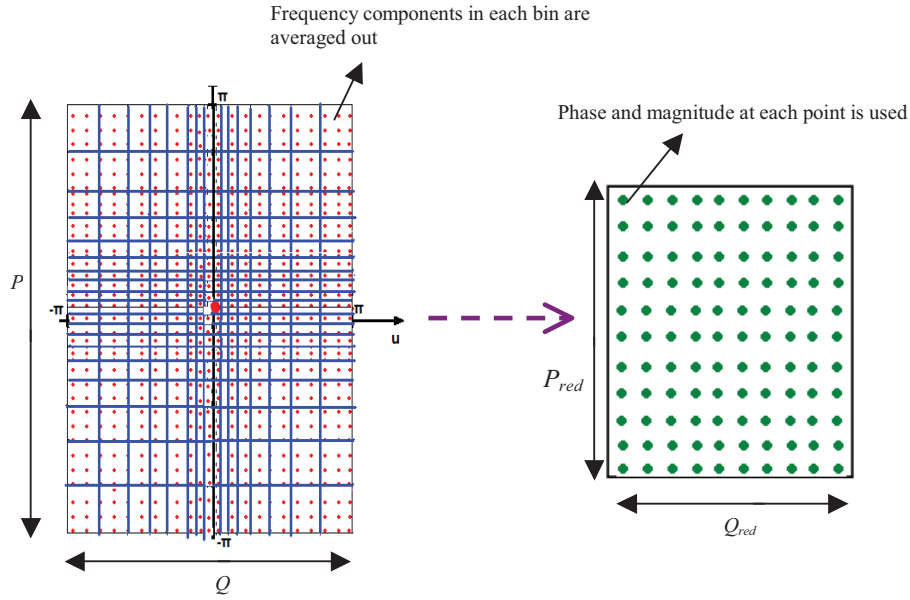Phase and magnitude at each point is used

Fig. 4.    Representative diagram of the non-uniform binning of the DFT coefficients. Notice that the bin sizes are biggest for higher frequency components.

from the CSIQ database [46]. The two original images were compressed using JPEG 2000 to obtain the distorted images shown in (b)–(d) and (f)–(h), respectively. The compression level along each row increases from left to right equally for both images. We have indicated the respective subjective scores (in the form of DMOS, lower indicates better subjective quality) for the distorted images. As can be seen, for a same amount of distortion, the DMOS scores are higher for the images in the first row as compared to those in the second row. This means that the textured image can tolerate more distortion than the relatively smooth image. Another example is shown in Fig. 3. As can be observed, the "Baboon" image is highly textured and the increased amount of distortion (or error) does not necessarily imply the same loss of perceived quality. This is because a large amount of distortion is masked due to texture and its visibility is reduced. Based on the foregoing discussion and analysis, it is evident that for assessing visual quality, the unequal sensitivity of the HVS to distortions should be taken into consideration. To tackle this, we propose a simple yet effective strategy as explained next.

We divide the spectrum into non-uniform bins such that the bin size is bigger for higher frequency and smaller at lower frequency. A representative diagram is shown in Fig. 4, where each red dot represents a DFT coefficient. In this figure, the DC component is at the center and higher frequency components lie away from the center as indicated. Next, we obtain the average of frequency coefficients in each bin.

With the said procedure, we obtain a reduced space representation of the spectrum in which the higher frequency components are represented by the average of the components in each bin. This can also be taken as a special case of downsampling, wherein the frequency components in each bin are represented by just one sample (the average). On the other hand, the lower frequency components (starting with the DC) are not binned. Effectively, this means that we analyze the lower frequency components at finer resolution, while higher

frequency components at coarser resolution. As an illustration of the effectiveness of this, we have indicated in Fig. 3, the objective quality scores from SSIM, PSNR, and the proposed $Q_{\text{Phase}}$ and $Q_{\text{mag}}$ [defined later in (7) and (8), respectively]. Note that SSIM, $Q_{\text{Phase}}$ and $Q_{\text{mag}}$ predict scores in the range [0, 1] with 1 denoting best quality and 0 indicating worse quality. The reader will notice that the images shown in Fig. 3(b)–(d) are of similar visual quality to the original image (a). Therefore, objective predictions should be close to 1 in case of SSIM, $Q_{\text{Phase}}$ and $Q_{\text{mag}}$ and a large number in case of PSNR. However, both PSNR and SSIM tend to overestimate the error. On the other hand, $Q_{\text{Phase}}$ and $Q_{\text{mag}}$ are better (both predict scores closer to 1) because of higher emphasis on the distortion in lower frequency components.

The 2-D DFT, $Y(u, v)$, of the image $y(n_1, n_2)$ (size $N$ by $N$) is defined as

$$Y(u, v) = \frac{1}{N} \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} y(n_1, n_2) e^{-j2\pi \left( \frac{un_1 + vn_2}{N} \right)} \qquad (1)$$

where $n_1$ and $n_2$ denote the spatial coordinates, and $u$ and $v$ are the frequency coordinates. $Y(u, v)$ is, in general, a complex number consisting of the real and imaginary parts. Using Euler's formula, we can express $Y(u, v)$ as

$$Y(u, v) = |Y(u, v)| e^{j\phi(u,v)} \qquad (2)$$

where $|Y(u, v)|$ represents the magnitude, and $\varphi(u, v)$ denotes the phase such that

$$|Y(u, v)| = \sqrt{(re(Y(u, v))^2 + (im(y(u, v))^2} \qquad (3)$$

and

$$\phi(u, v) = \arctan \left( \frac{im(Y(u, v))}{re(Y(u, v))} \right) \qquad (4)$$

with $re(.)$ and $im(.)$ denoting real and imaginary parts, respectively.

We now describe the procedure of determining the quality of a distorted image compared with the reference image.

We first divide the image into non-overlapping blocks of size $P \times Q$. Next, we obtain the 2-D DFT coefficients of each block. We then group the DFT coefficients via non-uniform binning as demonstrated in Fig. 4. As mentioned, lower frequency components are not binned, while the higher frequency components are represented by the averaged (mean) sample in each bin. In this paper, we assumed square blocks, i.e., $P = Q$. An important point is that we do not throw away higher frequency components (like JPEG compression, which discards higher frequency coefficients) but only group them in non-uniform bins. In this way, we obtain a better reduced representation of the image to assess its quality. Finally, the phase and magnitude are extracted (for both reference and distorted images) from the reduced space representation and their similarity scores computed. An additional advantage of such non-uniform binning is that quality can be predicted using lesser information from the reference image and thus useful for RR quality measurement.

Let $P_{ij}^{(\text{ref})}$, $M_{ij}^{(\text{ref})}$ denote the phase and magnitude of the $j^{\text{th}}$ block in the reference image while $P_{ij}^{(\text{dis})}$, $M_{ij}^{(\text{dis})}$ denote that for the distorted image. The phase and magnitude similarity scores for each image block are obtained as

$$q_{\text{Phase}}^{(j)} = \frac{1}{N_{\text{red}}} \sum_{i=1}^{N_{\text{red}}} \frac{2 P_{ij}^{(\text{ref})} P_{ij}^{(\text{dis})} + C}{\left(P_{ij}^{(\text{ref})}\right)^2 + \left(P_{ij}^{(\text{dis})}\right)^2 + C} \quad (5)$$

$$q_{\text{mag}}^{(j)} = \frac{1}{N_{\text{red}}} \sum_{i=1}^{N_{\text{red}}} \frac{2 M_{i}^{(\text{ref})} M_{ij}^{(\text{dis})} + C}{\left(M_{ij}^{(\text{ref})}\right)^2 + \left(M_{ij}^{(\text{dis})}\right)^2 + C} \quad (6)$$

where $C$ is a constant used to avoid division by zero and $N_{\text{red}} = P_{\text{red}} \times Q_{\text{red}}$. The phase (and magnitude) similarity score for the whole image is obtained by averaging the scores over all the image blocks (let $N_{\text{block}}$ be the number of blocks) to obtain the two overall scores as

$$Q_{\text{Phase}} = \frac{1}{N_{\text{block}}} \sum_{j=1}^{N_{\text{block}}} q_{\text{Phase}}^{(j)} \quad (7)$$

$$Q_{\text{mag}} = \frac{1}{N_{\text{block}}} \sum_{j=1}^{N_{\text{block}}} q_{\text{magnitude}}^{(j)} \quad (8)$$

Note that $0 \le Q_{\text{Phase}}, Q_{\text{mag}} \le 1$ with 0 indicating no similarity (worst quality) and 1 implying perfect similarity (highest quality).

As mentioned in the introduction, a few existing works use the phase information directly and one such method has been proposed in [42]. We denote it as direct phase similarity (DPS) measure in the remaining paper. In this method, image quality is computed by measuring the Pearson correlation coefficient between the phase of the reference and distorted images. Therefore, DPS scores are in the range [0, 1] with 0 denoting worse quality and 1 denoting perfect quality. To demonstrate the effectiveness of the proposed method [denoted as $Q_{\text{combined}}$ and defined in (9)] in comparison to DPS, we show four distorted images in Fig. 5 which have been taken
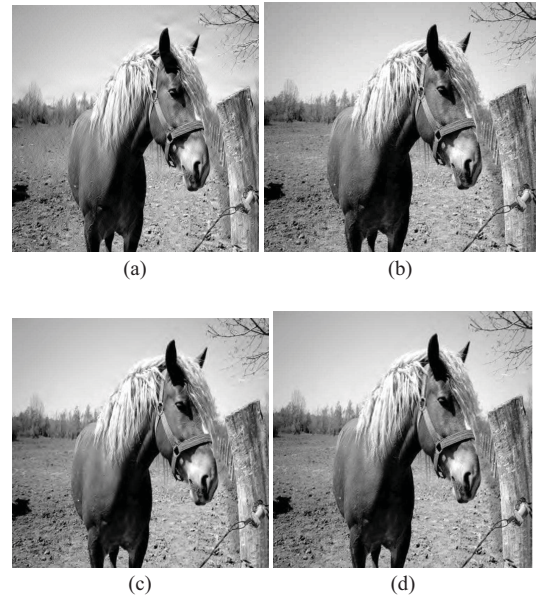


Fig. 5. Images taken from A57 database [49]. The subjective scores, DPS, PSNR, and proposed method's scores are given for each image. (a) DMOS = 0.5250, DPS = 0.7735, PSNR = 28.9830, $Q_{\text{combined}}$ = 41.5661. (b) DMOS = 0.1080, DPS = 0.7572, PSNR = 29.0030, $Q_{\text{combined}}$ = 32.4690. (c) DMOS = 0.1870, DPS = 0.6726, PSNR = 28.8700, $Q_{\text{combined}}$ = 35.5580. (d) DMOS = 0.1370, DPS = 0.7044, PSNR = 29.0560, $Q_{\text{combined}}$ = 35.0577.
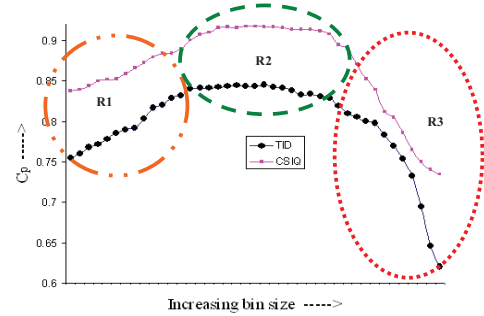


Fig. 6. Variation in the prediction performance with different bin sizes.

from A57 database [49]. The subjective scores in the form of DMOS (smaller means higher image quality) have also been indicated for each image. As can be seen, DPS scores are not consistent with subjective opinion, while the scores from proposed $Q_{\text{combined}}$ are closer to subjective viewing results. Note that $Q_{\text{combined}}$ scores are in the form of DMOS due to training with a database with DMOS (refer to Section II-C for details). We have also indicated the PSNR values for these images for comparison. Both DPS and PSNR do not explicitly account for the fact that HVSs' sensitivity to error in different frequency components is not the same. As opposed to this, the proposed method is more sensitive to error (distortion) in lower frequency.

At the same time, the error in higher frequency is not simply ignored but analyzed such that its overall impact is lower.

It may also be pointed out that the bin size in Fig. 4 increases in non-linear fashion. However, the exact bin size would depend on the chosen block size and also on the size of the reduced space representation that is required. To examine the variation in the prediction per-

formance with bin size, we show the variation of the Pearson correlation coefficients ($C_P$) for $Q_{\text{combined}}$ with increasing bin size in Fig. 6. Note that we have plotted the results for the two biggest image databases namely TID and CSIQ (their details are given in Section III-A). In this figure, the origin point indicates DPS, which means no binning i.e., using all phase points individually to determine the visual quality. Based on the shape of the curve shown in Fig. 6, we can identify three distinct regions.

The first one, marked as $R1$ in the plot, shows the positive impact that the non-uniform binning has and one can notice the increase in prediction accuracies as the bin size is increased. This happens because apart from the error in lower frequency (which is more important), the effect of changes in higher frequency is also incorporated (but at a lower resolution due to large bin size). The next region identified in Fig. 6 is marked as $R2$, where the prediction performance is nearly constant for both the databases. As bin size is increased further, there is considerable dip in the prediction accuracies (this is marked as $R3$ in Fig. 6). This can be explained because with very large bin size the even the lower frequency components start to get binned (i.e., averaged out), which means we tolerate more error in these components (which is obviously not in line with the fact that error in lower frequency usually has larger impact on quality).

We also found that frequency grouping in terms of percentage can be used for determining the approximate bin sizes. Starting from the highest frequency, we first bin the 25% highest frequency components. Then we bin the next 20% higher frequency components, then 18%, 10%, 6%, and 4%. The remaining 17% components (mainly lower frequency ones) are not binned. Of course such grouping will fix the size of the reduced space and for further information reduction, bigger percentage of the high frequency components should be binned. Also note that such grouping based on empirical results provides approximate guidelines and as future work, we would like to look into adaptive binning i.e., depending on the frequency content. Nevertheless, as can be seen from region $R2$, the performance is nearly constant (the variation in $C_P$ values is no more than 0.0045 and 0.0081 for TID and CSIQ databases, respectively) and thus our method is reasonably robust to small changes in the bin size (in region $R2$). We now describe the procedure to compute the quality score.

### C. Computing Quality Using Reduced-Space Representation

As pointed out earlier, with the reduction procedure shown in Fig. 4, we obtain a reduced-space representation of the image. That is, an image block of $P \times Q$ pixels will be represented by $(2 \times P_{\text{red}} \times Q_{\text{red}})$ coefficients, where $P_{\text{red}} < P$, $Q_{\text{red}} < Q$, and the factor of two is because we need both phase and magnitude at each point. We can further reduce the information required by using the fact that for real sequences, the 2-D DFT is symmetric. Because, the image $y(n_1, n_2)$ is real valued $Y(u, v)$ exhibits complex conjugate symmetry i.e., $Y(u, v) = Y^*(-u, -v)$. Therefore, the same magnitude information will be repeated because $|Y(u, v)| = |Y(-u, -v)|$. On the other hand, for phase we have $\phi(-u, -v) = -\phi(u, v)$,

i.e., phase differs only by the sign and so discarding the phase with negative sign does not have any impact. By this we do not imply that the symmetric phase does not provide any information, merely, for our purpose in this paper, it is not useful because we only need the similarity between the phases of reference and distorted images.

In this paper, we used a block size of 128, i.e., $P = Q = 128$ while we can choose different $P_{\text{red}}$ and $Q_{\text{red}}$. First, we selected $P_{\text{red}} = Q_{\text{red}} = 31$. This means that in this case we need only 11.73% of the information from the reference image to compute the quality. Due to exploiting the symmetry, we further reduce the information required to only about 6% of the information from the reference image to compute $Q_{\text{Phase}}$ and $Q_{\text{mag}}$, which have been defined in (7) and (8). Now these two need to be integrated into an overall quality score. To that end, we employed linear regression to obtain overall quality score. Although more sophisticated regression techniques may be employed, we chose linear regression as it is computationally simple and effective for our task. Assuming that $Q_{\text{combined}}$ denotes the overall quality score, we can express the solution as

$$Q_{\text{combined}} = w_1 Q_{\text{Phase}} + w_2 Q_{\text{mag}} + b \qquad (9)$$

where $w_1$ and $w_2$ are the respective weights for $Q_{\text{Phase}}$ and $Q_{\text{mag}}$ while $b$ is the intercept (a constant). Because phase plays a more crucial role in determining the change in the image *structure*, it is expected to have a larger impact on the overall quality and so $|w_1| > |w_2|$. However, to determine the exact contribution (weights) of each term, it will be more convincing to obtain them via training instead of *ad-hoc* selection. Let $\{X_1, X_2, \ldots, X_l\}$ and $\{y_1, y_2, \ldots, y_l\}$ denote the training set. Here, each $X_i = \left[ (Q_{\text{Phase}})_i, (Q_{\text{magnitude}})_i \right]$ represents the 2-D vector consisting of the phase and magnitude similarity scores, and each $y_i$ is the associated subjective score (i.e., target value) for the $i^{\text{th}}$ image. Given the training data $(X_1, y_1), \ldots, (X_l, y_l)$, we find the weight vector $W = (w_1, w_2)$ and $b$ by solving the following optimization problems:

$$\min_{W, b} \sum_{i=1}^{l} \left( y_i - \left( W^T X_i + b \right) \right)^2.$$

Therefore, $(W^T X_i + b)$ approximates the training data by minimizing the sum of squared errors. We used the A57 database [49] for training and as a result of which $w_1 = -10.57$, $w_2 = -5.59$ and $b = 16.14$. As expected, $Q_{\text{Phase}}$ has a larger impact (contribution) than $Q_{\text{mag}}$. It is also easy to see that with this set of $w_1$, $w_2$, and $b$, $Q_{\text{combined}}$ will be close to zero for the perfect quality image (because $Q_{\text{Phase}} = Q_{\text{mag}} = 1$) and increase as image quality decreases. Hence, $Q_{\text{combined}}$ will predict DMOS because the training database A57 provides DMOS as the subjective scores.

As mentioned, $Q_{\text{combined}}$ requires about 6% reference image information. We can further reduce the required reference information with small loss in prediction accuracy. To that end, we first obtain the averaged phase and magnitude over

all the image blocks as

$$P_i^{(\text{ref})} = \frac{1}{N_{\text{block}}} \sum_{j=1}^{N_{\text{block}}} P_{ij}^{(\text{ref})} \qquad (10)$$

$$P_i^{(\text{dis})} = \frac{1}{N_{\text{block}}} \sum_{j=1}^{N_{\text{block}}} P_{ij}^{(\text{dis})} \qquad (11)$$

$$M_i^{(\text{ref})} = \frac{1}{N_{\text{block}}} \sum_{j=1}^{N_{\text{block}}} M_{ij}^{(\text{ref})} \qquad (12)$$

$$M_i^{(\text{dis})} = \frac{1}{N_{\text{block}}} \sum_{j=1}^{N_{\text{block}}} M_{ij}^{(\text{dis})}. \qquad (13)$$

We then calculate phase and magnitude similarities $Q_{\text{Phase}}^{(1)}$ and $Q_{\text{mag}}^{(1)}$ as

$$Q_{\text{Phase}}^{(1)} = \frac{1}{N_{\text{red}}} \sum_{i=1}^{N_{\text{red}}} \frac{2P_i^{(\text{ref})} P_i^{(\text{dis})} + C}{\left(P_i^{(\text{ref})}\right)^2 + \left(P_i^{(\text{dis})}\right)^2 + C} \qquad (14)$$

$$Q_{\text{mag}}^{(1)} = \frac{1}{N_{\text{red}}} \sum_{i=1}^{N_{\text{red}}} \frac{2M_i^{(\text{ref})} M_{ij}^{(\text{dis})} + C}{\left(M_{ij}^{(\text{ref})}\right)^2 + \left(M_{ij}^{(\text{dis})}\right)^2 + C}. \qquad (15)$$

The overall quality $Q^{(1)}$ is then determined via linear regression-based combination of $Q_{\text{Phase}}^{(1)}$ and $Q_{\text{mag}}^{(1)}$. Note that both $Q^{(1)}$ and $Q_{\text{combined}}$ are developed using $P_{\text{red}} = 31$ and $Q_{\text{red}} = 31$. However, $Q^{(1)}$ is different from $Q_{\text{combined}}$ in that it uses the phase and magnitude similarity between averaged image blocks. On the other hand, $Q_{\text{combined}}$ uses the phase and magnitude similarity between the individual image blocks as shown in (5)–(8).

As another example of reduction in the required number of coefficients from the reference image, we develop another algorithm following the same procedure as for $Q_{\text{Phase}}^{(1)}$ and $Q_{\text{mag}}^{(1)}$ [i.e., using (10)–(15)]. The only difference is that in this case we use $P_{\text{red}} = 31$ and $Q_{\text{red}} = 25$ (instead of $P_{\text{red}} = 31$ and $Q_{\text{red}} = 31$) in order to reduce the information further. Let $Q_{\text{Phase}}^{(2)}$ and $Q_{\text{mag}}^{(2)}$, respectively, denote the phase and magnitude similarities for this case. Note that we need only 400 phase coefficients from the reference image for computing $Q_{\text{Phase}}^{(2)}$ (similarly we require 400 magnitude coefficients from the reference image to calculate $Q_{\text{mag}}^{(2)}$). To further demonstrate the effectiveness of the reduced-space representation and its potential for scalability, we select $P_{\text{red}} = Q_{\text{red}} = 15$. Following the similar procedure as outlined for $Q_{\text{Phase}}^{(1)}$, we arrive at $Q_{\text{Phase}}^{(3)}$. Note that we again use only the phase information (i.e., 120 phase coefficients). Finally, we develop $Q_{\text{Phase}}^{(4)}$ and $Q_{\text{Phase}}^{(5)}$ using only the phase. We first use $P_{\text{red}} = Q_{\text{red}} = 15$. Next we use the average of the coefficients in $2 \times 2$ window. Using this we obtain $Q_{\text{Phase}}^{(4)}$ and $Q_{\text{Phase}}^{(5)}$ which, respectively, require 48 and 36 coefficients. Obviously, the averaging in $2 \times 2$ window will result in loss of prediction accuracy but this is done only to make the required number of coefficients in $Q_{\text{Phase}}^{(4)}$ and $Q_{\text{Phase}}^{(5)}$ the same as those in DNT [57] (requires 48 coefficients) and RR SSIM (requires 36 coefficients). The reader will notice that the proposed methods $Q_{\text{combined}}$, $Q^{(1)}$, $Q_{\text{Phase}}^{(2)}$, $Q_{\text{Phase}}^{(3)}$, $Q_{\text{Phase}}^{(4)}$,

and $Q_{\text{Phase}}^{(5)}$ are all obtained via systematic averaging ($Q_{\text{Phase}}^{(4)}$ and $Q_{\text{Phase}}^{(5)}$ use additional averaging) outlined in Fig. 4. The only difference being the use of different values of $P_{\text{red}}$ and $Q_{\text{red}}$.

## III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we present the experimental results to assess the prediction performance of the proposed scheme. Note that all the results, in this paper, are for the luminance component of the image. We also include the results for PSNR, MSSIM[2] [59], VSNR [16], VIF [15], PSNR-HVS-M [44], and DPS [42] all of which are FR schemes.

### A. Databases and Performance Assessment Criteria

In this paper, we used total of nine subjectively rated image and video databases, namely the LIVE image database [45], CSIQ database [46], IVC database [47], Toyama database [48], A57 database [49], TID database [50], WIQ database [51], LIVE video database [45], and the EPFL video database [53]. We provide a brief summary of these databases in Table I and refer the reader to the respective references for more details.

As mentioned, we used A57 database for training with the remaining image and video databases as test sets. Note that none of the images in A57 database appear in the remaining databases. For reporting the results for A57 database, we used WIQ database as the training set. Thus, training and test contents do not overlap and there is no parameter *optimization* toward any of the test databases. A five-parameter logistic mapping between the objective outputs and the subjective scores was also employed, following the video quality experts group (VQEG) Phase-I/II test and validation method [54]. The experimental results are reported in terms of four criteria, namely Pearson linear correlation coefficient $C_P$ (for prediction accuracy), Spearman correlation coefficient $C_S$ (for monotonicity), Kendall rank correlation coefficient $C_K$, and the root mean squared error (RMSE), between the subjective score and the objective prediction (after logistic transformation). A better quality metric has higher $C_P$, $C_K$, $C_S$, and lower RMSE.

### B. Performance Comparison

Table V presents the comparative results for the seven image and two video databases. One can observe that the proposed method performs better (in many cases) or is very competitive when compared to the FR schemes. This is significant given that $Q_{\text{combined}}$ requires only about 6% (of the total number of pixels) of the reference information in contrast to the FR schemes (which need the complete reference information). Another observation from Table V is that some existing metrics are less consistent in that they do not perform well for all the databases. For instance, VSNR does well on A57 but its performance is relatively low for other databases, VIF performs well on three databases but performs rather poorly on A57. By contrast, the proposed scheme is more consistent in its performance.

TABLE I

MAJOR CHARACTERISTICS OF THE NINE SUBJECTIVELY RATED
DATABASES USED IN THIS PAPER

| | No. of reference images/videos | No. of distorted images/videos | No. of distortion types | Typical image/ frame size | Subjective score format (Range) |
|---|---|---|---|---|---|
| LIVE | 29 | 779 | 5 | 768 × 512 | DMOS (0-100) |
| CSIQ | 30 | 866 | 6 | 512 × 512 | DMOS (0-1) |
| IVC | 10 | 185 | 4 | 512 × 512 | MOS (1-5) |
| Toyama | 14 | 168 | 2 | 512 × 768 | MOS (1-5) |
| A57 | 3 | 54 | 6 | 512 × 512 | DMOS (0-1) |
| TID | 25 | 1700 | 17 | 512 × 384 | MOS (0-9) |
| WIQ | 7 | 80 | 1 | 512 × 512 | DMOS (0-100) |
| LIVE video | 10 | 150 | 4 | 768 × 432 | DMOS (0-100) |
| EPFL video | 6 | 78 | 1 | 352 × 288 | MOS (1-5) |

TABLE II

$C_P$ VALUES FOR THE FOUR DISTORTION LEVELS IN THE TID DATABASE.
LEVEL 1 IMPLIES LOWEST DISTORTION LEVEL WHILE LEVEL 4 MEANS
THE HIGHEST DISTORTION LEVEL

| | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|---|
| MS-SSIM | 0.8259 | 0.7612 | 0.7003 | 0.7656 |
| VIF | 0.5645 | 0.5221 | 0.8181 | 0.8867 |
| VSNR | 0.6264 | 0.6431 | 0.4818 | 0.6486 |
| PSNR | 0.5742 | 0.3241 | 0.3601 | 0.3601 |
| PSNR-HVS-M | 0.4732 | 0.5336 | 0.4957 | 0.5514 |
| DPS | 0.5389 | 0.4520 | 0.7344 | 0.7900 |
| $Q_{combined}$ | 0.7898 | 0.6983 | 0.7324 | 0.8358 |

TABLE III

COMPARISON OF $C_P$ VALUES ACHIEVED BY PHASE AND MAGNITUDE

| Database/ Algorithm | LIVE | TID | Toyama | IVC | CSIQ |
|---|---|---|---|---|---|
| $Q_{mag}^{(2)}$ | 0.8810 | 0.6893 | 0.8831 | 0.8469 | 0.7598 |
| $Q_{Phase}^{(2)}$ | 0.9423 | 0.8053 | 0.9084 | 0.9046 | 0.8815 |

TABLE IV

PERFORMANCE COMPARISON OF THE PROPOSED SCHEME WITH
RR SSIM [58]. THE RESULTS HAVE BEEN COMPUTED AFTER
USING THE LOGISTIC MAPPING BETWEEN THE OBJECTIVE
AND SUBJECTIVE SCORES

| Criteria | Algorithm | LIVE | TID | Toyama | IVC | A57 | CSIQ |
|---|---|---|---|---|---|---|---|
| $C_P$ | RR SSIM [58] | 0.9194 | 0.7231 | 0.8051 | 0.8177 | 0.7044 | 0.8426 |
| | $Q_{Phase}^{(5)}$ | 0.8968 | 0.7682 | 0.8134 | 0.7400 | 0.8036 | 0.8576 |
| $C_S$ | RR SSIM [58] | 0.9129 | 0.7210 | 0.8003 | 0.8154 | 0.7301 | 0.8527 |
| | $Q_{Phase}^{(5)}$ | 0.9073 | 0.7547 | 0.8067 | 0.7356 | 0.7973 | 0.7917 |
| $C_K$ | RR SSIM [58] | 0.7349 | 0.5236 | 0.6090 | 0.6164 | 0.5345 | 0.6540 |
| | $Q^{(5)}$ | 0.7334 | 0.5611 | 0.6108 | 0.5355 | 0.6198 | 0.6211 |
| RMSE | RR SSIM [58] | 11.3026 | 0.9270 | 0.7423 | 0.7014 | 0.1744 | 0.1413 |
| | $Q_{Phase}^{(5)}$ | 12.0863 | 0.8592 | 0.7279 | 0.8195 | 0.1456 | 0.1345 |

To gain more insights into such behavior of quality metrics, we performed additional analysis using the TID database. It is our hypothesis that the variation in performance of quality metrics over different databases is partly due to the distortion levels. To confirm this, we observed the performance of different metrics for the four distortion levels of the TID database. The first level (Level 1) denotes just noticeable or near threshold distortion, while the fourth level (Level 4) indicates higher distortions. Table II presents the $C_P$ values for the prediction performance of different metrics on the four distortion levels. We can see that VIF, VSNR, and PSNR-HVS-M perform better for the fourth distortion level (i.e., higher amount of distortion), while they are relatively poor for lower distortion levels. Also we find that there is a large variation in prediction accuracies for DPS, VIF, and PSNR-HVS-M as we go from Levels 1–4. On the other hand, MSSIM, VSNR, and $Q_{combined}$ are more consistent for the four levels with $Q_{combined}$ being overall better.

In addition to the overall performance, the proposed method in general performed well for individual distortion types. As an example, we have presented the $C_P$ values in Table VII for some typical distortion types, including JPEG, JPEG 2000, additive white noise, and fast fading.

We have already mentioned that phase information is generally more crucial than magnitude. To verify this further, as an example we present the individual results for $Q_{Phase}^{(1)}$, $Q_{mag}^{(1)}$, and $Q^{(1)}$ separately in Table VI (RMSE is omitted as

---

[2]We include results only for MSSIM because it is usually better than the single scale SSIM [6].

it leads to similar conclusions as from other criteria). One can see that $Q_{Phase}^{(1)}$ gives higher correlation with the subjective scores across all the databases. Nevertheless, we note that $Q_{mag}^{(1)}$ also plays a role. It is therefore not surprising that $Q^{(1)}$ (which is a linear combination of $Q_{Phase}^{(1)}$ and $Q_{mag}^{(1)}$) achieves the best results for each database. Even though the improvement in some cases (over $Q_{Phase}^{(1)}$) is smaller, the consistency in improvement for all the databases indicates that both play a role in overall quality score determination. Further experimental evidence of this is given in Table III, which provides the correlation values (only $C_P$ values are shown) achieved by $Q_{Phase}^{(2)}$ and $Q_{mag}^{(2)}$ on the five biggest image databases. On expected lines, $Q_{Phase}^{(2)}$ performs better than $Q_{mag}^{(2)}$. Thus, $Q_{Phase}^{(2)}$ is effective for reducing the reference information on one hand and achieving reasonably high prediction accuracy on the other. The prediction accuracy of $Q_{Phase}^{(2)}$ is also reported in Table V. It is expectedly slightly worse than $Q$ and $Q^{(1)}$ but still achieves reasonably good overall performance given the fact that it needs only 400 coefficients from the reference image. Furthermore, the results for $Q_{Phase}^{(3)}$ are also given in Table V. While it gives lower correlations as compared to other schemes, the performance drop is within a reasonable range. Of course the most crucial advantage of $Q_{Phase}^{(3)}$ is with regards to its requirement of the reference information. The performance of $Q^{(1)}$, $Q_{Phase}^{(2)}$, and $Q_{Phase}^{(3)}$ on individual distortion types is also presented in Table VII. One can observe the scalability of the proposed method i.e., the degradation in prediction performance is graceful with reduction in reference information.

TABLE V

PERFORMANCE COMPARISON OF THE PROPOSED SCHEME WITH FR METRICS. THE RESULTS HAVE BEEN COMPUTED
AFTER USING THE LOGISTIC MAPPING BETWEEN THE OBJECTIVE AND SUBJECTIVE SCORES

| Database | Criteria | PSNR | PSNR-HVS-M | MS-SSIM | VSNR | VIF | DPS | $Q_{\text{combined}}$ | $Q^{(1)}$ | $Q_{\text{Phase}}^{(2)}$ | $Q_{\text{Phase}}^{(3)}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LIVE | $C_S$ | 0.8756 | 0.9295 | 0.9513 | 0.9280 | 0.9632 | 0.9292 | 0.9563 | 0.9479 | 0.9454 | 0.9287 |
| | $C_K$ | 0.6865 | 0.7659 | 0.8044 | 0.7625 | 0.8270 | 0.7571 | 0.8190 | 0.7992 | 0.7932 | 0.7664 |
| | $C_P$ | 0.8723 | 0.9251 | 0.9409 | 0.9237 | 0.9598 | 0.9246 | 0.9537 | 0.9450 | 0.9422 | 0.9228 |
| | RMSE | 13.3597 | 10.3722 | 9.2593 | 10.4694 | 7.6670 | 10.4058 | 8.2193 | 8.9325 | 9.1485 | 10.5488 |
| TID | $C_S$ | 0.5794 | 0.6128 | 0.8542 | 0.7049 | 0.7496 | 0.7059 | 0.8338 | 0.8210 | 0.7804 | 0.7847 |
| | $C_K$ | 0.4210 | 0.4764 | 0.6568 | 0.5345 | 0.5863 | 0.5189 | 0.6425 | 0.6259 | 0.5869 | 0.5907 |
| | $C_P$ | 0.5726 | 0.6051 | 0.8451 | 0.6823 | 0.8090 | 0.7549 | 0.8441 | 0.8302 | 0.8052 | 0.8023 |
| | RMSE | 1.1003 | 1.0685 | 0.7173 | 0.9810 | 0.7888 | 0.8801 | 0.7195 | 0.7482 | 0.7957 | 0.8010 |
| Toyama | $C_S$ | 0.6132 | 0.8480 | 0.8874 | 0.8608 | 0.9077 | 0.9203 | 0.9148 | 0.9001 | 0.9029 | 0.8590 |
| | $C_K$ | 0.4443 | 0.6568 | 0.7029 | 0.6745 | 0.7315 | 0.7541 | 0.7384 | 0.7171 | 0.7224 | 0.6696 |
| | $C_P$ | 0.6353 | 0.8580 | 0.8922 | 0.8704 | 0.9138 | 0.9264 | 0.9184 | 0.9061 | 0.9091 | 0.8604 |
| | RMSE | 0.9664 | 0.6428 | 0.5652 | 0.6160 | 0.5084 | 0.4711 | 0.4951 | 0.5295 | 0.5213 | 0.6378 |
| A57[3] | $C_S$ | 0.6189 | 0.8962 | 0.8414 | 0.9355 | 0.6223 | 0.4443 | 0.8937 | 0.8802 | 0.9181 | 0.8697 |
| | $C_K$ | 0.4309 | 0.7261 | 0.6478 | 0.8031 | 0.4589 | 0.3148 | 0.7191 | 0.7051 | 0.7443 | 0.6939 |
| | $C_P$ | 0.6347 | 0.8749 | 0.8575 | 0.9497 | 0.6157 | 0.4745 | 0.9147 | 0.9093 | 0.9294 | 0.9053 |
| | RMSE | 0.1899 | 0.1190 | 0.1264 | 0.0769 | 0.1937 | 0.2163 | 0.0993 | 0.1023 | 0.0907 | 0.1044 |
| IVC | $C_S$ | 0.6884 | 0.8832 | 0.8980 | 0.7993 | 0.8964 | 0.8819 | 0.8943 | 0.8905 | 0.8960 | 0.7881 |
| | $C_K$ | 0.5218 | 0.6935 | 0.7203 | 0.6053 | 0.7158 | 0.6853 | 0.7114 | 0.7033 | 0.7150 | 0.5823 |
| | $C_P$ | 0.7196 | 0.8905 | 0.9108 | 0.8034 | 0.9028 | 0.8941 | 0.9046 | 0.9003 | 0.9048 | 0.7935 |
| | RMSE | 0.8460 | 0.5544 | 0.5029 | 0.7255 | 0.5239 | 0.5456 | 0.5192 | 0.5304 | 0.5190 | 0.7414 |
| CSIQ | $C_S$ | 0.8005 | 0.8179 | 0.9133 | 0.8104 | 0.9195 | 0.7831 | 0.9237 | 0.9344 | 0.8082 | 0.8197 |
| | $C_K$ | 0.5984 | 0.6430 | 0.7393 | 0.6237 | 0.7537 | 0.5951 | 0.7619 | 0.7773 | 0.6472 | 0.6247 |
| | $C_P$ | 0.7998 | 0.8137 | 0.8990 | 0.7993 | 0.9277 | 0.8376 | 0.9171 | 0.9336 | 0.8815 | 0.8687 |
| | RMSE | 0.1576 | 0.1526 | 0.1150 | 0.1578 | 0.0980 | 0.1434 | 0.1047 | 0.0940 | 0.1240 | 0.1295 |
| WIQ | $C_S$ | 0.6257 | 0.7261 | 0.7360 | 0.6558 | 0.6918 | 0.6631 | 0.8418 | 0.8360 | 0.8271 | 0.7518 |
| | $C_K$ | 0.4626 | 0.5569 | 0.5645 | 0.4873 | 0.5246 | 0.5069 | 0.6519 | 0.6500 | 0.6386 | 0.5575 |
| | $C_P$ | 0.7549 | 0.7632 | 0.7761 | 0.7625 | 0.7333 | 0.7352 | 0.8547 | 0.8511 | 0.8481 | 0.7766 |
| | RMSE | 15.0235 | 14.8022 | 14.4442 | 14.8199 | 15.5734 | 15.5267 | 11.8914 | 12.0274 | 12.1378 | 14.4301 |
| LIVE video | $C_S$ | 0.5431 | 0.6889 | 0.7389 | 0.6710 | 0.5662 | 0.3654 | 0.7481 | 0.7487 | 0.7393 | 0.7397 |
| | $C_K$ | 0.3818 | 0.5179 | 0.5579 | 0.4977 | 0.3948 | 0.2561 | 0.5561 | 0.5581 | 0.5484 | 0.5492 |
| | $C_P$ | 0.5583 | 0.6947 | 0.7447 | 0.6878 | 0.5875 | 0.4379 | 0.7619 | 0.7623 | 0.7611 | 0.7590 |
| | RMSE | 9.1072 | 7.9262 | 7.3262 | 7.9687 | 8.8833 | 9.8689 | 7.1104 | 7.0844 | 7.1568 | 7.2270 |
| EPFL video | $C_S$ | 0.6869 | 0.8760 | 0.9220 | 0.8631 | 0.6866 | 0.7206 | 0.9301 | 0.9268 | 0.9187 | 0.9117 |
| | $C_K$ | 0.5058 | 0.6754 | 0.7642 | 0.6757 | 0.5178 | 0.5385 | 0.7749 | 0.7669 | 0.7590 | 0.7349 |
| | $C_P$ | 0.6907 | 0.8865 | 0.9499 | 0.8890 | 0.7681 | 0.7224 | 0.9438 | 0.9422 | 0.9356 | 0.9289 |
| | RMSE | 0.9753 | 0.6240 | 0.4216 | 0.6176 | 0.8636 | 0.9326 | 0.4458 | 0.4520 | 0.4711 | 0.4995 |

Finally, we compare the performance of the proposed scheme with DNT [57] and RR SSIM [58]. The prediction performances of $Q_{\text{Phase}}^{(4)}$ and DNT for LIVE image database are presented in Table VIII ($C_P$ and $C_S$ values are presented). It may be pointed out that DNT requires training and its authors have reported the experimental results for two training cases: 1) training with LIVE database and 2) training with A57 database. For fair comparison with the proposed scheme, we have included the results derived from [57] with A57 database as the training set and LIVE image database as the test set. We find that $Q_{\text{Phase}}^{(4)}$ performs well and is overall better.

Finally, the results for $Q_{\text{Phase}}^{(5)}$ and RR SSIM are presented in Table IV. For RR SSIM, we have reported the results as provided by its authors for six image databases. We can see that $Q_{\text{Phase}}^{(5)}$ performs better than RR SSIM for A57, CSIQ, and TID databases and achieves competitive performance on LIVE and Toyama databases.

In summary, we have presented six algorithms namely $Q_{\text{combined}}$, $Q^{(1)}$, $Q_{\text{Phase}}^{(2)}$, $Q_{\text{Phase}}^{(3)}$, $Q_{\text{Phase}}^{(4)}$, and $Q_{\text{Phase}}^{(5)}$ which, respectively, require approximately 1/17, 1/200, 1/490, 1/1640, 1/4096, and 1/5460 of the reference information (for image resolution of $512 \times 384$). These algorithms perform well and

TABLE VI

RESULTS FOR THE PHASE AND MAGNITUDE SCORES SEPARATELY. THE RESULTS FOR THEIR COMBINATION
HAVE ALSO BEEN PRESENTED FOR A QUICK GLANCE

| Criteria/ Database | $C_P$ | | | $C_S$ | | | $C_K$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $Q_{\text{Phase}}^{(1)}$ | $Q_{\text{magnitude}}^{(1)}$ | $Q^{(1)}$ | $Q_{\text{Phase}}^{(1)}$ | $Q_{\text{magnitude}}^{(1)}$ | $Q^{(1)}$ | $Q_{\text{Phase}}^{(1)}$ | $Q_{\text{magnitude}}^{(1)}$ | $Q^{(1)}$ |
| LIVE | 0.9413 | 0.8803 | 0.9450 | 0.9434 | 0.8798 | 0.9479 | 0.7980 | 0.6998 | 0.7992 |
| TID | 0.8135 | 0.6853 | 0.8302 | 0.7883 | 0.7050 | 0.8210 | 0.5960 | 0.5384 | 0.6259 |
| Toyama | 0.8992 | 0.8789 | 0.9061 | 0.8874 | 0.8748 | 0.9001 | 0.7011 | 0.6893 | 0.7171 |
| A57 | 0.8996 | 0.7470 | 0.9093 | 0.8840 | 0.7393 | 0.8802 | 0.7002 | 0.5694 | 0.7051 |
| IVC | 0.8927 | 0.8389 | 0.9003 | 0.8820 | 0.8315 | 0.8905 | 0.6955 | 0.6342 | 0.7033 |
| CSIQ | 0.8811 | 0.7549 | 0.9336 | 0.8067 | 0.7686 | 0.9344 | 0.6480 | 0.5894 | 0.7773 |
| WIQ | 0.8470 | 0.8342 | 0.8511 | 0.8284 | 0.8179 | 0.8360 | 0.6418 | 0.6297 | 0.6500 |

TABLE VII

PERFORMANCE COMPARISON FOR TYPICAL DISTORTION TYPES. THE RESULTS HAVE BEEN COMPUTED AFTER
USING THE LOGISTIC MAPPING BETWEEN THE OBJECTIVE AND SUBJECTIVE SCORES

| Distortion Type | Database | PSNR | PSNR-HVS-M | MS-SSIM | VSNR | VIF | DPS | $Q_{\text{combined}}$ | $Q^{(1)}$ | $Q_{\text{Phase}}^{(2)}$ | $Q_{\text{Phase}}^{(3)}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| JPEG | LIVE | 0.8897 | 0.9485 | 0.9812 | 0.9735 | 0.9859 | 0.9742 | 0.9773 | 0.9704 | 0.9690 | 0.9559 |
| | TID | 0.8703 | 0.9720 | 0.9607 | 0.9379 | 0.9547 | 0.9308 | 0.9469 | 0.9282 | 0.9231 | 0.9308 |
| | CSIQ | 0.8788 | 0.9576 | 0.9815 | 0.9487 | 0.9882 | 0.9695 | 0.9771 | 0.9704 | 0.9686 | 0.9644 |
| JPEG 2000 | LIVE | 0.8997 | 0.9200 | 0.9706 | 0.9641 | 0.9760 | 0.9583 | 0.9637 | 0.9546 | 0.9478 | 0.9321 |
| | TID | 0.8672 | 0.9669 | 0.9753 | 0.9531 | 0.9730 | 0.9629 | 0.9700 | 0.9604 | 0.9600 | 0.9510 |
| | CSIQ | 0.9463 | 0.9680 | 0.9785 | 0.9561 | 0.9776 | 0.9618 | 0.9722 | 0.9654 | 0.9627 | 0.9502 |
| Blur | LIVE | 0.7835 | 0.8869 | 0.9591 | 0.9369 | 0.9740 | 0.9412 | 0.9737 | 0.9627 | 0.9560 | 0.9310 |
| | TID | 0.8736 | 0.9143 | 0.9512 | 0.9277 | 0.9401 | 0.8857 | 0.9413 | 0.9229 | 0.9149 | 0.8968 |
| | CSIQ | 0.9081 | 0.9553 | 0.9669 | 0.9342 | 0.9717 | 0.9427 | 0.9728 | 0.9678 | 0.9606 | 0.9510 |
| Additive white noise | LIVE | 0.9857 | 0.9865 | 0.9725 | 0.9816 | 0.9841 | 0.9757 | 0.9847 | 0.9710 | 0.9692 | 0.9563 |
| | TID | 0.9341 | 0.9363 | 0.8021 | 0.7577 | 0.8725 | 0.6750 | 0.7144 | 0.6510 | 0.6827 | 0.5944 |
| | CSIQ | 0.8978 | 0.9433 | 0.9465 | 0.9260 | 0.9606 | 0.8703 | 0.9212 | 0.9004 | 0.9101 | 0.8720 |
| Fastfading | LIVE | 0.8897 | 0.9093 | 0.9284 | 0.9055 | 0.9613 | 0.9488 | 0.9431 | 0.9461 | 0.9380 | 0.9087 |
| | TID | 0.8536 | 0.9257 | 0.8386 | 0.7797 | 0.8372 | 0.7359 | 0.8369 | 0.7769 | 0.7796 | 0.7690 |

TABLE VIII

PERFORMANCE COMPARISON OF THE PROPOSED SCHEME WITH [57] FOR LIVE IMAGE DATABASE. THE RESULTS
HAVE BEEN COMPUTED AFTER USING THE LOGISTIC MAPPING BETWEEN THE OBJECTIVE AND SUBJECTIVE SCORES

| Criteria | Algorithm | All data | JP2(1) | JP2(2) | JPG(1) | JPG(2) | Noise | Blur | Fastfading |
|---|---|---|---|---|---|---|---|---|---|
| $C_P$ | DNT [57] | 0.8930 | 0.9115 | 0.9422 | 0.8501 | 0.9354 | 0.9401 | 0.8773 | 0.9243 |
| | $Q_{\text{Phase}}^{(4)}$ | 0.9009 | 0.9031 | 0.9362 | 0.8850 | 0.9623 | 0.9512 | 0.8931 | 0.8897 |
| $C_S$ | DNT [57] | 0.9093 | 0.9081 | 0.9239 | 0.8389 | 0.8734 | 0.9316 | 0.8608 | 0.9237 |
| | $Q_{\text{Phase}}^{(4)}$ | 0.9031 | 0.9140 | 0.9225 | 0.8915 | 0.8831 | 0.9435 | 0.8858 | 0.8888 |

are usually better or very competitive with FR schemes (and the two RR schemes). Importantly, the degradation in prediction performance is graceful with the reduction in reference information across all the databases. This enables scalability of the proposed method, which is a crucial advantage.

### C. Further Discussion

In the previous sections, we have shown the effectiveness of the proposed method with regards to its prediction accuracy and scalability. As stated before, these are achieved primarily by accounting for the unequal sensitivity of the HVS to changes/distortions in different frequency components. To further examine the impact of unequal emphasis on the different frequency components as exploited in the proposed method, we have presented the results for DPS in Table V and we make the following observations.

1) As already mentioned, we can regard DPS as similar to PSNR (or MSE) because it uses each phase point for computing image quality, while PSNR uses each pixel. However, DPS performs better than PSNR for most databases. This suggests that phase conveys more precise information regarding *structural* changes than pixel.

2) The proposed $Q_{\text{combined}}$, $Q^{(1)}$, $Q_{\text{Phase}}^{(2)}$, $Q_{\text{Phase}}^{(3)}$ are overall better and more consistent than DPS across databases. As mentioned before DPS is just the direct phase similarity measure, so the results clearly demonstrate

the positive impact of using the reduced-space representation, which leads to objective predictions that are better aligned with HVSs perception. It also confirms that discrimination of signal contents is an important aspect toward more accurate quality prediction.

3) A closer look at Table V reveals that DPS actually performs quite well for LIVE, Toyama, and IVC image databases, while its performance is relatively poor on TID, A57, CSIQ, WIQ, and the two video databases. This can be explained by considering the distortion levels in the databases. In LIVE, Toyama, and IVC image databases, the distortion levels are relatively higher (i.e., suprathreshold) and more clearly visible. Due to higher amounts of distortion, any change in the visual signal usually corresponds to a similar magnitude of the reduction in visual quality and hence the prediction accuracy of DPS is reasonable. In contrast to this, the distortion levels in databases, such as TID, A57, CSIQ, and WIQ is lower and many images are with near-threshold distortions (just noticeable). In this case, the change in the signal due to the distortion may not necessarily imply the same loss of visual quality (for example as shown in Fig. 3, the effect of distortion is masked). As a further confirmation of this, one can observe from Table II that $C_P$ values for DPS are much lower (in comparison to other schemes except PSNR) for the first two distortion levels as compared to higher distortion levels i.e., there is a large difference in prediction accuracies of DPS for lower and higher distortion levels. Hence, DPS is overall less effective, while the proposed method tackles this much better as already explained.

The reader will recall that we used a block size of 128, i.e., $P = Q = 128$. We then used different values of $P_{\text{red}}$ and $Q_{\text{red}}$ to obtain the reduced-space representation and thus obtained a group of algorithms, which are suitable for RR scenarios. We also experimented with smaller block sizes. It was found that performance usually degraded with smaller block size. We propose two plausible reasons for this observation. First, when small blocks are employed, they are assumed to be independent which may not always be true. A more global Fourier analysis, on the other hand, can tackle the interaction/dependencies between the blocks better. Second, with decreasing block size, the number of blocks will obviously increase. It is quite possible that in such case the useful information about change in quality may be suppressed due to averaging over a large number of blocks.

## IV. CONCLUSION

Phase and magnitude are important information in images, while phase has been known to convey more useful information regarding important features, such as edges, contours. However, their use for visual quality assessment was largely unexplored. In this paper, we first used the phase and magnitude together (as a comprehensive way) to compute visual quality. We obtained an effective reduced-space representation of the image (or video frame) by non-uniform binning of the high frequency components. This was based on the fact that the human eye can tolerate more error (distortion) in high frequencies (such as texture) and error in smooth (low frequency) area is more annoying. In addition, since phase is more important with regards to image *structure*, we further explored the scalability of the proposed method by using only the phase of the reduced-space representation. A thorough experimental verification of the effectiveness of the proposed method was done, using nine publicly available image and video databases. As a unique advantage, the proposed method can serve as both a FR metric and an RR metric (with additional scalability), while the existing metrics are either FR or RR but not both.

## REFERENCES

[1] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R Standard BT.500-11, Jun. 2002.

[2] S. Winkler, "Perceptual video quality metrics-a review," in *Digital Video Image Quality and Perceptual Coding*, H. R. Wu and K. R. Rao, Eds. Boca Raton, FL: CRC Press, 2005, ch. 5.

[3] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 660–668, Sep. 2008.

[4] W. Lin and M. Narwaria, "Perceptual image quality assessment: Recent progress and trends," *Proc. SPIE*, vol. 7744, p. 774403, Jul. 2010.

[5] W. Lin and C. Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Imag. Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[6] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Imag. Process.*, vol. 13, no. 4, pp. 1–4, Apr. 2004.

[7] Y. Huang, T. Ou, P. Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614–1624, Nov. 2010.

[8] B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207–220.

[9] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 525–536, Jul. 1974.

[10] F. X. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. Commun.*, vol. 30, no. 7, pp. 679–1692, Jul. 1982.

[11] S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 179–206.

[12] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 163–178.

[13] C. Lambrecht and O. Verscheure, "Perceptual quality measure using a spatio-temporal model of the human visual system," *Proc. SPIE Digital Video Compress.: Algorithms Technol.*, vol. 2668, pp. 450–461, Jan. 1996.

[14] S. Winkler, "A perceptual distortion metric for digital color video," *Proc. SPIE Human Vis. Electron. Imag.*, vol. 3644, pp. 175–184, Jan. 1999.

[15] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Imag. Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[16] D. Chandler and S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Imag. Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[17] M. Narwaria and W. Lin, "Scalable image quality assessment based on structural vectors," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, Oct. 2009, pp. 1–6.

[18] A. M. Eskicioglu, A. Gusev, and A. Shnayderman, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Trans. Imag. Process.*, vol. 15, no. 2, pp. 422–429, Feb. 2006.

[19] M. Sendashonga and F. Labeau, "Low complexity image quality assessment using frequency domain transforms," in *Proc. Int. Conf. Imag. Process.*, 2006, pp. 385–388.

[20] C. Wee, R. Paramesran, R. Munundan, and X. Jiang, "Image quality assessment by discrete orthogonal moments," *Pattern Recognit.*, vol. 43, no. 12, pp. 4055–4068, 2010.

[21] M. Liu and X. Yang, "Image quality assessment using contourlet transform," *Opt. Eng.*, vol. 48, no. 10, p. 107201, 2009.

[22] A. Oppenheim and J. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.

[23] X. Ni and X. Huo, "Statistical interpretation of the importance of phase information in signal and image reconstruction," *Stat. Probab. Lett.* vol. 77, no. 4, pp. 447–454, 2007.

[24] P. Kovesi, "Phase preserving denoising of images," in *Proc. Australian Pattern Recognit. Soc. Conf.*, 1999, pp. 212–172.

[25] M. C. Morrone and D. C. Burr, "Feature detection in human vision: A phase-dependent energy model," *Proc. Royal Soc., London Ser. B*, vol. 235, no. 1280, pp. 221–245, 1988.

[26] K. Gegenfurtner, D. Braun, and F. Wichmann, "The importance of phase information for recognizing natural images," *J. Vis.*, vol. 3, no. 9, p. 519, 2003.

[27] B. C. Hansen and R. F. Hess, "Structural sparseness and spatial phase alignment in natural scenes," *J. Opt. Soc. Amer.*, vol. 24, no. 7, pp. 1873–1885, 2007.

[28] W. Hsiao and R. Millane, "Effects of spectral amplitude and phase errors on image reconstruction," *Proc. SPIE*, vol. 5562, no. 27, pp. 175–180, 2004.

[29] W. Hsiao and R. Millane, "Effects of Fourier-plane amplitude and phase errors on image reconstruction. I. Small amplitude errors," *J. Opt. Soc. Amer. A*, vol. 24, no. 10, pp. 3180–3188, 2007.

[30] M. Thomson, D. Foster, and R. Summers, "Human sensitivity to phase perturbations in natural images: A statistical framework," *Perception*, vol. 29, no. 9, pp. 1057–1069, 2000.

[31] G. Blanchet, L. Moisan, and B. Rouge, "Measuring the global phase coherence of the image," in *Proc. IEEE Int. Conf. Imag. Process.*, Jan. 2008, pp. 1176–1179.

[32] C. Juglin and D. Hines, "The phase correlation image alignment method," in *Proc. Int. Conf. Cybern. Soc.*, 1975, pp. 163–65.

[33] T. Alieva and M. Calvo, "Importance of the phase and amplitude in the fractional Fourier domain," *J. Opt. Soc. Amer.*, vol. 20, no. 3, pp. 533–541, 2003.

[34] J. Bretel, T. Caelli, R. Hilz, and I. Rentschler, "Modelling perceptual distortion: Amplitude and phase transmission in the human visual system," *Human Neurobiol.*, vol. 1, no. 1, pp. 61–67, 1982.

[35] Y. Tadmor and D. Tolhurst, "Both the phase and the amplitude spectrum may determine the appearance of natural images," *Vis. Res.*, vol. 33, no. 1, pp. 141–145, 1993.

[36] K. Ito, T. Aoki, E. Kosuge, R. Kawamata, and I. Kashima, "Medical image registration using phase-only correlation for distorted dental radiographs," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[37] K. Ito, T. Aoki, H. Nakajima, K. Kobayashi, and T. Higuchi, "A palmprint recognition algorithm using phase-based image matching," in *Proc. IEEE Int. Conf. Imag. Process.*, Oct. 2006, pp. 2669–2672.

[38] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal Saliency detection using phase spectrum of quaternion fourier transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[39] A. Sao and B. Yegnanarayana, "On the use of phase of the Fourier transform for face recognition under variations in illumination," *Signal, Imag. Video Process.*, vol. 4, no. 3, pp. 353–358, 2009.

[40] S. Mitra, M. Savvides, and A. Brockwell, "Modeling phase spectra using Gaussian mixture models for human face identification," in *Pattern Recognition and Image Analysis* (Lecture Notes in Computer Science), vol. 3687. New York: Springer-Verlag, 2005, pp. 174–182.

[41] S. Rajagopalan and R. Robb, "Phase based image quality assessment," *Proc. SPIE Med. Imag. Percept., Observer Perform., Technol. Assessment*, vol. 5749, pp. 373–382, Feb. 2005.

[42] P. Skurowski and A. Gruca, "Image quality assessment using phase spectrum correlation," in *Computer Vision and Graphics* (Lecture Notes in Computer Science), vol. 5337. New York: Springer-Verlag, 2009, pp. 80–89.

[43] W. Lin, L. Dong, and P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 900–909, Jul. 2005.

[44] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *Proc. 3rd Int. Workshop Video Process. Qual. Metrics Consumer Electron.*, Jan. 2007, pp. 1–4.

[45] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack. *Image and Video Quality Assessment Research at LIVE* [Online]. Available: http://live.ece.utexas.edu/research/quality/

[46] E. C. Larson and D. M. Chandler. *Categorical Image Quality (CSIQ) Database* [Online]. Available: http://vision.okstate.edu/csiq

[47] A. Ninassi, P. L. Callet, and F. Autrusseau. *Subjective Quality Assessment-IVC Database* [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb

[48] Y. Horita, K. Shibata, Y. Kawayoke, and Z. Sazzad. *MICT Image Quality Evaluation Database* [Online]. Available: http://mict.eng.u-toyama.ac.jp/mict/index2.html

[49] D. M. Chandler and S. S. Hemami. *VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images* [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html

[50] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti, "Color image database for evaluation of image quality metrics," in *Proc. Int. Workshop Multimedia Signal Process.*, Oct. 2008, pp. 403–408.

[51] U. Engelke, H.-J. Zepernick, and M. Kusuma. (2010). *Wireless Imaging Quality Database* [Online]. Available: http://www.bth.se/tek/rcg.nsf/pages/wiq-db

[52] R. Kakarala, "Signal processing approach to Fourier analysis of ranking data: The importance of phase," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1518–1527, Apr. 2011.

[53] F. Simone, M. Naccari, M. Tagliasacchi, F. C. Dufaux, S. Tubaro, and T. Ebrahimi, "Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel," in *Proc. 1st Int. Workshop Qual. Multimedia Experience*, 2009, pp. 204–209.

[54] Y. Wang. (2003, Aug.). *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II* [Online]. Available: http://www.vqeg.org

[55] G. Chen, C. Yang, and S. Xie, "Gradient-based structural similarity for image quality assessment," in *Proc. Int. Conf. Imag. Process.*, 2006, pp. 2929–2932.

[56] G. Cheng, J. Huang, C. Zhu, Z. Liu, and L. Cheng, "Perceptual image quality assessment using a geometric structural distortion model," in *Proc. Int. Conf. Imag. Process.*, 2010, pp. 325–328.

[57] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[58] A. Rehman and Z. Wang, "Reduced-reference SSIM estimation," in *Proc. IEEE Int. Conf. Imag. Process.*, Sep. 2010, pp. 289–292.

[59] Z. Wang, E. Simoncelli, and A. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2003, pp. 1398–1402.

[60] S. Cakir and A. Cetin, "Mel-cepstral feature extraction methods for image representation," *Opt. Eng.*, vol. 49, no. 9, p. 097004, 2010.

[61] V. Ojansivu and J. Heikkilä, "Image registration using blur-invariant phase correlation," *IEEE Signal Process. Lett.*, vol. 14, no. 7, pp. 449–452, Jul. 2007.

[62] K. Castleman, *Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[63] Y. Fang, W. Lin, C. Lau, and B. Lee, "A visual attention model combining top-down and bottom-up mechanisms for salient object detection," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 1293–1296.

[64] N. Skarbnik, C. Sagiv, and Y. Zeevi, "Edge detection and skeletonization using quantized localized phase," in *Proc. 17th Eur. Signal Process. Conf.*, 2009, pp. 1542–1546.

[65] T. Huang, J. Burnett, and A. Deczky, "The importance of phase in image processing filters," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 23, no. 6, pp. 529–542, Dec. 1975.

[66] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, Mar. 2010.

**Manish Narwaria** received the B.Tech. degree in electronics and communication engineering from Amrita Vishwa Vidyapeetham, Coimbatore, India, in 2008. He is currently pursuing the Ph.D. degree with the School of Computer Engineering, Nanyang Technological University, Singapore.

His current research interests include image/video and speech quality assessment, pattern recognition, and machine learning.

**Weisi Lin** (M'92–SM'98) received the B.Sc. and M.Sc. degrees from Zhongshan University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree from Kings College, London University, London, U.K., in 1992.

He is currently an Associate Professor and the double degree (Business and Computing) Program Coordinator with the School of Computer Engineering, Nanyang Technological University, Singapore. He is a Chartered Engineer. He has authored over 190 scholarly publications. He has been a recipient of over $3.8 million in research grant funding since 1997. His current research interests include image processing, video compression, perceptual visual and audio modeling, computer vision, and multimedia communication.

Dr. Lin is a fellow of the Institution of Engineering and Technology. He serves as an Associate Editor for the IEEE TRANSACTION ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Visual Communication and Image Representation*.

**Ian McLoughlin** received the Ph.D. degree in audio signal processing from the University of Birmingham, Birmingham, U.K., in 1997, funded by Philips/Simoco Telecommunications, Cambridge, U.K.

He was with the GEC Hirst Research Centre, a part of General Electric Company Inc., and later Her Majesty's Government Communications Centre. In 1998, he began over four years of lecturing in Singapore, then five years as a Principal Engineer with Tait Electronics, Group Research, Christ Church, New Zealand. In 2006, he was an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He was involved in the field of speech research and embedded systems for 20 years.

**Sabu Emmanuel** received the B.E. degree in electronics and communication engineering from Regional Engineering College, Durgapur, India, in 1988, the M.E. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1998, and the Ph.D. degree in computer science from the National University of Singapore, Singapore, in 2002.

He is currently an Assistant Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include multimedia and software security and surveillance video processing.

Dr. Emmanuel has served as the Guest Editor for special issues of several journals and also Reviewer for several journals, such as *Springer Multimedia Systems*, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

**Liang-Tien Chia** (M'90) received the B.S. and Ph.D. degrees from the Loughborough University of Technology, Loughborough, U.K., in 1990 and 1994, respectively.

He is an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He was the Director of the Centre for Multimedia and Network Communications, School of Computer Engineering, Nanyang Technological University, from 2002 to 2007, and is currently the Head of the Division of Computer Communications. He has published over 100 refereed research papers. His current research interests include the areas of internet-related research with emphasis on the semantic web, multimedia understanding through media analysis, annotation, and adaptation, multimodal data fusion, and multimodality ontology for multimedia.