Image quality assessment using two-dimensional complex mel-cepstrum

Serdar Cakir
A. Enis Cetin

Image quality assessment using two-dimensional complex mel-cepstrum

Serdar Cakir* and A. Enis Cetin
Bilkent University, Department of Electrical and Electronics Engineering, Bilkent, Ankara TR 06800, Turkey

Abstract. Assessment of visual quality plays a crucial role in modeling, implementation, and optimization of image- and video-processing applications. The image quality assessment (IQA) techniques basically extract features from the images to generate objective scores. Feature-based IQA methods generally consist of two complementary phases: (1) feature extraction and (2) feature pooling. For feature extraction in the IQA framework, various algorithms have been used and recently, the two-dimensional (2-D) mel-cepstrum (2-DMC) feature extraction scheme has provided promising results in a feature-based IQA framework. However, the 2-DMC feature extraction scheme completely loses image-phase information that may contain high-frequency characteristics and important structural components of the image. In this work, "2-D complex mel-cepstrum" is proposed for feature extraction in an IQA framework. The method tries to integrate Fourier transform phase information into the 2-DMC, which was shown to be an efficient feature extraction scheme for assessment of image quality. Support vector regression is used for feature pooling that provides mapping between the proposed features and the subjective scores. Experimental results show that the proposed technique obtains promising results for the IQA problem by making use of the image-phase information. © 2016 SPIE and IS&T [DOI: 10.1117/1.JEI.25.6.061604]

Keywords: image quality assessment; image phase; two-dimensional complex mel-cepstrum; two-dimensional mel-cepstrum; cepstral features; support vector regression.

Paper 16366SS received Apr. 30, 2016; accepted for publication Jul. 12, 2016; published online Aug. 4, 2016.

1 Introduction and Related Work

Visual-quality assessment is an important process that plays a key role in the modeling, implementation, optimization, and testing of image-/video-processing algorithms and multimedia services. Despite the recent advances in coding technologies, the transmitted data suffer from both lossy source encoding and the losses induced by transmission channels.[1] This results in a degradation of quality in the images and videos. Therefore, it is crucial to establish a criterion to measure the perceived quality. Since the statement "quality" is generally considered as a subjective term, the subjective evaluation is formally defined to be the most accurate and reliable tool in the assessment of visual quality if the number of subjects is sufficiently large.[2] However, the subjective evaluation is impractical, tedious, and unsuitable to be used in real-time applications. Moreover, the environmental conditions of the subjects and the contents of the images[3] directly affect the evaluation process. Therefore, it is important to develop an objective quality measure and significant amount of research has been carried out to obtain such a measure.[4] The visual-quality metrics can be divided into two main categories: (1) "signal fidelity measures" and (2) "perceptual visual-quality metrics (PVQM)."

The signal fidelity measures including mean absolute error, mean square error (MSE), signal-to-noise ratio (SNR), and peak SNR (PSNR) are simple and well-defined measures.[5] However, these measures perform poorly when the noise is not additive.[6] Although these measures respond appreciably to the changes in picture quality, it has been acknowledged by Girod[7] that the signal fidelity measures do not coincide well with the human visual system (HVS). The inconsistency between the MSE-based measures and the perceived quality was also illustrated by Dosselmann and Young[8] through the equal MSE hypersphere concept.

The PVQMs can be divided into three subcategories based on their dependency on the reference information, namely: (1) full-reference, (2) reduced-reference, and (3) no-reference PVQMs. The PVQMs which use some part of the reference signal instead of using it as a whole, are called reduced-reference metrics.[9] If the whole reference signal is used in the evaluation of the PVQM, then this PVQM is called a full-reference metric.[10] The no-reference metrics use only the test signal and, therefore, they are called no-reference metrics.[11,12]

The structural similarity index (SSIM) developed by Wang et al.[13] is one of the most common full-reference metrics available in the literature. The method tries to model the structural losses in a given image. The multiscale extension of SSIM, namely MS-SSIM,[14] achieves better performance than ordinary SSIM. As the first version of the SSIM, the universal quality index (UQI) had been proposed by Wang and Bovik[15] to model the loss of quality in terms of (1) "loss of correlation," (2) "luminance distortion," and (3) "contrast distortion." Information theory has also been applied to image quality assessment and the visual information fidelity (VIF) metric that tries to quantify the information shared by the reference and distorted images has been proposed by Sheikh and Bovik.[16] Damera-Venkata et al.[17] proposed the noise quality measure (NQM) full-reference

*Address all correspondence to: Serdar Cakir, E-mail: cakir@ee.bilkent.edu.tr

framework that tries to model the degradations in the distorted images using reference images. A multistage metric, visual signal-to-noise ratio (VSNR) is proposed by Chandler and Hemami[18] for image quality assessment. In the first stage of VSNR computation, the visibility of the distortion in the image is examined with a contrast thresholding-based scheme. In the second stage, perceived contrast and visual property of global precedence-based framework is applied, when the distortion in the image is above the threshold.

Feature extraction-based PVQMs are developed to model the degradations in an image. Local variance- and correlation-based structural schemes, which are important milestones in image quality assessment,[13] can also be considered as a feature extraction-based method. Moreover, discrete cosine transform (DCT)- and discrete wavelet transform (DWT)-based transform domain features are also used to measure the quality of the images.[19] Promising results obtained by DCT- and DWT-based frameworks encourage researchers to develop quality measures that work in the transform domain.[20–22] As another feature extraction scheme, Shnayderman et al.[23] used singular value decomposition (SVD) to define a quality measure for the images degraded by different levels and different types of distortions. The method computes the distance between the eigenvalues of image subregions on the reference and distorted images. The distances calculated at different subregions are accumulated to obtain a quality score. In another work, Narwaria and Lin[24] used the SVD-based features together with a machine-learning framework.

After the features are extracted from the images, a final objective quality score based on the selected features has to be determined. The process of selecting or integrating features to obtain a quality score is generally called "feature pooling," which is an important concept in the PVQM framework. Feature pooling is performed by using simple summation,[25] weighted combination,[26] and other types of pooling regimes[27] in the literature. Recently, Narwaria and Lin[24] have used regression-based machine-learning techniques for feature pooling. However, the utilization of machine-learning theory in the image quality assessment problem is not new. Researchers have proposed machine-learning approaches to define quality measures[28,29] in the literature. Due to their generalization capabilities and ability to handle high-dimensional data, machine-learning-based techniques have the potential to provide better solutions for PVQM. A full-reference PVQM evaluation composed of feature detection and pooling stages is proposed by Narwaria and Lin.[30] In the feature detection stage, SVDs of the original and perturbed images are calculated. After detection of the features, feature pooling is carried out using the support vector regression (SVR). The main aim is to evaluate the deviations present in the singular vectors of the original and perturbed images.

Narwaria et al.[31] have proposed another PVQM framework, which is similar to the one proposed in their earlier study.[30] The authors used two-dimensional (2-D) mel-cepstrum (2-DMC) features instead of SVD in the feature extraction stage of the PVQM.[31] The 2-DMC features, which originate from the cepstral analysis, have been used for image representation by Cakir and Cetin.[32,33] Cepstral feature extraction techniques have drawn attention due to their representative power in speech-processing applications

including speech and sound recognition and speaker identification.[34–37] In the literature, the 2-D cepstrum, which is the natural extension of one dimensional (1-D) cepstrum, was used for shadow detection, echo removal, automatic intensity control, enhancement of repetitive features, and cepstral filtering.[38–40] The 2-DMC-based IQA scheme proposed by Narwaria et al.[31] was shown to outperform the baseline techniques; however, the method does not consider the image-phase information that may be a significant component in the quality assessment problem.

The phase of the Fourier transform (FT) is extremely important in image representation.[41,42] Oppenheim et al.[41] carried out a group of experiments on the images to observe the effects of FT phase information in image synthesis. In almost all images, edges and texture are preserved in the synthesis even if the spectral magnitude is totally ignored and set to a constant value. Due to its significance in image representation, FT phase is often utilized in various image-processing applications including image quality assessment.[43,44] These techniques try to integrate the image-phase information with the magnitude by using a rule-based scheme.

In this article, a full-reference PVQM that contains feature-detection and feature-pooling operations is proposed. A set of features, the 2-D complex mel-cepstrum, for feature extraction is introduced. The proposed 2-D complex mel-cepstrum features contain image-phase information, which is obtained using the complex logarithm of the FT in mel-scale. Feature pooling is carried out using SVR, which is a machine-learning framework. This paper is organized as follows. In Sec. 2, the 2-DMC extraction procedure is reviewed. The proposed 2-D mel-complex cepstrum feature-extraction scheme is described in Sec. 3. The SVR-based feature-pooling regime is presented in Sec. 4. Experimental studies and concluding remarks are presented in Secs. 5 and 6, respectively.

2 Two-Dimensional Mel-Cepstrum Feature Extraction

A 2-DMC has originated from the ordinary 2-D cepstrum of a 2-D signal. A 2-D cepstrum $\Psi(a, b)$ of a 2-D image $y(a, b)$ can be defined as follows:

$$\Psi(a, b) = F_2^{-1}\{\log[|Y(u, v)|^2]\}, \tag{1}$$

where (a, b) denotes 2-D cepstral frequency coordinates, F_2^{-1} denotes 2-D inverse discrete-time Fourier transform (IDTFT), and $Y(u, v)$ is the 2-D discrete-time Fourier transform (DTFT) of the image $y(a, b)$. In general, fast Fourier transform (FFT) algorithm is preferred to compute the DTFT.

The 2-DMC divides the DTFT domain into different-sized regions by making use of the structural nature of the Fourier transform. In other words, the DTFT domain is divided into nonuniform bins in a logarithmic manner as shown in Fig. 1 and the energy $|G(m, n)|^2$ of each bin is computed as follows:

$$|G(m, n)|^2 = \sum_{u,v \in B(m,n)} |Y(u, v)|^2, \tag{2}$$

where $Y(u, v)$ is the discrete Fourier transform (DFT) of $y(a, b)$ and $B(m, n)$ is the cell of the logarithmic grid corresponding to the (m, n)-th location. The bin sizes are smaller at low frequencies compared to high frequencies.
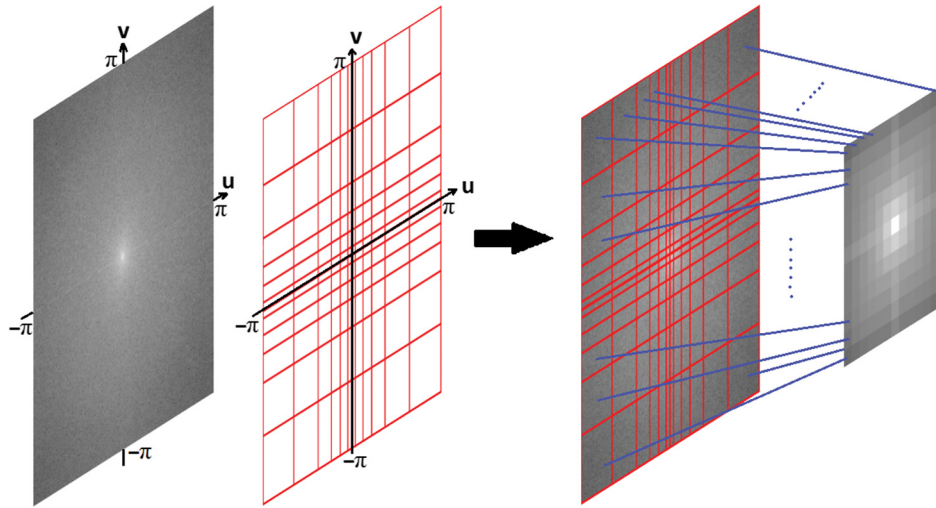
Fig. 1 A representative nonuniform 2-DMC grid in the DTFT domain. Cell sizes are smaller at low frequencies compared to high frequencies. The resultant matrix has smaller dimensions than the original spectrum since each frequency bin produces only one coefficient.

The concept of dividing the transform domain into different-sized regions is similar to the mel-cepstrum computation in speech processing. Similar to speech signals, most natural images are low-pass in nature. Therefore, there is more signal energy at low frequencies compared to high frequencies. Logarithmic division of the DFT grid in such a manner emphasizes high frequencies. After DFT gridding, 2-D mel-frequency cepstral coefficients $\Psi^{mel}(p, q)$ are computed using inverse DFT (IDFT) as follows:

$$\Psi^{mel}(p, q) = F_2^{-1}\{\log[|G(m, n)|^2]\}. \quad (3)$$

Due to the logarithmic grid shown in Fig. 1, the size of the forward DFT used to compute $Y(u, v)$ is larger than the size of the IDFT. Similar to the use of mel-cepstrum in speech processing, to emphasize certain bands, different weights may also be assigned to different frequency bins. The dimensions of the 2-DMC sequence, which is computed using the IDFT, is smaller than the original image because several DFT values are grouped together in each bin. Note that the image is selected to be a square image for ease of representation. The steps of the 2-DMC-based feature extraction scheme are summarized as follows.

- $N \times N$ 2-D DFT of input images are calculated. Before the DFT calculation, the images are padded with zeros up to the size that is the closest power of 2 to take advantage of the FFT algorithm during DFT computation.
- The nonuniform DTFT grid is applied to the resultant DFT matrix and the energy $|G(m, n)|^2$ of each cell is computed. Each cell of the grid can also be weighted with a coefficient. The data size is $M \times M$, where $M \leq N$.
- Logarithm of cell energies $|G(m, n)|^2$ are computed.
- 2-D IDFT of the $M \times M$ data is computed to get the $M \times M$ mel-cepstrum sequence.

In an image, the contributions to the high frequencies are generally made by edges and structural features. For the purpose of extracting better representative features, the cells of the 2-D DFT grid, which correspond to the high-frequency components, are multiplied with higher weights compared to those used for the low-frequency cells in the grid. This ensures that high-frequency components are further emphasized.

An important feature is the invariance of the cepstrum to pixel amplitude changes. In this manner, robustness to illumination invariance can be achieved. Let $Y(u, v)$ denote the 2-D DTFT of a given image $y(a, b)$ and $y_2(a, b) = \alpha y(a, b)$ has a DTFT $\alpha Y(u, v)$ for any real constant α. The log spectrum of $\alpha Y(u, v)$ is given as follows:

$$\log[|\alpha Y(u, v)|] = \log(|\alpha|) + \log[|Y(u, v)|] \quad (4)$$

and the corresponding cepstrum is given as follows:

$$\Psi_2(a, b) = \kappa\delta(a, b) + \Psi(a, b), \quad (5)$$

where $\delta(a, b) = 1$ for $a = b = 0$ and $\delta(a, b) = 0$, otherwise. Therefore, the cepstrum values except at the (0,0) location (DC term) do not vary with amplitude changes. Since the FT magnitudes of $y(a, b)$ and $y(a - c, b - d)$ are the same, the 2-D cepstrum and mel-cepstrum are shift invariant features.

Another important characteristic of the 2-D cepstrum is symmetry with respect to $\Psi(a, b) = \Psi(-a, -b)$. As a result only half of the 2-D cepstrum or $M \times M$ 2-DMC coefficients is enough when IDFT is used. The 2-DMC method that is originated from the ordinary 2-D cepstrum also has the same properties.

A more detailed discussion and derivations for 2-DMC can be found in our earlier studies.[32,33] Due to its representative properties and the success in the face recognition problem,[32] the 2-DMC feature extraction scheme was proposed for the image quality assessment problem and achieved promising results.[31]

3 Two-Dimensional Complex Mel-Cepstrum-Based Feature Extraction

Although the 2-DMC feature extraction scheme has achieved promising results for image representation,[31–33] the method
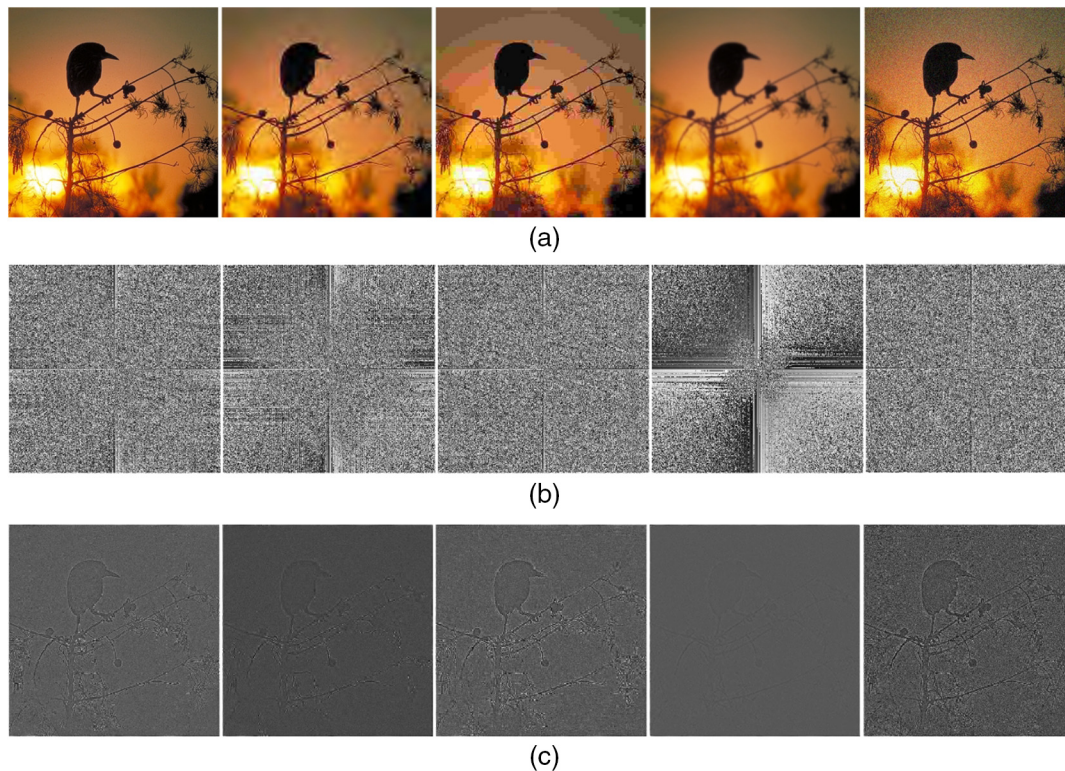
Fig. 2 The illustration of FT phase information and phase-only synthesis of reference and distorted images. (a) Reference Image, JPEG2000 compressed image, JPEG compressed image, Gaussian blurred image, and the noisy image (image size: 512 × 512). (b) FT phase information of reference, JPEG2000 compressed, JPEG compressed, Gaussian blurred, and the noisy images. (c) Image synthesis using only image-phase information. Magnitude of the FT is set as unit matrix.

completely ignores the image-phase information that may possess superior representation capability. The experiments carried out by Oppenheim et al.[41] revealed that by using only the FT phase information, high-frequency components such as edges and texture are preserved in the synthesis even if the spectral magnitude is totally ignored and set to a constant value. The FT phase information of a set of images and the synthesis results obtained by using the approach proposed by Oppenheim et al.[41] for these images are presented in Fig. 2.

By looking at the synthesis results presented in Fig. 2, it can be observed that the image-phase information contains descriptive high-frequency characteristics, since the synthesis image preserves most of the edges and texture details of the input image. Moreover, high-frequency characteristics of the image are still present in the image-phase information after exposure to different types of distortions. The illustrations of FT phase matrices in Fig. 2 also show that different types of distortions cause the FT phase information to distribute in different manners in the spectrum. Since we deal with different types and different levels of distortion in image quality assessment problem, the FT phase information, which disperses differently in the spectrum under different degradation types, may be a useful feature for signal representation. In addition, frequency binning using a nonuniform grid in the 2-DMC computation can be applied to FT phase information to extract phase characteristics in local frequency bands. Therefore, due to its possible contribution to signal representation and its discriminative distribution in the spectrum, we plan to integrate FT phase information

with the 2-DMC features to use it in a feature-based image quality assessment scheme. Here, we try to seek the possibility of deriving 2-D complex mel-cepstrum features based on complex cepstrum theory.[41]

A 1-D complex cepstrum was defined by Oppenheim and his coworkers.[41,45,46] The natural 2-D extension of the complex cepstrum was also proposed to be used in image-processing applications such as echo removal, automatic intensity control, and enhancement of repetitive features.[39] Recalling that $Y(u, v)$ is the DTFT of the image $y(a, b)$, the $Y(u, v)$ may contain artifacts commonly known as "cross-effect" due to the implicit assumption about the image's periodicity. However, natural images contain significant discontinuities especially across the frame borders that may cause cross-shape artifacts in the Fourier spectrum. Therefore, before further processing, the image is exposed to a "periodization" step[47] to eliminate the artifacts that degrade the spectrum. The periodization algorithm[47] used herein decomposes the input image into two components, one of which is a periodic image while the other is a smooth component except for sharp transitions near the frame borders. The main idea behind the algorithm is finding a linear decomposition, which provides a periodic component whose spectrum is close to the spectrum of the input image as well as being artifact free. However, the latter component captures all the edge artifacts arising from the implicit periodization of the input image. When the input image is directly padded with zeros to produce an image of size power of twos, this padding operation may cause undesired artifacts in the image. However, prior periodization eliminates the undesired
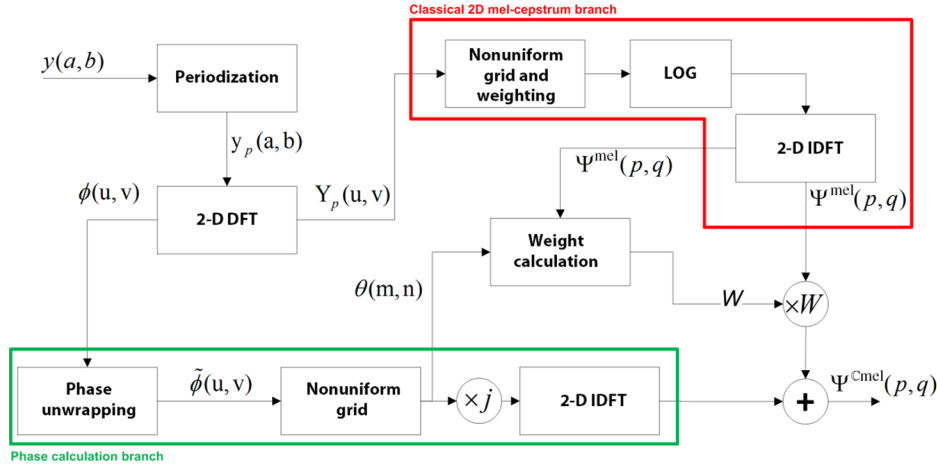
Fig. 3 Flow diagram of the proposed 2-D complex mel-cepstrum feature extraction scheme.

effects caused by direct padding. The $Y_p(u, v)$ is calculated as the DTFT of the periodized image $y_p(a, b)$ and used in the 2-D complex cepstrum computation as follows:

$$
\begin{aligned}
\Psi^{\mathbb{C}}(a, b) &= F_2^{-1}\{\log[Y_p(u, v)]\} \\
&= F_2^{-1}\{\log[|Y_p(u, v)|e^{j\phi(u,v)}]\} \\
&= F_2^{-1}\{\log[|Y_p(u, v)|] + \log(e^{j\phi(u,v)})\} \\
&= F_2^{-1}\{\log[|Y_p(u, v)|]\} + F_2^{-1}\{j\phi(u, v)\}.
\end{aligned} \quad (6)
$$

In the final calculation step of Eq. (6), we have the real 2-D cepstrum of the image on the first summand and inverse FT of the image phase on the second summand. The right summand is due to the complex logarithm operation, which has to be continuous. Therefore, the discontinuities in phase should be unwrapped to compute the complex cepstrum correctly. In the literature, several phase-unwrapping algorithms have been developed to calculate complex cepstrum in an appropriate manner.[46,48–53] The phase unwrapping is carried out by adding multiples of 2π to appropriate phase locations for a smooth phase function. The phase-unwrapping algorithm used in this work is based on the local phase gradient. Each element of the phase is checked as to whether the sign of the phase gradient is consistent with its consecutive neighbors. If the sign of the phase gradient changes, the phase element is modified by adding/subtracting integer multiples of 2π. Otherwise, the phase element remains unchanged. This way, undesired phase transitions are eliminated and the phase is modified to be smooth.

In this article, the 2-D complex mel-cepstrum is defined for feature extraction from images. The 2-DMC is known to be an efficient feature extraction technique in face recognition[32] and image quality assessment[31] problems. The 2-D complex mel-cepstrum is computed similarly to the 2-DMC by combining discrete Fourier domain coefficients in logarithmic bins. However, both magnitude and phase terms are averaged in the 2-D complex mel-cepstrum. As in regular 2-D complex cepstrum, phase terms in a given logarithmic bin must be unwrapped before averaging. Let $\tilde{\phi}(u, v)$ be the unwrapped phase of the image FT. In each logarithmic bin, unwrapped phase coefficients are averaged as follows:

$$
\theta(m, n) = \frac{1}{\mu_{m,n}} \sum_{u,v \in B(m,n)} \tilde{\phi}(u, v), \quad (7)
$$

where $\theta(m, n)$ is the mean phase value of the (m, n)-th bin and $\mu_{m,n}$ is the number of 2-D DFT coefficients in the corresponding bin $B(m, n)$. By using the logarithmic bin energy values and phase information defined in Eqs. (2) and (7), the proposed 2-D complex mel-cepstrum coefficients $[\Psi^{\mathbb{C}mel}(p, q)]$ are calculated as follows:

$$
\begin{aligned}
\Psi^{\mathbb{C}mel}(p, q) &= F_2^{-1}\{\log[|G(m, n)|e^{j\theta(m,n)}]\} \\
&= F_2^{-1}\{\log[|G(m, n)|] + \log(e^{j\theta(m,n)})\} \\
&= F_2^{-1}\{\log[|G(m, n)|]\} + F_2^{-1}\{j\theta(m, n)\} \\
\Psi^{\mathbb{C}mel}(p, q) &= W \cdot \Psi^{mel}(p, q) + F_2^{-1}\{j\theta(m, n)\}.
\end{aligned} \quad (8)
$$

The 2-D complex mel-cepstrum computation integrates the $\Psi^{mel}(p, q)$ with the inverse FT of the unwrapped phase information as shown in Eq. (8). To establish an appropriate balance between the energy and phase terms in the computation [Eq. (8)], $\Psi^{mel}(p, q)$ should be calculated by using the appropriate weight. The weight (W) used in $\Psi^{mel}(p, q)$ computation is calculated as follows:

$$
\begin{aligned}
W &= \underset{W^*}{\arg\min}\{|\bar{\theta} - W^* \cdot \bar{\Psi}^{mel}|\}, \\
\bar{\theta} &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \theta(i, j), \\
\bar{\Psi}^{mel} &= \frac{1}{N^2} \sum_{p=1}^{N} \sum_{q=1}^{N} \Psi^{mel}(p, q).
\end{aligned} \quad (9)
$$

In this way, the domination of either the energy or phase components is avoided in the overall 2-D complex mel-cepstrum computation.

The flow diagram of the proposed 2-D complex mel-cepstrum feature extraction scheme is presented in Fig. 3. The implicit periodization is carried out as a preprocessing step before the computation of cepstral coefficients.

By looking at the flow diagram presented in Fig. 3, one can say that the proposed 2-D complex mel-cepstrum
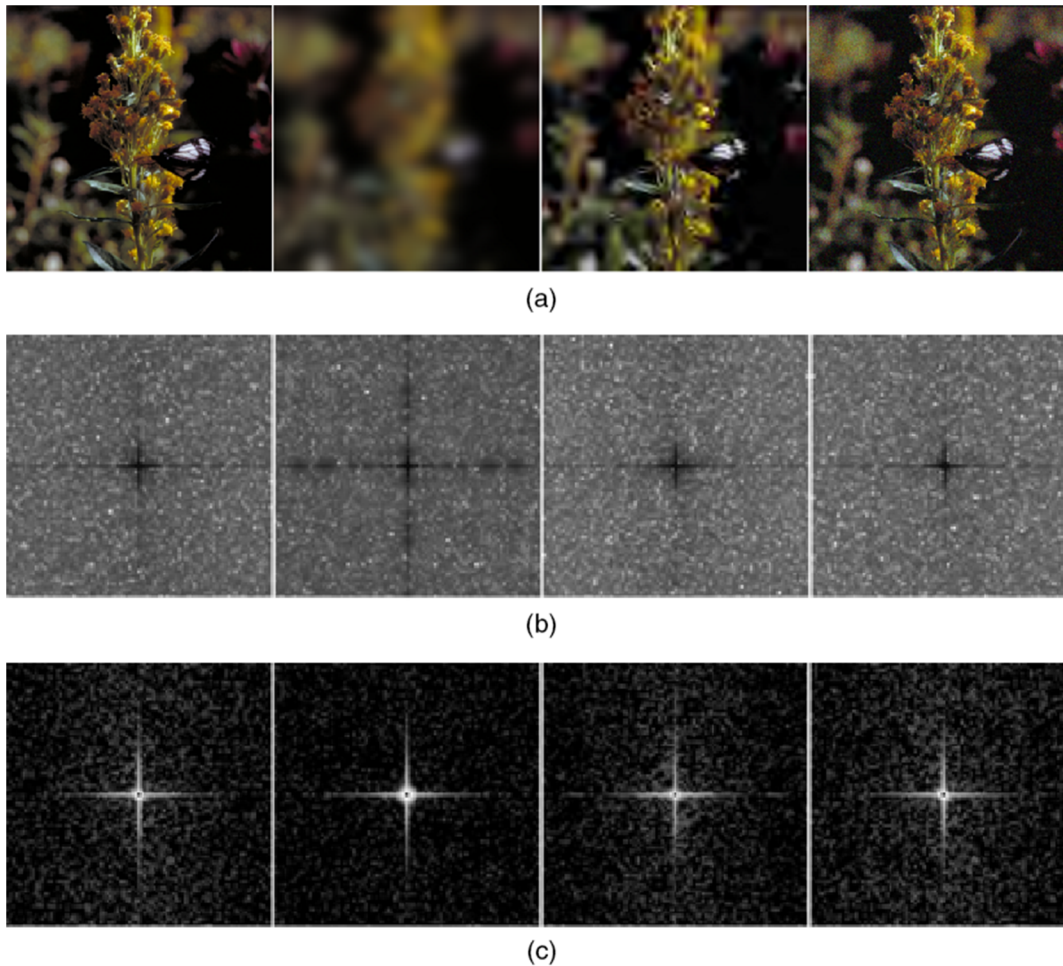
Fig. 4 Example of 2-DMC and 2-D complex mel-cepstrum features extracted from reference and distorted images. (a) Reference image, Gaussian blurred image, JPEG2000 compressed image, and the noisy image (image size: 512×512). (b) 2-DMC coefficients corresponding to reference, Gaussian blurred, JPEG2000 compressed, and the noisy images (feature size: 85×85). (c) 2-D complex mel-cepstrum coefficients corresponding to reference, Gaussian blurred, JPEG2000 compressed, and the noisy images (feature size: 85×85).

scheme introduces a phase branch (left branch in Fig. 3) in the calculations. The classical 2-DMC computation $[\Psi^{\text{mel}}(p, q)]$ is carried out in the right branch of Fig. 3. At the end, the 2-D complex mel-cepstrum features are obtained by the integration of the classical 2-DMC with the phase information. Similar to the classical mel-cepstrum computation, the proposed technique also provides a reduced feature representation by making use of nonuniform binning in both magnitude and phase components. Recall that the 2-D cepstrum is symmetric with respect to $\Psi(a, b) = \Psi(-a, -b)$ and enables a further dimensionality reduction in the feature space. This is caused by the conjugate-symmetric structure of the FT coefficients and it is clear that the amplitude of the complex FT pairs are the same. But, the proposed 2-D complex mel-cepstrum features do not provide dimensionality reduction since the phase information is not symmetric with respect to DC term, i.e., (0,0) frequency location. This is due to the phase differences between the conjugate-symmetric FT pairs and the nature of the phase-unwrapping algorithm. The main aim of the phase-unwrapping algorithm is eliminating the local discontinuities in the image phase. While eliminating these discontinuities, it does

not try to preserve the overall symmetry since it may damage the phase information.

In Fig. 4, examples of 2-D complex mel-cepstrum coefficient matrices extracted from both reference and distorted images are presented. The images are taken from CSIQ image dataset[54] and the magnitude of the 2-D complex mel-cepstrum matrices are calculated and then saturated for visualization since the proposed features are complex-valued matrices.

4 Feature Pooling with Support Vector Regression

To calculate objective quality scores from the cepstral feature space, a regression framework is carried out to perform mapping between the high-dimensional feature space and subjective scores. The regression framework used in this study is SVR, namely ϵ-SVR. Instead of reimplementing the SVR algorithm, the library for support vector machines (LIBSVM) package[55] which provides an effective implementation and easy-to-use software, is used. The SVR accepts the features in terms of vectors, and therefore, the cepstral feature matrices $[\Psi^{\text{mel}}(p, q), \Psi^{\mathbb{C}\text{mel}}(p, q)]$ are converted into vectors before the SVR algorithm. Let ϑ_R and ϑ_D be the cepstral

feature vectors extracted from the reference and distorted images, respectively. The image degradations, i.e., distortions, are modeled by defining the feature difference vector as follows:

$$\mathbf{x} = |\boldsymbol{\vartheta}_R - \boldsymbol{\vartheta}_D|. \tag{10}$$

More specifically, let $\mathbf{x}_i$, $i = 1,2,\ldots,n_{\text{tr}}$ denote the difference feature vector corresponding to the i'th image in the training dataset and n_{tr} denotes the number of images used for training. In ϵ-SVR, the main aim is to find a function f that maps the feature vectors to subjective quality evaluation score (s_i) while pushing the standard deviation smaller than ϵ. While satisfying these requirements, the function $f(\mathbf{x}_i)$ should also be as flat as possible.[56] The function to be learned is $f(\mathbf{x}) = \boldsymbol{\beta}^T \varphi(\mathbf{x}) + \gamma$, where $\varphi(\mathbf{x})$ is a nonlinear function of $\mathbf{x}$, $\boldsymbol{\beta}$ is weight vector, and γ is bias term. The weight vector $\boldsymbol{\beta}$ is calculated as follows:

$$|s_i - f(\mathbf{x}_i)| \leq \varepsilon, \quad \boldsymbol{\beta} = \sum_{i=1}^{n_{\text{sv}}} (\eta_i^* - \eta_i)\varphi(\mathbf{x}_i), \tag{11}$$

where n_{sv} denotes the number of support vectors obtained in the regression process and η_i^*, η_i are the Lagrange multipliers used in the function optimization. If the optimization problem is solved, the function to be learned becomes as follows:

$$
\begin{aligned}
f(\mathbf{x}) &= \boldsymbol{\beta}^T \varphi(\mathbf{x}) + \gamma \\
&= \sum_{i=1}^{n_{sv}} (\eta_i^* - \eta_i)\varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}) + \gamma \\
&= \sum_{i=1}^{n_{sv}} (\eta_i^* - \eta_i) K(\mathbf{x}_i, \mathbf{x}) + \gamma \\
&= \sum_{i=1}^{n_{sv}} (\eta_i^* - \eta_i) \exp(-\rho\|\mathbf{x}_i - \mathbf{x}\|^2) + \gamma.
\end{aligned}
\tag{12}
$$

Note that, the kernel function $K(\mathbf{x}_i, \mathbf{x})$ is selected as radial basis function kernel.

After the function is learned, the quality score Q is obtained by $Q = f(\mathbf{x})$. In this way, the SVR maps the high-dimensional image data to a single quality score.

5 Experimental Studies

In this work, the proposed technique is compared with baseline techniques over eight different image quality databases. Each feature extraction scheme is followed by a feature-pooling phase that is carried out by using SVR. The quality scores obtained for each feature extraction scheme and image quality database are compared with three correlation-based measures. In Sections 5.1 and 5.2, the experimental procedures are explained in detail.

5.1 Image Quality Databases

In the performance evaluation of different IQA schemes, eight different image databases (A57,[57] TID2008,[58] IRCCyN/IVC,[59] LIVE,[60] CSIQ,[54] Toyoma,[61] VCL@FER,[62] and WIQ[63]) are used. Basic properties of each dataset, namely, number of reference images, number of distorted images, number of distortion types, number of distortion levels, image properties, and information about subjective evaluation are listed in Table 1. Some of the attributes in Table 1 are stated as "Varying" since these attributes may change for different reference images.

5.2 Experimental Procedures and Results

In the experiments, a k-fold cross-validation procedure is followed and the image databases used in the experiments are divided into equally sized partitions. These partitions are formed such that none of the partitions has distorted images obtained using the same reference image to prevent testing with the images used in the training phase. In k-fold cross validation, the overall dataset is divided into k partitions. At each experimental step, the partition used in the test phase is changed and the remaining $k - 1$ partitions are used in the training phase of the SVR. The experiments are repeated until every single partition is used in the test

Table 1 Information about image quality datasets.

Database	# Reference images	# Distortion types	# Distortion levels	# Distorted images	Image properties			Subjective evaluation	
					Image size	Color	Bit depth	Score type	Scale
A57	3	6	3	54	512 × 512	Gray scale	8-bit	PD	[0,1]
TID2008	25	17	4	1700	512 × 384	RGB	24-bit	MOS	[0,9]
IRCCyN/IVC	10	4	Varying	185	512 × 512	RGB	24-bit	MOS	[1,5]
LIVE	29	5	Varying	779	768 × 512	RGB	24-bit	DMOS	[1,100]
CSIQ	30	6	4/5	866	512 × 512	RGB	24-bit	DMOS	[0,1]
Toyama	14	2	6	168	768 × 512	RGB	24-bit	MOS	[1,5]
VCL@FER	23	4	6	552	Varying	RGB	24-bit	MOS	[1,100]
WIQ	7	1	Varying	80	512 × 512	Gray Scale	8-bit	DMOS	[1,100]

Note: PD, perceived distortion; MOS, mean opinion score; DMOS, difference mean opinion score.

Table 2 Performance of the objective quality metrics for different image quality databases.

Database	Crit.	Objective quality measures										
		PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	NQM	IFS	VSI	2-DMC	Proposed
A57	C_P	0.7073	**0.9502**	0.6228	0.4257	0.8507	0.5584	0.8178	0.8995	0.9057	0.9242	**0.9383**
	C_S	0.6189	**0.9355**	0.6223	0.4143	0.8394	0.5398	0.7968	0.8617	0.9030	0.9016	**0.9312**
	C_K	0.4309	**0.8031**	0.4589	0.2854	0.6478	0.3936	0.5890	0.6842	0.7373	0.7490	**0.7883**
TID2008	C_P	0.5734	0.6823	0.8084	0.6413	0.8424	0.6473	0.6135	**0.8810**	**0.8762**	0.8226	0.8235
	C_S	0.5834	0.7015	0.7491	0.6272	0.8526	0.5851	0.6236	**0.8903**	**0.8979**	0.8026	0.8027
	C_K	0.4256	0.5323	0.5861	0.4562	0.6539	0.4255	0.4600	**0.7009**	**0.7123**	0.6122	0.6152
IRCCyN/IVC	C_P	0.6698	0.7306	0.8940	0.7885	0.8920	0.8255	0.7857	**0.9401**	0.9104	0.9294	**0.9481**
	C_S	0.6427	0.7347	0.8876	0.7750	0.8827	0.8196	0.7769	**0.9326**	0.8979	0.9174	**0.9270**
	C_K	0.4784	0.5469	0.7031	0.5887	0.6978	0.6202	0.5796	0.7695	0.7194	**0.7830**	**0.7955**
LIVE	C_P	0.9078	0.9427	**0.9666**	0.9424	0.9560	0.9177	0.8713	**0.9599**	0.9559	0.9509	0.9570
	C_S	0.9124	0.9465	**0.9695**	0.9479	0.9624	0.9136	0.8627	**0.9645**	0.9623	0.9501	0.9536
	C_K	0.7346	0.7989	**0.8477**	0.8006	0.8342	0.7495	0.6799	**0.8417**	0.8328	0.8148	0.8269
CSIQ	C_P	0.8000	0.7685	0.8996	0.7929	0.8937	0.8292	0.7773	**0.9576**	0.9279	0.9360	**0.9545**
	C_S	0.8033	0.7786	0.8935	0.7795	0.8847	0.8236	0.7719	**0.9582**	0.9423	0.9281	**0.9497**
	C_K	0.6022	0.5930	0.7101	0.5921	0.7005	0.6250	0.5745	**0.8165**	0.7857	0.7724	**0.8079**
Toyama	C_P	0.6440	0.8755	**0.9186**	0.7901	0.8919	0.7202	0.8746	0.8710	0.7131	0.9039	**0.9300**
	C_S	0.6153	0.8660	**0.9114**	0.7809	0.8872	0.7181	0.8721	0.8668	0.7015	0.8875	**0.9155**
	C_K	0.4476	0.6833	0.7383	0.5883	0.7044	0.5336	0.6831	0.6772	0.5159	**0.7422**	**0.7736**
VCL@FER	C_P	0.8186	0.4186	0.6363	0.8419	0.8178	0.7902	**0.9396**	0.8394	0.6429	**0.9310**	0.9272
	C_S	0.8053	0.4620	0.6348	0.8419	0.8354	0.7871	**0.9398**	0.8637	0.6005	**0.9290**	0.9241
	C_K	0.6141	0.3243	0.4482	0.6479	0.6501	0.5895	**0.7752**	0.6736	0.4563	**0.7624**	0.7556
WIQ	C_P	0.7609	0.6537	0.7121	0.7881	0.7575	0.6967	0.8161	0.6404	0.8243	**0.8845**	**0.9429**
	C_S	0.6257	0.5945	0.6549	0.7195	0.7146	0.6084	0.7644	0.5795	**0.8043**	0.7811	**0.8289**
	C_K	0.4626	0.4193	0.5025	0.5261	0.5326	0.4360	0.5803	0.4119	0.6170	**0.6442**	**0.7124**

phase. At the end of the experiments, the results obtained for each repetition are used to calculate the overall average of the results. The A57, TID2008, IRCCyN/IVC, LIVE, CSIQ, Toyama, VCL@FER, and WIQ image databases are divided into 3, 5, 10, 6, 6, 7, 5, and 7 partitions, respectively. For each image dataset, we tried to construct partitions, which contain the same number of images, depending on the number of reference images present in the database.

The quality measures produced by the proposed and the baseline techniques are evaluated by measuring the similarity between these measures and the subjective scores. To measure this similarity, three commonly used criteria, namely: (1) Pearson linear correlation coefficient (C_P), (2) Spearman's Rho correlation coefficient (C_S), and (3) Kendall's Tau correlation coefficient (C_K) are used. The objective quality metric resulting in the highest values of these correlation coefficients with the subjective scores is determined to be the best quality assessment metric. Also, to establish a mapping between the objective quality metric results and subjective scores, a five-parameter logistic function[64] is used before the evaluation of correlation-based similarity criteria. The performance of the proposed scheme is compared with
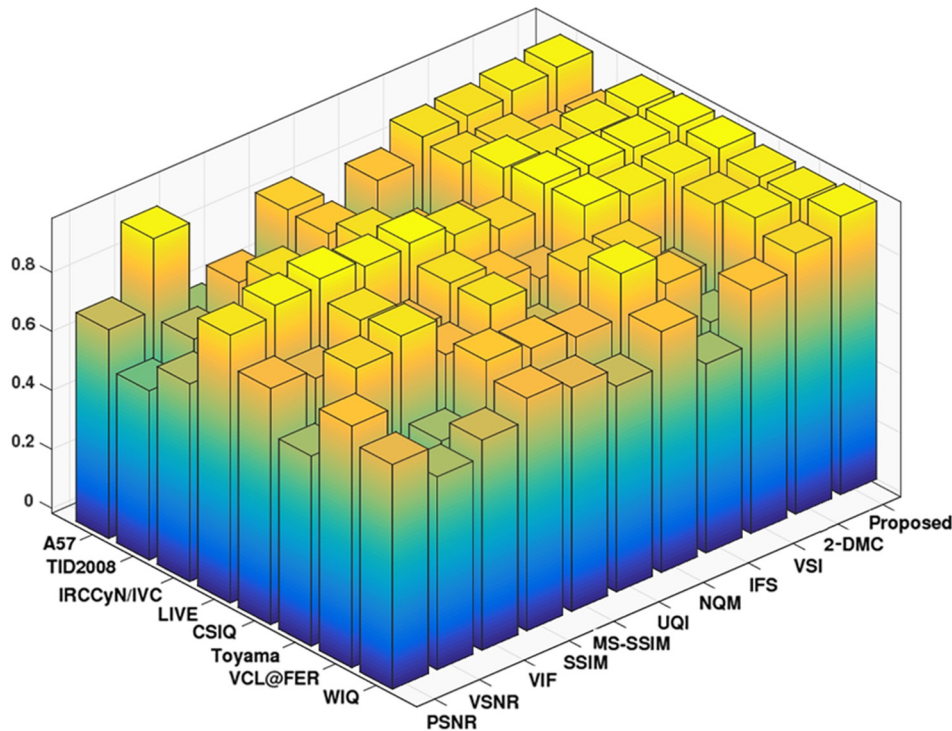
Fig. 5 The illustration of C_p values corresponding to each dataset and objective quality measure.

the baseline techniques consisting of PSNR, VSNR,[18] VIF,[16] SSIM,[13] MS-SSIM,[14] UQI,[15] NQM,[17] independent feature similarity (IFS),[65] visual saliency-based index (VSI),[66] and 2-DMC.[31,32] The results obtained for each objective measure corresponding to each image dataset are presented in Table 2. To increase the readability of the results, similarity measures providing the first and second best performance are written in bold font.

To visualize the correlation values, C_p's corresponding to each objective metric and image dataset are illustrated in Fig. 5.

By looking at the results presented in Table 2 and Fig. 5, one can say that PSNR provides the worst performance on all of the databases. In the literature, it is generally stated that PSNR cannot model HVS adequately,[7] and therefore, the low performance of the PSNR metric is expected. The VSNR outperformed other objective measures in the A57 database but it did not provide promising results on the remaining databases. The VIF metric is another objective measure that showed significant performance variations on different image databases. The VIF, outperforming other objective measures on LIVE and Toyama databases, showed unsatisfactory performance on the remaining databases. Structural similarity-based measures, namely SSIM and MS-SSIM, provided moderate performance on the image databases. However, MS-SSIM outperformed other objective measures (except IFS and VSI) on the TID2008 database. The UQI measure did not show a satisfactory performance on any database. The NQM, especially efficient on additive types of noise, performs quite well on the VCL@FER database. However, it fails to provide acceptable performance on the other datasets. IFS measure outperforms the other objective measures on the TID2008, IVC, LIVE and CSIQ databases. However, the performance of the IFS scheme decreases

dramatically on the remaining databases. The visual saliency-based VSI method provides satisfactory performance on TID2008 and LIVE image databases but the performance of the VSI method does not meet the expectations on the remaining image datasets. Different from the other objective measures, cepstral methods, namely 2-DMC and the proposed 2-D complex mel-cepstrum, achieved acceptable performance on all of the databases. The proposed technique outperformed other baseline techniques especially on IVC, CSIQ, Toyama, and WIQ image databases.

To evaluate the overall performance of each objective metric, the similarity results listed in Table 2 are used to compute the average performance. The averaging is carried out in two ways: (1) ordinary and (2) weighted averaging. In ordinary averaging, all similarity measures corresponding to each image dataset have equal impact on the final result. However, on the weighted averaging, the similarity measures corresponding to each dataset are multiplied with the weights that are proportional to the number of distorted images in the corresponding dataset. The average performance results of each objective measure are presented in Table 3. The standard deviation is another important criterion for the consistency of the objective quality measures. The measure that deviates less than the other measures can be determined to be more consistent. Therefore, the standard deviation of the similarity measures are also presented in Table 3. As another performance criterion, the correlation results of the objective quality measures are sorted to obtain a ranking for each correlation type (C_P, C_S, C_K). The method that has the highest correlation value achieves the first rank. For example, by looking at the results provided in Table 2, VSNR obtains the first rank for C_P measure on A57 image dataset. Similarly, each objective quality metric obtains a ranking between [1,11], for each correlation type on each of the

Table 3 The average, standard deviation, and average ranking of the performance of the objective quality metrics. In order to increase the readability of the results, objective quality measures providing the first and second best performance, are written in bold font.

Averaging scheme	Criteria	Objective quality measures										
		PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	NQM	IFS	VSI	2-DMC	Proposed
Ordinary averaging	Average C_P	0.7352	0.7528	0.8073	0.7514	0.8628	0.7482	0.8120	0.8736	0.8446	**0.9103**	**0.9277**
	Average C_S	0.7009	0.7524	0.7904	0.7358	0.8574	0.7244	0.8010	0.8647	0.8387	**0.8872**	**0.9041**
	Average C_K	0.5245	0.5876	0.6244	0.5607	0.6777	0.5466	0.6152	0.6969	0.6721	**0.7350**	**0.7594**
	Deviation C_P	0.1007	0.1648	0.1256	0.1453	0.0559	0.1076	0.0902	0.0970	0.1041	**0.0382**	**0.0406**
	Deviation C_S	0.1135	0.1569	0.1320	0.1488	0.0657	0.1253	0.0883	0.1147	0.1193	**0.0580**	**0.0527**
	Deviation C_K	0.1052	0.1592	0.1374	0.1395	0.0786	0.1146	0.0890	0.1239	0.1229	**0.0656**	**0.0635**
Weighted averaging	Average C_P	0.7203	0.7246	0.8366	0.7619	0.8722	0.7594	0.7562	**0.9029**	0.8658	0.8915	**0.8991**
	Average C_S	0.7182	0.7387	0.8112	0.7535	0.8762	0.7309	0.7559	**0.9084**	0.8715	0.8784	**0.8855**
	Average C_K	0.5429	0.5731	0.6474	0.5782	0.6962	0.5553	0.5788	**0.7418**	0.7073	0.7132	**0.7261**
	Deviation C_P	0.1018	0.1672	0.1290	0.1457	0.0567	0.1082	0.1060	0.1013	0.1063	**0.0426**	**0.0497**
	Deviation C_S	0.1148	0.1575	0.1336	0.1497	0.0683	0.1255	0.0992	0.1228	0.1237	**0.0587**	**0.0559**
	Deviation C_K	0.1069	0.1598	0.1394	0.1406	0.0807	0.1149	0.0961	0.1318	0.1279	**0.0691**	**0.0717**
Average ranking		9.1667	7.7917	5.5833	7.8750	5.0417	8.5833	7.1250	4.2083	4.8333	**3.4583**	**2.3333**

image datasets. The average ranking is calculated by averaging the individual rankings through correlation types and image datasets. The average ranking of the correlation values are presented in the last row of Table 3.

The average correlation results corresponding to each objective quality measure is displayed in Fig. 6. In Fig. 6, the minimum and maximum correlation values are denoted by upper and lower triangles. The line between the triangles denote the value interval for each correlation coefficient. Also, the length of the interval gives an impression about the deviation values listed in Table 3.

The results presented in Table 3 and Fig. 6 show that the proposed metric obtains promising results in the overall performance evaluations. The proposed metric outperforms baseline techniques when the performance is evaluated using ordinary averaging. When weighted averaging is used in performance evaluation, the IFS metric obtains better results than the proposed regime due to the size of TID2008 dataset in which the IFS outperforms other metrics. However, the IFS metric does not achieve a consistent regime throughout the datasets. The proposed quality scheme not only achieves promising results in average similarity measures, but also provides a consistent regime, i.e., low standard deviation of performance on the overall databases. As expected, the baseline techniques showing large performance deviations on different datasets result in higher deviation values, which can be observed in Table 3. The proposed complex cepstrum-based measure also achieves the highest average ranking, which is another sign of better performance. The overall performance evaluations show that cepstral methods, especially the proposed 2-D complex mel-cepstrum, provides

significant performance increases on certain databases while obtaining acceptable performance on the remaining databases.

To investigate the generalization capability of the proposed quality assessment scheme, an experiment is carried out by using the IVC dataset. Recall that the IVC dataset contains 10 reference images. In the experiment, the algorithm is trained with a single image by changing the reference image at each time. When a single reference image is used in training, the remaining nine images are used for testing. This way, the SVR-based algorithm is trained with very few samples. To quantify the performance of the experiment, average correlation coefficients are calculated. The average correlation values (C_p, C_s, C_k) obtained by the proposed algorithm on the IVC dataset are 0.8665, 0.8609, and 0.6704, respectively. It is obvious that the performance of single-image training scheme is worse than the classical $k - 1$ partition-based training. However, the proposed feature-based IQA framework provides promising results even in the case in which there is a single image in the training set. This experiment revealed the generalization capability of the proposed method, which may enable an efficient representation under circumstances that lack sufficient training data.

To further analyze the performance of objective quality metrics on different distortion types, another set of experiments are carried out. The LIVE image dataset,[60] containing five different distortion types, namely, JPEG2000 compression, JPEG compression (JPEG), additive white noise (WN), Gaussian blur (GB), and fast fading (FF), is used for this single distortion-based evaluation.
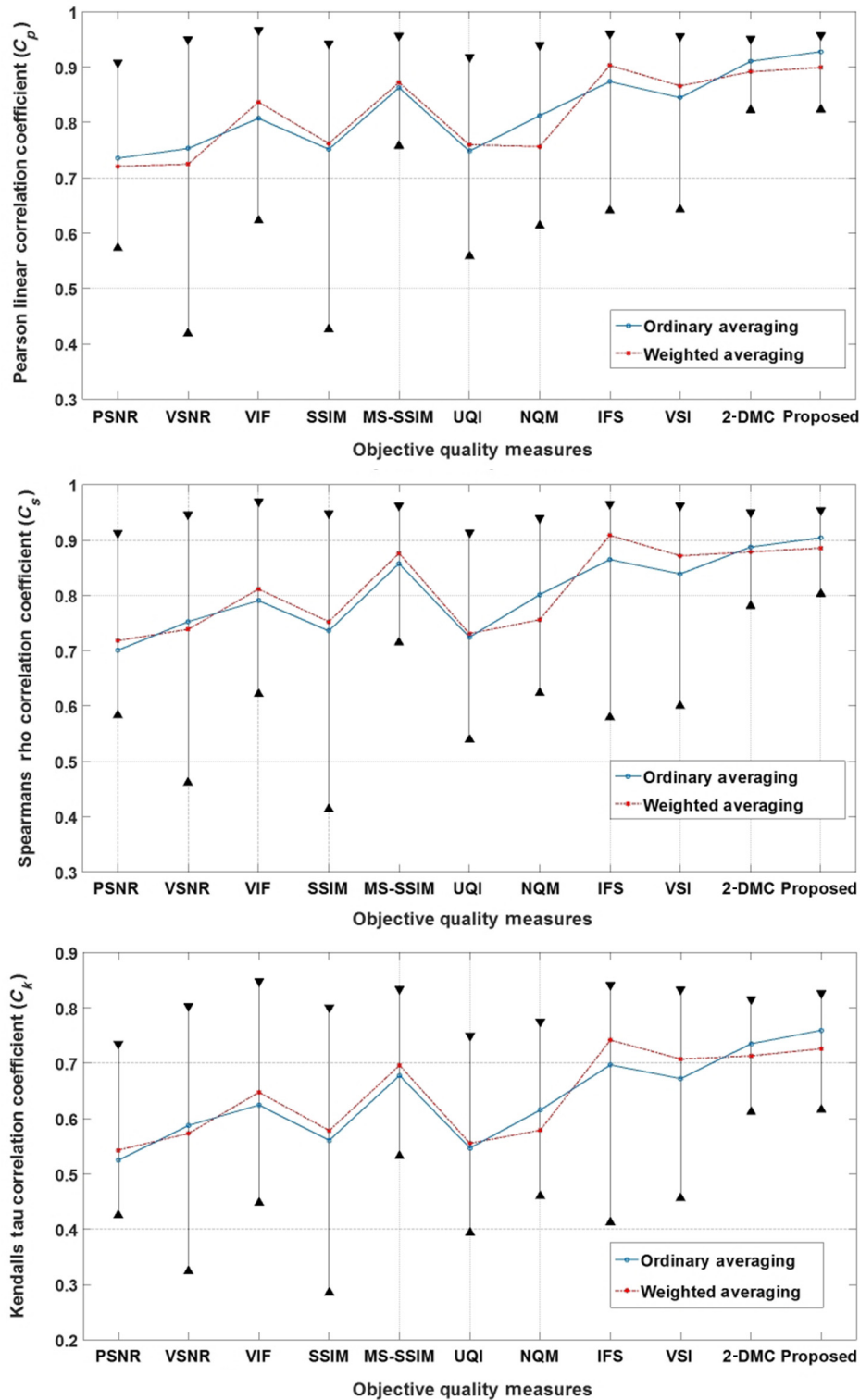
Fig. 6 Average correlation results corresponding to each objective quality measure.

Single distortion-based quality evaluation is performed on the LIVE image dataset, because the objective quality metrics that provide the leading performance on the overall quality evaluation (IFS, VSI, 2-DMC, and proposed) suffer performance losses on the LIVE image dataset. In this way, the effect of each distortion type on the overall performance can be examined. The performance of the VIF metric on different distortion types is also evaluated since the VIF

metric outperforms the other objective metrics on the overall quality evaluation of the LIVE image dataset. The correlation results corresponding to each degradation type are listed in Table 4.

By looking at the results presented in Table 4, one can conclude that the VIF metric coincides well with the subjective scores. In other words, the correlation results corresponding to the VIF metric are higher than the result

Table 4 The average, standard deviation, and average ranking of the performance of the objective quality metrics. In order to increase the readability of the results, objective quality measures providing the first and second best performance, are written in bold font.

Distortion type	Criteria	Objective quality measures				
		VIF	IFS	VSI	2-DMC	Proposed
JPEG2000	C_P	**0.9696**	**0.9694**	0.9604	0.9509	0.9571
	C_S	**0.9597**	0.9564	**0.9605**	0.9314	0.9456
	C_K	**0.8209**	0.8131	**0.8153**	0.7894	0.8124
JPEG	C_P	**0.9846**	**0.9778**	0.9761	0.9620	0.9636
	C_S	**0.9791**	**0.9728**	0.9712	0.9532	0.9585
	C_K	**0.8551**	**0.8474**	0.8449	0.8385	0.8466
WN	C_P	0.9858	**0.9883**	0.9835	**0.9859**	0.9852
	C_S	0.9786	**0.9805**	0.9785	0.9796	**0.9803**
	C_K	0.9231	**0.9324**	0.9265	0.9172	**0.9301**
GB	C_P	0.9728	0.9665	0.9527	**0.9735**	**0.9791**
	C_S	**0.9688**	0.9654	0.9562	0.9675	**0.9792**
	C_K	0.8863	0.8763	0.8447	**0.8914**	**0.9090**
FF	C_P	**0.9650**	0.9515	0.9430	0.9466	**0.9537**
	C_S	**0.9614**	**0.9459**	0.9403	0.9445	0.9454
	C_K	**0.8341**	**0.8123**	0.7865	0.7889	0.8046

obtained by other objective measures. Except for the results obtained for WN and GB, the VIF outperforms other metrics on the remaining distortion types. The results in Table 4 also reveal that the proposed complex mel-cepstrum technique obtained comparable results with the VIF metric. Moreover, it outperforms the baseline metrics in the presence of the GB type of distortion. Although the proposed technique achieves satisfactory performance on the WN, GB, and FF types of degradation, it fails to provide acceptable performance on the compression type of degradation (JPEG2000, JPEG). Taking into consideration the overall quality evaluations presented in Tables 2 and 3, one can conclude that slight performance losses in single distortion results lead to performance reduction on the overall quality evaluation.

The experiments also revealed that the proposed 2-D complex mel-cepstrum outperforms classical 2-DMC features in nearly all of the tests by making use of image-phase information, which contains structural details and high-frequency components. Therefore, our motivation to enhance 2-DMC features with the image phase is validated through the experiments.

6 Conclusion

In this article, a 2-D complex mel-cepstrum feature extraction scheme is proposed for image quality assessment.

The proposed feature extraction framework integrates image-phase information with the classical mel-cepstrum computation to achieve an appropriate balance between FT magnitude and phase for image representation. The complex mel-cepstrum-based features are fed into the SVR-based feature-pooling technique to obtain objective quality scores. The experimental studies demonstrate that the proposed feature extraction technique outperforms baseline techniques on several datasets while achieving an acceptable performance on the remaining datasets. In addition, the proposed feature extraction framework provides the best average performance when the average is computed over all of the datasets using an ordinary averaging scheme. The promising results obtained through large-scale experimentation reveal the effectiveness and representative power of the 2-D complex mel-cepstrum feature extraction scheme. In future work, the proposed 2-D complex mel-cepstrum feature extraction technique is intended to be used in a no-reference image quality assessment framework.

References

1. U. Engelke and H.-J. Zepernick, "Perceptual-based quality metrics for image and video services: a survey," in *3rd EuroNGI Conf. on Next Generation Internet Networks*, pp. 190–197 (2007).
2. Radio-communication assembly, "Recommendation ITU-R BT.500-11 methodology for the subjective assessment of the quality of television pictures," (2002).
3. I. van der Linde and R. M. Doe, "Influence of affective image content on subjective quality assessment," *J. Opt. Soc. Am. A* **29**(9), 1948–1955 (2012).
4. S. Winkler and P. Mohandas, "The evolution of video quality measurement: from PSNR to hybrid metrics," *IEEE Trans. Broadcast.* **54**(3), 660–668 (2008).
5. A. Eskicioglu and P. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.* **43**(12), 2959–2965 (1995).
6. S. Karunasekera and N. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Trans. Image Process.* **4**(6), 713–724 (1995).
7. B. Girod, "Digital images and human vision," in *What's Wrong with Mean-Squared Error?*, pp. 207–220, MIT Press, Cambridge, Massachusetts (1993).
8. R. Dosselmann and X. D. Yang, "A comprehensive assessment of the structural similarity index," *Signal Image Video Process.* **5**(1), 81–91 (2011).
9. S. Wolf, "Measuring the end-to-end performance of digital video systems," *IEEE Trans. Broadcast.* **43**(3), 320–328 (1997).
10. H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
11. H. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Process Lett.* **4**(11), 317–320 (1997).
12. S. Gabarda and G. Cristóbal, "No-reference image quality assessment through the von Mises distribution," *J. Opt. Soc. Am. A* **29**(10), 2058–2066 (2012).
13. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
14. Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. of the Thirty-Seventh Asilomar Conf. on Signals, Systems and Computers*, Vol. 2, pp. 1398–1402 (2003).
15. Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process Lett.* **9**(3), 81–84 (2002).
16. H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
17. N. Damera-Venkata et al., "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.* **9**(4), 636–650 (2000).
18. D. Chandler and S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.* **16**(9), 2284–2298 (2007).
19. M. Sendashonga and F. Lebeau, "Low complexity image quality assessment using frequency domain transforms," in *Proc. of IEEE Int. Conf. on Image Processing*, pp. 385–388 (2006).
20. C.-Y. Wee et al., "Image quality assessment by discrete orthogonal moments," *Pattern Recognit.* **43**(12), 4055–4068 (2010).
21. L. Junfeng et al., "Image quality assessment based on nonsubsampled contourlet transform," in *29th Chinese Control Conf.*, pp. 2665–2670 (2010).

22. Z. Haddad et al., "Image quality assessment based on wave atoms transform," in *Proc. of IEEE Int. Conf. on Image Processing*, pp. 305–308 (2010).
23. A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Trans. Image Process.* **15**(2), 422–429 (2006).
24. M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Networks* **21**(3), 515–519 (2010).
25. A. Shnayderman, A. Gusev, and A. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Trans. Image Process.* **15**(2), 422–429 (2006).
26. M. Miyahara, K. Kotani, and V. Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Trans. Commun.* **46**(9), 1215–1226 (1998).
27. S. Winkler, "Perceptual distortion metric for digital color video," *Proc. SPIE* **3644**, 175 (1999).
28. P. Carrai et al., "Image quality assessment by using neural networks," in *Proc. of IEEE Int. Symp. on Circuits and Systems*, Vol. 5, pp. V–253–V–256 (2002).
29. W. Ding et al., "Image and video quality assessment using neural network and SVM," *Tsinghua Sci. Technol.* **13**(1), 112–116 (2008).
30. W. Lin and M. Narwaria, "Perceptual image quality assessment: recent progress and trends," *Proc. SPIE* **7744**, 774403 (2010).
31. M. Narwaria, W. Lin, and A. E. Cetin, "Scalable image quality assessment with 2d mel-cepstrum and machine learning approach," *Pattern Recognit.* **45**(1), 299–313 (2012).
32. S. Cakir and A. E. Cetin, "Mel-cepstral feature extraction methods for image representation," *Opt. Eng.* **49**(9), 097004 (2010).
33. S. Cakir, "Cepstral methods for image feature extraction," Master's Thesis, Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey (2010).
34. S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust. Speech Signal Process.* **28**(4), 357–366 (1980).
35. H. Yang, D. Huang, and L. Cai, "Perceptually weighted mel-cepstrum analysis of speech based on psychoacoustic model," *IEICE Trans. Inf. Syst.* **E89-D**(12), 2998–3001 (2006).
36. V. Tyagi and C. Wellekens, "On desensitizing the mel-cepstrum to spurious spectral components for robust speech recognition," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 529–532 (2005).
37. T. Kitamura and S. Takei, "Speaker recognition model using two-dimensional mel-cepstrum and predictive neural network," in *Proc. of Fourth Int. Conf. on Spoken Language*, Vol. 3, pp. 1772–1775 (1996).
38. B. U. Toreyin and A. E. Cetin, "Shadow detection using 2D cepstrum," *Proc. SPIE* **7338**, 733809 (2009).
39. J. K. Lee et al., "The complex cepstrum applied to two-dimensional images," *Pattern Recognit.* **26**(10), 1579–1592 (1993).
40. Y. Yeshurun and E. Schwartz, "Cepstral filtering on a columnar image architecture: a fast algorithm for binocular stereo segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(7), 759–767 (1989).
41. A. Oppenheim and J. Lim, "The importance of phase in signals," *Proc. IEEE* **69**(5), 529–541 (1981).
42. A. E. Cetin and R. Ansari, "Convolution-based framework for signal recovery and applications," *J. Opt. Soc. Am. A* **5**(8), 1193–1200 (1988).
43. P. Skurowski and A. Gruca, "Image quality assessment using phase spectrum correlation," in *Proc. of the Int. Conf. on Computer Vision and Graphics: Revised Papers, ICCVG 2008*, pp. 80–89, Springer-Verlag (2009).
44. M. Narwaria et al., "Fourier transform-based scalable image quality measure," *IEEE Trans. Image Process.* **21**(8), 3364–3377 (2012).
45. R. Schafer, "Echo removal by discrete generalized linear filtering," PhD Thesis, MIT (1968).
46. J. Tribolet, "A new phase unwrapping algorithm," *IEEE Trans. Acoust. Speech Signal Process.* **25**(2), 170–177 (1977).
47. L. Moisan, "Periodic plus smooth image decomposition," *J. Math. Imaging Vision* **39**(2), 161–179 (2011).
48. D. C. Ghiglia and L. A. Romero, "Minimum Lp-norm two-dimensional phase unwrapping," *J. Opt. Soc. Am. A* **13**(10), 1999–2013 (1996).
49. K. Steiglitz and B. Dickinson, "Phase unwrapping by factorization," *IEEE Trans. Acoust. Speech Signal Process.* **30**(6), 984–991 (1982).
50. G. Fornaro et al., "Global and local phase-unwrapping techniques: a comparison," *J. Opt. Soc. Am. A* **14**(10), 2702–2708 (1997).
51. M. Costantini, "A novel phase unwrapping method based on network programming," *IEEE Trans. Geosci. Remote Sens.* **36**(3), 813–821 (1998).
52. J. Bioucas-Dias and G. Valadao, "Phase unwrapping via graph cuts," *IEEE Trans. Image Process.* **16**(3), 698–709 (2007).
53. B. Marendic, Y. Yang, and H. Stark, "Phase unwrapping using an extrapolation-projection algorithm," *J. Opt. Soc. Am. A* **23**(8), 1846–1855 (2006).
54. E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imaging* **19**(1), 011006 (2010).
55. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011).
56. B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, Massachusetts (2001).
57. D. M. Chandler and S. S. Hemami, "A57 Dataset," 2007, http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html (March 2016).
58. N. Ponomarenko et al., "TID2008—a database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectron.* **10**, 30–45 (2009).
59. P. Le Callet and F. Autrusseau, "Subjective quality assessment IRCCyN/IVC database" (2005), http://www.irccyn.ec-nantes.fr/ivcdb/ (July 2016).
60. H. R. Sheikh et al., "Image and Video Quality Assessment Research at LIVE," 2004, http://live.ece.utexas.edu/research/quality (July 2016).
61. Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "MICT image quality evaluation database," 2011, http://mict.eng.u-toyama.ac.jp/database_toyama/ (December 2016).
62. A. Zaric et al., "VCL@FER image quality assessment database," *Automatika* **53**(4), 344–354 (2012).
63. U. Engelke et al., "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging," *Signal Process. Image Commun.* **24**(7), 525–547 (2009).
64. H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.* **15**(11), 3440–3451 (2006).
65. H.-W. Chang et al., "Perceptual image quality assessment by independent feature detector," *Neurocomputing* **151**, Part 3, 1142–1152 (2015).
66. L. Zhang, Y. Shen, and H. Li, "VSI: a visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.* **23**(10), 4270–4281 (2014).

Serdar Cakir received his BS degree in electrical and electronics engineering from Osmangazi University and his MS degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2008 and 2010, respectively. He also continues his PhD studies at the Department of Electrical Engineering, Bilkent University. In 2010, he joined Advance Technologies Research Institute operating under the Scientific and Technological Research Council of Turkey, where he is a senior research scientist. His research interests include image/video processing, computer vision, pattern recognition, and infrared imagery.

A. Enis Cetin studied electrical engineering at the Middle East Technical University. After getting his BSc degree, he got his MSE and PhD degrees in systems engineering from the Moore School of Electrical Engineering at the University of Pennsylvania. He has been with Bilkent University, Turkey, since 1989. He is a fellow of IEEE. His research interests include signal and image processing, human–computer interaction using vision and speech, and audiovisual multimedia databases.