

# Exploring V1 by modeling the perceptual quality of images

**Fan Zhang**

Image and Visual Computing Lab, Lenovo Group Ltd.  
Hong Kong SAR, China



**Wenfei Jiang**

Taobao Video, Taobao (China) Software, Co., Ltd  
Shanghai, China



**Florent Autrusseau**

LUNAM Universite, Universite de Nantes IRCCyN UMR  
CNRS 6597, Polytech Nantes, rue Christian Pauc, BP  
50609, 44306 Nantes, France



**Weisi Lin**

Nanyang Technological University  
Singapore



We propose an image quality model based on phase and amplitude differences between a reference and a distorted image. The proposed model is motivated by the fact that polar representations can separate visual information in a more independent and efficient manner than Cartesian representations in the primary visual cortex (V1). We subsequently estimate the model parameters from a large subjective data using maximum-likelihood methods. By comparing the various model hypotheses on the functional form about the phase and amplitude, we find that: 1) discrimination of visual orientation is important for quality assessment and yet a coarse level of such discrimination seems sufficient; 2) a product-based amplitude-phase combination before pooling is effective, suggesting interesting viewpoint about the functional structure of the simple cells and complex cells in V1.

Keywords: V1, image quality assessment, phase, independent subspace analysis, additive log-logistic model

## Introduction

Adaptations are reflected in the structure and function of the human visual system, and many functional traits of the primary cortex are specialized for the environment or niche that we human beings normally occupy. (Olshausen & Field, 1996; Karklin & Lewicki, 2008; Esock et. al., 2009). As we evaluate the quality of a distorted image compared to the reference image, our assessment originates from the difference between the neural responses. The early stages of the human visual system are believed to abstract critical differences and feed the following steps. Although it is an ill-posed problem to infer the underlying physiological process from the overt behavior judgment, a computational model that mimics some properties of visual neurons and predicts image quality accurately, nevertheless, may offer plausible arguments on how the neurons co-act.

The human ability to perceive images emerges in the ventral stream, beginning with simple and complex cells in the primary visual cortex (V1) at the first stage. Simple cells respond selectively to bars and edges at a particular location, of a particular orientation, and with particular bands of spatial frequencies. They act like edge detectors in the standard model (Hubel and Wiesel, 1962). However, since their responses saturate with intense stimuli, they are alternatively argued to act basically

as phase detectors within their contrast saturation ranges (Zetsche & Krieger, 1999). Complex cells differ from simple cells by showing relative invariance to a phase shift of the stimulus (i.e. a small translation perpendicular to the orientation of the stimulus) and responding as the magnitudes of multiple simple cells' responses (Adelson & Bergen, 1985; Hyvriänen & Hoyer, 2000, 2001), and hence are often described as amplitude detectors. The second stage of the ventral system is modeled, for example, by computing products between pairs of V1 responses (both simple and complex) and averaging these products over local regions (Freeman & Simoncelli, 2011).

Images may be decomposed into phase spectra and power spectra by spatial-frequency analysis, e.g. by the Fourier transform. Phase-only reconstruction (Oppenheim & Lim, 1981) and power spectra swapping (Piotrowski & Campbell, 1982) preserve the essential image identity; this is usually interpreted as phase spectra containing more information than the global power spectra, also termed as "phase dominance". By contrast, it is well known that phase-invariant features are important for visual recognition (Sampat, et. al., 2009), since the human visual system should cope with small local deformations of stimuli. Such a paradox raises the question we discuss in this paper: what kind of phase is essential and how is the essential phase fused with the amplitude in the human visual system?

Previous works have investigated the human response to artifact detection (Clarke, 2012) and image classification (Wichmann, 2006; Joubert, et al., 2009) under either phase or amplitude degradations. Recent functional imaging may detect the neural activity directly (Issa, et. al., 2008), but its spatial resolution is not as fine as neural organization and needs to be supplemented by mathematical models, for instance, multivariate analysis (Kamitani & Tong, 2005). In our study, the question of amplitude-phase co-action is explored by a computational model for image quality assessment (IQA). We model subjective image quality as a function of amplitude-phase combinations, test this model on a large collection of subjective image quality databases, and find the essential phase and the effective amplitude-phase combination leading to the most accurate quality prediction.

We use the independent subspace analysis (ISA, Hyvärinen & Hoyer, 2000, 2001) as the functional form of simple and complex cells, and thus the amplitude and the phase are simulated by the magnitude and the phase angle of the ISA response vector, respectively. Similar to independent component analysis (ICA, Jutten & Héroult, 1991), ISA is a method for finding underlying factors or bases from multidimensional statistical data, as reviewed briefly in Appendix A. What distinguishes ISA from ICA is that it searches for statistically independent subspaces consisting of separate groups of bases. The resultant bases and subspaces learnt from natural image data resemble the receptive fields (RFs) of simple and complex cells.

Another merit is that the discriminability on the locations and orientations of the ISA bases can be gradually suppressed by progressively re-configuring the number of ISA subspaces. We find that, for IQA, an important factor is the phase difference that discriminates coarse changes in orientations (e.g. between vertical and horizontal), but the phase difference that further discriminates the finer orientations is less important. Phase difference provides compressive but essential information for IQA.

We further compare the amplitude-only and the phase-only candidate models, as well as various combinations of phase and amplitude, including the sum-based, the product-based, and the polynomial-based candidates. The results show that the product-based model achieves the best accuracy, with a concise form similar to the recent ventral model (Freeman & Simoncelli, 2011). Thus, the amplitude-phase combination must occur prior to the pooling (i.e. counting together the neural responses across visual field) in functional structure. The optimal model parameters confirm that both amplitude and phase are indispensable for IQA. Additionally, the product-

based model may explain human perception using negative film in photographic images.

## Phase and amplitude

Phase and amplitude, for 1D (temporal) signal, indicate the location and the strength of harmonic components at a (temporal) frequency respectively. For 2D signals, phase may represent the location of the harmonic component given both a frequency and an orientation. When stimuli translate perpendicularly to the orientation, the phase varies but the amplitude does not. If the phase is extended to both the location and the orientation of a harmonic component given a frequency, the amplitude remains invariant while the phase captures a small deformation of stimuli, no matter if the stimuli translate or rotate.

In V1, the deformation-invariant characteristic emerges in complex cells. Such functional traits can be simulated by ISA. ISA estimates the receptive fields (RFs) of simple cells from natural images, based on the energy model about complex cells and the sparsity of complex cells' responses. ISA assumes that each complex cell receives inputs from a special group of simple cells, and constitutes a concept of "subspace" (refer to Appendix A). Consequently, the ISA configuration includes the subspace size (i.e., the number of simple cells linked to a complex cell) and the total number of the subspaces (i.e., complex cells). For image patches with 8 by 8 pixels, the ISA is typically configured as ten 4D subspaces (Hyvärinen & Hoyer, 2000, 2001), named as 4D×10 here (The notation  $J D \times T$  means there are  $T$  subspaces and the size of each subspace is  $J$ ). Despite the lack of physiological evidence, ISA supports flexible configurations and can simulate a theoretically optimal model of V1 at various ratios of complex cells to simple cells and various numbers of complex cells. Of course, such simulations only take account of the participating neurons, without consideration of the redundancy trait of the neural system.

Figure 1 (a) shows how the ISA bases evolve as the number of subspaces decreases (where groups of ISA bases are learnt from natural image patches at various numbers of subspaces; the subspaces are delimited by red frames; the number of subspaces decreases from left to right; and the subspace size multiplied with the number of subspaces is approximately equal to the pixel dimension of image patches so as to obtain nearly complete ISA bases). Having fewer subspaces implies that the modeled RF area is larger (i.e. a lower discriminability on locations) with a wider range of orientations that the RFs of each subspace subtend (i.e., a lower discriminability on orientations). For example, the RFs only occupy a small region for Case 2D×32, while the RFs spread all over

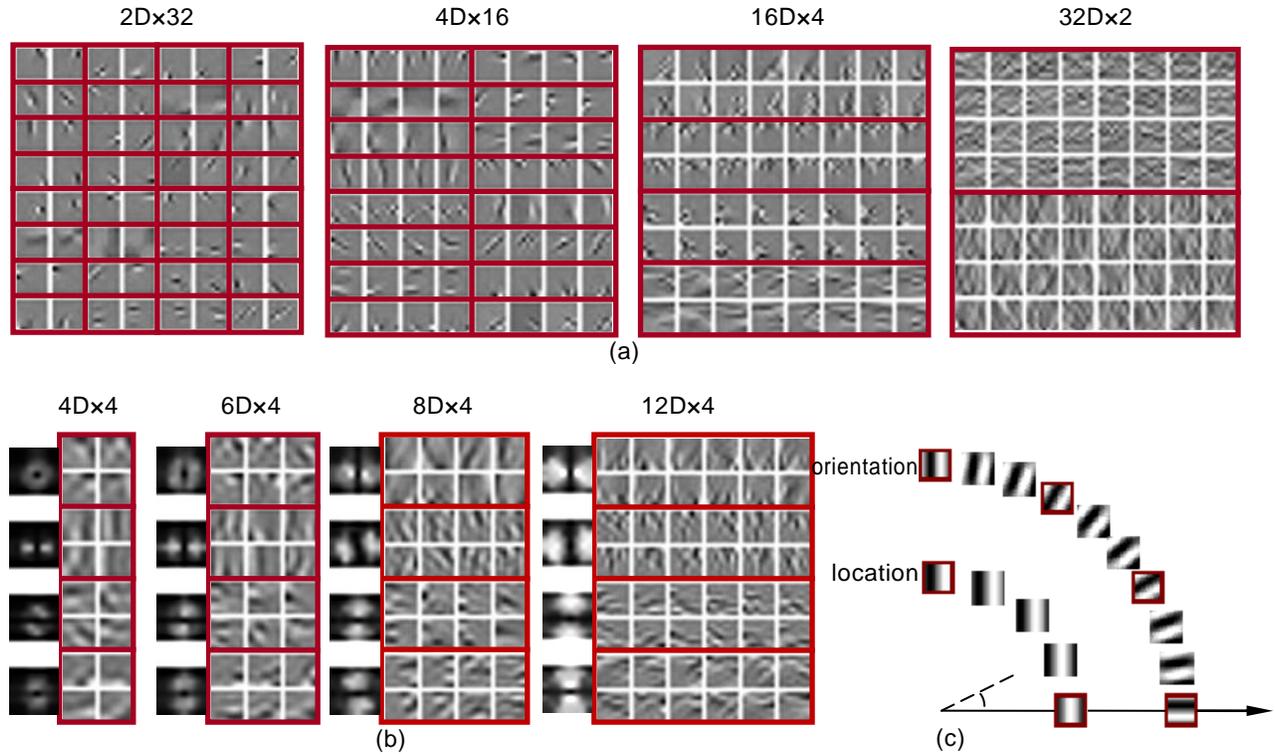


Figure 1 Phase of ISA bases. (a) Complete ISA bases with decreasing numbers of subspaces leading to decreasing discriminability on orientations and locations. (b) Incomplete ISA bases with increasing subspace sizes and four subspaces maintaining consistent discriminability on orientations (with spectral wings) and locations (in spatial area), but gradually accumulating high spatial-frequency components (decentralizing spectrum). (c) Phase discriminating orientations (as the outer sector) and locations (as inner the sector). The bases without red frame are obtained by linearly interpolating on the nearest two bases with red frame. We set the size of ISA bases as 8 by 8 pixels and follow other default settings in FastICA toolbox of Hyvärinen's group.

the patches for Case 32Dx2. Nevertheless, the RFs show diversity in orientation and locations across subspaces for Case 2Dx32, while the RFs are grouped into two coarse orientations (the vertical and horizontal) for Case 32Dx2, in accordance with the orientation anisotropies of human visual cortex (Essock et al., 2009). Moreover, each RF pair shares the same orientation for Case 2Dx32, while the RFs of the same subspace still have diverse orientations and locations for Case 32Dx2. To conclude, it gets more difficult to discriminate the orientation, location or both of ISA bases when the numbers of subspaces are configured decreasingly.

Figure 1(b) shows that the ISA bases maintain consistent discriminability on orientations and locations with varying subspace sizes (to obtain the ISA bases, the dimension of the image patches is reduced to fourfold the subspace size by principal component analysis. Hence, the attained ISA bases are incomplete and support only the principal components). To assist in the analysis, the average Fourier spatial-frequency spectrum of each subspace is also shown on the left. As the subspace size increases, three phenomena can be observed, *i*) HF components emerge and accumulate in the RFs as the spectrum decentralizes;

*ii*) the discriminability on orientations does not change too much since the spectral wings of the RFs keep nearly the same separated orientations; *iii*) the discriminability on locations remains as consistent as the effective regions in the RFs.

## Perceptual quality model

In a  $J$  dimensional ( $J$ -D) space, the polar representation with a scalar amplitude and a  $(J-1)$ -D phase provides an alternative to the Cartesian representation. The polar representation is more likely to independently separate visual information than the Cartesian representation. V1 might provide suitable substrates for amplitude-phase encoding (Zetsche & Krieger, 1999). Indeed, non-Cartesian cells are found in area V4 of macaque monkey (Gallant, et. al., 1993), and sensitive to shape, size, but not location (Gallant, 2000). It is natural for a neural realization to perceive image quality based on the amplitude difference and the phase difference from the distorted image to the original image. The amplitude difference is defined as

$$\rho = |s_r| - |s_d| \quad (1)$$

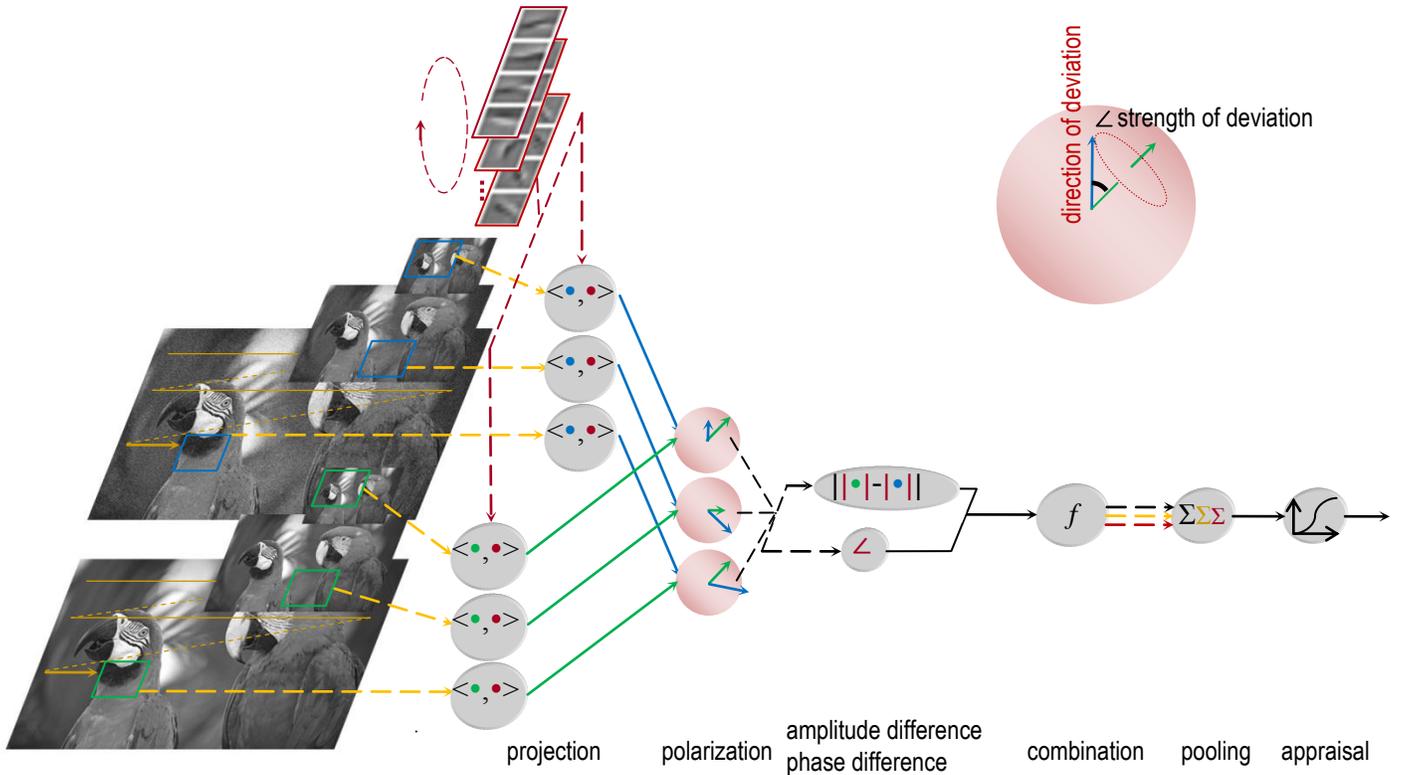


Figure 2 Illustration of the perceptual quality model, where the multi-scale reference image (green) and distorted image (blue) are projected onto the ISA subspaces (red) patch by patch, and then the amplitude and phase difference between patch pairs are combined into a quality score. Dashed arrows represents the iterations over image scales (black), spatial locations (yellow), or subspaces (red). Phase difference records only the strength but not the direction of the phase deviation as illustrated in the top-right subfigure. The images are reproduced from the photo, “Two macaws” by Steve Kelly, with permission from Eastman Kodak Company.

where  $\mathbf{s}_r$  and  $\mathbf{s}_d$  are the pair of ISA response vectors corresponding to the reference and distorted image, and operator  $|\cdot|$  calculates the magnitude of a vector.

Moreover, the phase difference is defined as the angle between the vector pair,

$$\theta = \arccos \frac{\langle \mathbf{s}_r, \mathbf{s}_d \rangle}{|\mathbf{s}_r| \cdot |\mathbf{s}_d|} \quad (2)$$

where  $\langle \cdot, \cdot \rangle$  calculates the inner product. The phase difference records how different the orientations of the vector pair are, and yet tells nothing about along which direction one vector deviates from the other. There is few physiological evidence for a wiring scheme adhering to (2), but a suitable substrate of phase difference might have been in front of our eyes given the existence of meridian-relative anisotropy of the human visual cortex, which refers to that the cortical response is enhanced when the local pattern orientation is coincident with the angular meridian compared with when it is tangential (Sasaki, et. al. 2006; Mannion, et. al., 2010). Phase difference is also associated with computations such as the establishment of optic flow (Geisler 1999) and geometric perspective (Bruce and Tsotsos 2006). Such efficient visual representation seems to be a common principle across species (O’Carroll et. al., 1996).

Within each subspace, the phase difference is equal to the inverse cosine of the classic cosine simi-

ilarity. It is worth noting a special case of the phase difference. When ISA converges to a single 64D subspace, the phase difference is equivalent to the structure comparison in the well-known Structure SIMilarity (SSIM metric; Wang, et. al., 2004, 2008). This is because the structure similarity is defined as the cosine similarity in the pixel space, and ISA, as an orthonormal transform, conserves the cosine similarity (see proof detail in Appendix B).

After V1, the middle stage of the ventral system combines the V1 responses (both simple and complex) (Ungerleider & Haxby, 1994). Let function  $f(\rho, \theta; \alpha, \beta, \gamma)$  represent the combination of the amplitude and the phase differences ( $f$  with parameters  $\{\alpha, \beta, \gamma\}$  will be instantiated later). The functional responses over all ISA subspaces and all image patches are pooled by summing. Considering that simple cells show selectivity of spatial-frequency, we down-sample an image progressively in a ratio of 1/2 to form a pyramid, calculate  $f(\rho, \theta)$  by using the same ISA bases and the identical  $\{\alpha, \beta, \gamma\}$  at each scale, and compute a weighted sum over all scales. The parameters  $\{\alpha, \beta, \gamma\}$  are adaptive for different scales so as to fit the data, and will be estimated by regression (see Appendix C).

After the ventral system, IQA may involve higher level vision stages, such as appraisal process (Scherer, Shorr, & Johnstone, 2001). Finally, subjective opi-

nions tend to “saturate” for very bad or very good image quality, termed as flooring and ceiling effects. It often happens for the subjective IQA protocols (ITU-R BT.500-11, P.910) and most psychological measurements (Aron, et. al., 2006) except, e.g., 2AFC experiment. It may be fitted by a sigmoid or Naka-Rushton-type curve. A log-logistic curve is adopted in video quality assessment standard ITU-T P.1202.2 Mode 2 (Zhang, et. al., 2013). It monotonically maps a distortion value  $f$  to a quality score  $q$

$$q = \frac{1}{1 + a[f(\rho, \theta; \alpha, \beta, \gamma)]^b}, \quad (a > 0) \quad (3)$$

where parameters  $a$  and  $b$  control the curve shape and thereby influence how much the flooring and ceiling effects impact on the distortion  $f$ . The flooring and ceiling effects depend on the context of the test materials, so  $a$  and  $b$  should be associated with each database. In this study, our regression method supports using a single set of  $\{\alpha, \beta, \gamma\}$  for all databases and meanwhile an adaptive set of  $\{a, b\}$  for each database (see details in Appendix C). Using adaptive  $\{a, b\}$  values can compensate the misaligned flooring and ceiling effects across multiple datasets. The functional form of V1 is related to only  $\{\alpha, \beta, \gamma\}$  but not  $\{a, b\}$ . Factually,  $\{a, b\}$  does not change the quality ranking for a database.

The proposed perceptual quality metric is depicted in Figure 2. From the psychophysical stimuli to the visual psychological responses, the metric mainly captures the phase-amplitude response of V1, their combination in the ventral system, subspace pooling, spatial pooling and spatial-frequency pooling, as well as the appraisal process (Scherer, Shorr, & Johnstone, 2001). For simplicity, we assume spatial pooling as a summation function, spatial-frequency pooling as a linear weighted sum, and treating this appreciating as a two-parameter log-logistic; we focus on how the amplitude and phase co-act in V1. The risk might be oversimplification of the unknown stages in the human visual system, but this shortcoming can be balanced and diminished by learning from a large dataset. Most importantly, the proposed metric belongs to the additive linear model (Hastie and Tibshirani, 1990), and thus the model parameters have a converged solution to maximum likelihood (see Appendix C).

## Evaluate image metrics on experimental data

We train the image quality model on existing experimental data. The data are collected from the publicly-available subjectively-rated image quality databases (Winkler, 2012). Each database contains natural images (both original and distorted) and corres-

ponding subjective quality opinions. A total of 11 databases are used, including LIVE, IVC, Toyama, TID2008, A57, WIQ, CSIQ, LAR, BA, FourierSB, and Meerwald. In Table 1, the test materials (including the number, distortion types, and format of images) and the experiment details (such as subjects, viewing setup, and subjective testing methods) are summarized. The image distortions include various kinds of additive noise, Gaussian blur, compression, channel loss, additive watermarks, etc. Figure 3 shows the most typical distortions. The subjective scores (i.e., Mean Opinion Scores, MOSs) represent the ground truth of image quality, and are used to evaluate the prediction accuracy of quality models. At first, we normalize all subjective scores to the range  $[0, 1]$  as shown in Figure 3; a MOS of 1 indicates the best quality while a MOS of 0 represents the worst quality.

The model accuracy is evaluated using two criteria, the likelihood (refer to Equation (7) in Appendix C) and the Spearman rank order correlation coefficient  $\rho_s$ . The likelihood (7) measures the “agreement” of the additive log-logistic model with the subjective quality scores. The coefficient  $\rho_s$  evaluates the ordinal match between the predicted and the subjective quality scores, and thus remains invariant with any a monotonic mapping of the data, including the two-parameter (i.e.,  $a$  and  $b$ ) log-logistic mapping in our model. For both criteria, the higher the value, the better the accuracy;  $\rho_s$  has a range of  $[-1, 1]$ , while the likelihood does not have that constant range and its value also depends on the number and the distribution of the data. Hence, we use  $\rho_s$  to quantify the accuracy of metrics, and use the likelihood to assist in the comparison.

It should be noted that the test conditions were not strictly consistent among the databases, and the subjective data needs to be aligned. Fortunately, the aforementioned additive log-logistic model can align the subjective data automatically and compensate the inconsistency to some extent. More importantly, these 11 databases provide ample data and reduce to the least the bias about image selection, distortion generation, subjects’ preference, etc.

We finally choose the most accurate and concise model from a group of candidates. For each candidate model, we optimize its parameters to fit the subjective data, and then compare the candidates at their best accuracy. The model candidates use various RFs (i.e. the ISA bases under different configurations) and neural co-actions (i.e., the amplitude-phase combination function  $f$ ). The ISA bases are trained offline using FastICA toolbox from Hyvärinen’s group, independent to the quality prediction or the model regression; actually the thirteen natural monochromatic images for ISA bases training are totally different from the images in the ten databases.

	Additive noise	Gaussian Blur	JPEG compression	Intensity shift	Contrast change
MOS	0.547	0.357 ☹	0.425	0.694	0.944 ☺
MSE	492	499	258	235 ☺	819 ☹
Proposed	0.527	0.314 ☹	0.501	0.757 ☺	0.714

Figure 3. Typical distortions of TID2008 databases with subjective/objective quality scores (Ponomarenko, Lukin, & Zelensky, 2008). For each row, the icons ☺ and ☹ respectively annotate the best and the worst score rated by the corresponding method.

We conduct the following two experiments with two questions in mind:

1) what kind of phase information is essential for IQA, in other words, to what extent IQA relies upon the discriminability on orientations and locations?

2) how do the amplitude and phase co-act during IQA, and what is the optimal form of the combination function  $f$ ?

## Experiment 1

To answer Question 1, we conduct Experiment 1 in two steps. First, we find the lowest number of subspaces that can suffice for an accurate IQA. Second, we seek for the narrowest subspaces.

By the first step, the impact of scalable phase information on IQA is evaluated. With a decreasing number of subspaces, the ISA base subtends wider orientations and takes bigger locations, that is, the phase of orientations and locations gradually gets more difficult to be discriminated. Therefore, more phase information is omitted by the quality model.

Specifically, the size and the number of subspaces are configured as follows:  $64D \times 1$ ,  $32D \times 2$ ,  $21D \times 3$ ,  $16D \times 4$ ,  $12D \times 5$ ,  $10D \times 6$ ,  $9D \times 7$ ,  $8D \times 8$ ,  $7D \times 9$ ,  $6D \times 10$ ,  $5D \times 12$ ,  $4D \times 16$ ,  $3D \times 21$ , and  $2D \times 32$ , respectively. Moreover, we adopt 64 ICA bases as Case “ $1D \times 64$ ” for comparison, where  $\theta = 0$  constantly. Here, we keep the ISA bases as complete as possible. Then, we instantiate the combination  $f(\rho, \theta)$  with a product:

$$d_t = |\rho|^\alpha \theta^\beta \quad (4)$$

Using the function (4) is because it works better than other combination candidates in Experiment 2.

The results are presented in Figure 4(a). As the number of subspaces grows (refer to Figure 1(a)), the metric accuracy increases steeply from a single subspace, saturates promptly around four subspaces, peaks persistently for more subspaces, and yet falls finally for ICA. It is thus concluded that four subspaces can suffice for IQA.

It is intuitive, for a reader, to argue that the first step above is not rigorous enough; although the number of subspaces decreases, each subspace expands and thus might compensate and sustain the metric accuracy. Let us move to the next step, to find the narrowest subspaces of necessity.

In this step, given four subspaces, the influence of subspace size on IQA is evaluated. As the subspace expands from a low to a high dimension, the number of ISA bases increases and turn to complete (refer to Figure 1(b)). Specifically, the subspaces size is configured as  $4D \times 4$ ,  $6D \times 4$ ,  $8D \times 4$ ,  $10D \times 4$ ,  $12D \times 4$ ,  $14D \times 4$  respectively, as well as  $16D \times 4$ .

The result is shown in Figure 4(b). As the size of the four subspaces expand, the metric accuracy increases from 4D slightly, peaks around 8D, and holds until the full dimension. Hence, in Experiment 1, the fact that the case with four subspaces is as good as those with more but narrower subspaces should not be attributed to the subspace size, but to the number of subspaces. Actually,  $8D \times 4$  bases support only half of the dimensions but guarantee an efficient IQA. This is not surprising because only the LF (low spatial-frequency) components are taken into account but the HF is ignored. On one hand V1 is not sensitive to the HF at the finest image scale; on the other hand the HF at current image scale is overrepresented by the LF at the finer image scale.

Indeed, the case of product-based  $8D \times 4$  is as accurate as the state-of-the-art metrics like FSIM (Zhang, et. al., 2011) and SIQM (Narwaria, et. al., 2012), and outperforms the CW-SSIM (Sampat, et. al., 2009) and MSE, as can be seen in Table 2. We split TID database into two subsets (Column 4 and 5), TID<sub>1-15</sub> with the first 15 distortion types and TID<sub>16-17</sub> with the last two types, namely “intensity shift” and “contrast change”, which globally adjust the mean and the variance of images respectively as shown in Figure 3. We separate TID<sub>16-17</sub> from the full set because the proposed metric is not good at it and we will discuss that later.

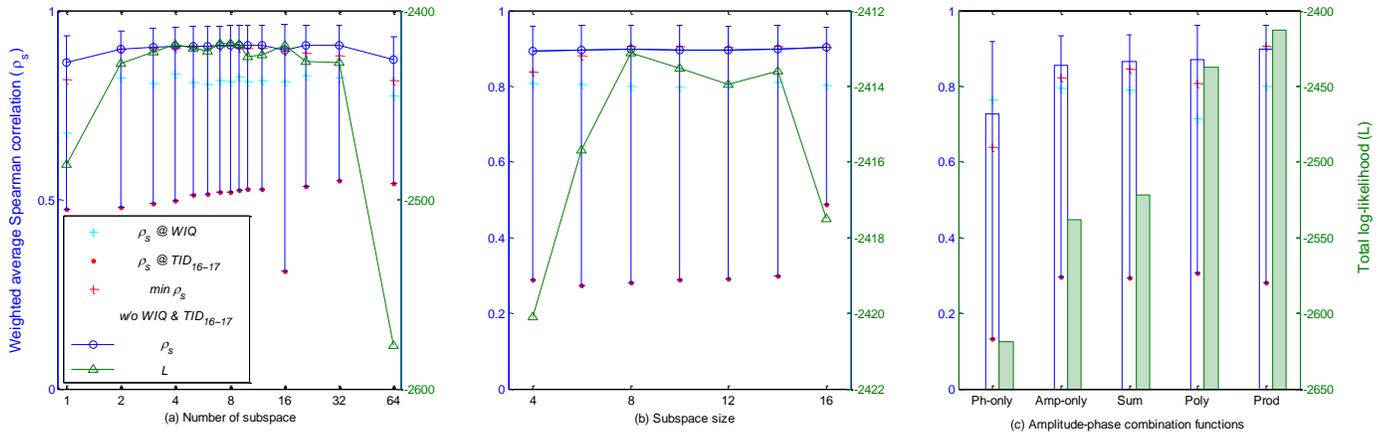


Figure 4 Metric accuracy with various configurations (a) various numbers of ISA subspaces, including the ICA-based case as 1Dx64 on the rightmost, (b) various ISA subspace sizes, and (c) various amplitude-phase combination functions. Green triangles mark the total log-likelihood, blue circles mark the average  $\rho_s$  on all databases, blue vertical bars depict the range of  $\rho_s$  on individual database; cyan crosses depict  $\rho_s$  on WIQ, red points depict  $\rho_s$  on TID<sub>16-17</sub>, while the lowest  $\rho_s$  excluding WIQ and TID<sub>16-17</sub> are depicted by red crosses.

Database	LIVE	IVC	Toyama	TID	A57	WIQ	CSIQ	LAR	BA	FourierSB	Meerwald
# of rated images	779	182	168	1700	54	80	866	120	120	210	120
# of distortion types	5	5	2	17	5	1	6	3	2	6	2
image type	color	Color	Color	color	Gray	gray	color	color	gray	gray	gray
resolution	~768x512	512x512	768x512	512x512	512x512	512x512	512x512	512x512	512x512	512x512	512x512
# of subjects	20~29	15	16	33	7	30	5~7	19	17	7	14
screen	21''CRT	CRT	17''CRT	19''LCD	Papers	17''LCD/CRT	LCD	CRT	24''LCD	'LCD	24''LCD
distance	2~2.5Hs	6Hs	4Hp	varying	4Hp	4~6Hs	80cm	4Hp	6Hs	6Hs	6Hs
rating method	ACR	DSIS	ACR	PC	MSCQS	DSCQS	MSCQS	DSIS	DSIS	DSIS	DSIS
subjective data	DMOS	MOS	Raw	MOS	DMOS	Raw	DMOS	Raw	Raw	Raw	Raw

Hs -- screen height, Hp -- picture height; ACR – absolute category rating, PC – pair comparison method (ITU-T P.910); DSIS – double stimulus impairment scale, DSCQS – double stimulus continuous quality scale, MSCQS – Multiple stimuli, continuous quality scale (ITU-R BT.500-11);

Table 1 Comparison of subjective-rated image quality database, where most information are cited in (Winkler, 2012)

Database	LIVE	IVC	Toyama	TID <sub>1-15</sub>	TID <sub>16-17</sub>	A57	WIQ	CSIQ	LAR	BA	FourierSB	Meerwald
<b>Proposed</b>	0.948	0.910	<b>0.925</b>	<b>0.907</b>	0.281	0.913	0.799	<b>0.961</b>	0.930	0.931	0.906	0.928
MSE	0.856	0.679	0.613	0.532	0.476	0.570	0.817	0.806	0.819	0.934	0.696	0.891
CW-SSIM	0.852	0.621	0.784	0.642	0.482	0.656	0.621	0.577	0.920	0.631	0.055	0.795
SIQM	0.956	0.894	0.915	0.831	0.807	0.894	0.842	0.924	0.892	<b>0.952</b>	0.846	<b>0.940</b>
FSIM	<b>0.963</b>	<b>0.926</b>	0.906	0.882	0.881	<b>0.918</b>	0.806	0.924	<b>0.958</b>	0.934	<b>0.914</b>	0.930

Table 2 Comparison of metrics' performance in terms of  $\rho_s$  over individual database

Completeness of ISA bases Scale <i>l</i> (Fine→Coarse)	Complete bases in Experiment 1			Incomplete bases in Experiment 2		
	Scale 1	Scale 2	Scale 3	Scale 1	Scale 2	Scale 3
$\alpha_l$ : exponent of amplitude error	1.96 ± 0.40	0.62 ± 0.16	1.08 ± 0.51	2.02 ± 0.32	0.53 ± 0.12	1.14 ± 0.07
$\beta_l$ : exponent of phase error	(2.2 ± 7.6) × 10 <sup>-4</sup>	0.45 ± 0.20	0.84 ± 0.22	0.60 ± 0.72	0.45 ± 0.14	0.57 ± 0.09
log $\gamma_l$ : weight of image scale	0	7.14 ± 0.46	6.73 ± 0.81	0	7.11 ± 0.24	5.97 ± 0.85

Table 3 Parameter consistency in terms of mean ± thrice standard deviation

## Experiment 2:

Experiment 2 evaluates the effectiveness of the amplitude-phase combinations and provides an answer to Question 2. We consider the product-based function (4) and the sum-based function:

$$f(\rho, \theta) = |\rho|^\alpha + \gamma\theta^\beta, \quad (\gamma > 0) \quad (5)$$

In (4) and (5),  $\alpha$  and  $\beta$  approximate the nonlinear responses to the amplitude and phase, respectively. When  $\alpha$  or  $\beta$  takes a value of 1, it is degenerated into a linear response, and when  $\alpha$  or  $\beta$  takes a small value near 0, it implies no response and the amplitude or phase term can be omitted. Note that we have assumed subspace pooling, spatial pooling, and spatial-frequency pooling all as summation functions. Accordingly, the product-based function suggests that the amplitude-phase combination is inseparable and thus prior to pooling, while the sum-based function implies that the phase and the amplitude may be decoupled and thus there is no priority between them. Such a difference not only reflects mathematical logic, but also has potentially implications for neural structure.

We take account of another two simpler forms, including the phase-only:

$$f(\rho, \theta) = \theta^\beta$$

and the amplitude-only:

$$f(\rho, \theta) = |\rho|^\alpha$$

which may be regarded as the special sum-based function (5) with  $\gamma \rightarrow \infty$  and  $\gamma \rightarrow 0$ , respectively. We furthermore consider the 3rd order polynomial approximation:

$$f(\rho, \theta) = \sum_{i=0}^3 \sum_{j=0}^{3-i} \gamma_{i,j} \rho^i \theta^j \quad (\gamma_{i,j} > 0)$$

As shown in Figure 4(c), the product-based model outperforms other models. Note that it employs less parameter than the sum-based model and the 3rd order polynomial model.

## Discussion

The experiments above exploit the computational models of IQA. The optimal form of the models suggests that a potentially sufficient polar representation is operative in the human visual system.

Polar representation does not reduce the dimension of signal, but the phase difference does. Note that the phase difference, as a scalar value of an angle, only records the strength of phase deviation but disregards the direction of phase deviation in the high dimensional subspace, as shown by the top-left illustration in Figure 2. Here the direction indicates

how stimuli deform within a range of locations and orientations. The higher the space dimension, the more the phase about the deviation direction is discarded.

Experiment 1 compares various ISA subspaces, which simulate diversified orientation preferences of V1. With more subspaces, the denser becomes the orientation pinwheel, where each simulated RF subtends a narrower range. Phase difference  $\theta$  records how far away the phase deviates within a limited range of orientations and locations, but is blind to the orientation or location from which the phase departs. Therefore with fewer subspaces, more phase information is disregarded by the quality model.

The result of Experiment 1 suggests that overall four subspaces are necessary and sufficient for IQA. This, on one hand, confirms the necessity of the orientation selectivity possessed by V1, and on the other hand implies that a very coarse orientation discriminability suffice for a visual task such as IQA. Nevertheless, the factual orientation preference map is much denser. For example, in the portion of cat V1 that represents the central visual field, the RFs subtend an angle of about  $1.2^\circ$  (Hubel and Wiesel 1962; Tusa, et. al., 1978). This is not surprising, due to the redundancy trait of neural system and the requirement of more challenging tasks like identification.

Experiment 1 further highlights Case 8D×4 for its efficiency (see Table 2 for the comparison with the state-of-the-art metrics), which suggests an interesting structure of V1. Every eight simple cells seem sufficient for the linkage to each complex cell, and the orientation-selective simple cells are grouped into four categories, including the vertical, horizontal, and oblique. Note that each simple cell is a phase detector (Zetzsche & Krieger, 1999) while a complex cell is an amplitude detector. Thus, despite the human visual system detects the total change of the phase detectors' responses, it disregards exactly which phase detectors contribute to the change. In other words, the deformation of stimuli in orientations and locations are detected but they are suppressed in the final analysis. Consequently, only an essential and compressive amount of phase information is factually encoded and forwarded.

It is also interesting to note that in the case of a single subspace, i.e. 64D×1, Experiment 1 links with the classic SSIM metric. Equation (4) is similar to SSIM as proved in Appendix B. In SSIM, the direction in which the phase deviates is totally neglected; in other words, equal values of phase difference across any orientations and any locations are perceived as being equal. The disadvantage of 64D×1 over the cases with more subspaces implies that SSIM could be refined with additional consideration of orientations.

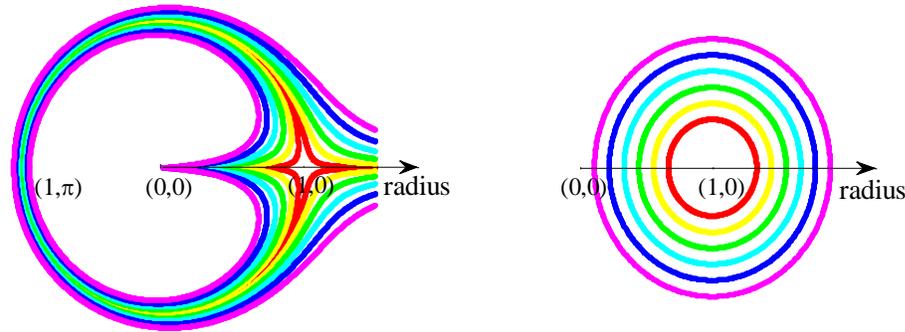


Figure 5 Iso-distance maps of  $(1, 0)$  under distance metric (a)  $d = |\rho|^{0.6}\theta^{0.5}$  and (b) MSE

Note again that we use a single set of parameters  $\{\alpha_l, \beta_l, \gamma_l | l = 1, 2, 3\}$  for all databases and adaptive parameters  $\{a_m, b_m | m = 1, 2, \dots, 10\}$  for the  $m$ -th database. Due to the redundancy between  $\{\gamma\}$  and  $a$ , we set  $\gamma_1 = 1$  constantly. That is, although we introduce 29 parameters, the degree of freedom for the proposed V1 model is only 8.

As listed in Table 3, most of the  $\{\alpha, \beta, \gamma\}$  are consistent under various configurations of ISA subspaces, except  $\beta_1$  for the phase at the finest scale.  $\beta_1$  can be omitted (i.e. set as 0) especially when complete ISA bases are configured, and hence the phase of the finest HF components is insignificant for IQA. This is in accordance with the view that amplitude is more important than phase at fine image scales (Field and Chandler, 2012).

Experiment 2 compares various amplitude-phase combinations, which simulate different co-actions between simple and complex cells. Among the candidates including summation, product, the 3rd order polynomial, etc., the product performs best. This result agrees with the product-based assumption on the ventral system (Freeman & Simoncelli, 2011). The phase-only and the amplitude-only models, as special cases of the sum-based model, are also compared, but their inferior performance confirms that both the phase and amplitude are indispensable for IQA.

The product-based combination suggests a viewpoint of perceptual distortion that is quite different from traditional quality metrics. We can explain this via the iso-distance map of metric Equation (4). Given a reference point  $o$  and a distance  $d$ , the iso-distance curve consists of all the points that are located at distance  $d$  from  $o$ . Let us consider a 2D polar coordinate system for simplification. Given a reference point with radius of 1 and phase angle of 0, noted by  $(1, 0)$ , its iso-distance map under metric Equation (4) is shown in Figure 5(a). A point has a distance of zero from the reference, as long as either its phase or its amplitude remains the same as the reference. This differs from the iso-distance map un-

der MSE (mean squared error) as shown in Figure 5(b), where a point moves farther away from the reference point unless both its amplitude and phase are equal to the reference's. Such a difference is because the metric of Eq. (4) employs a product to combine the amplitude and phase error while MSE approximates to a sum of two items related to the amplitude and phase error respectively (Hisao & Millane, 2004).

It is intuitive that the phase-invariant distortions impair images mildly, as human eyes often show adaptation and tolerance to the contrast adjustment or to high dynamic range mapping for images. Besides, it is partly reasonable, although counterintuitive, that the amplitude-invariant distortions only marginally degrade images. Amplitude-invariant distortions do not occur commonly, yet this is relevant in negative film photography. Generally, negative images (i.e. amplitude-invariant and inverse-phase images) are still recognizable. Of course this does not mean that a negative image deserves a rating of perfect quality, but it seems more unreasonable to evaluate a negative image worse than a random noise image like most existing metrics do.

Compared to summation, a product of terms is more common for combining two incommensurate quantities. Considering that the quality model pools the results of the combinations by summation, combining by product suggests that the combination and the pooling are not commutative. Otherwise, if combining by sum was preferred, the combination and the pooling would be commutative and combination followed by pooling would be equivalent to combination after pooling.

The proposed models appear to perform inconsistently on the 11 databases since there are large (blue vertical) error bars of  $\rho_s$  in Figure 4. This is mainly because of their low accuracy on the TID<sub>16-17</sub> and the WIQ database. If excluding them, the worst accuracies (marked by red crosses) are not significantly lower than the best ones. The proposed metric inaccurately measures “intensity shift”, because the

image mean is overlooked by using the ISA bases which are obtained from the whitened data with zero mean. “Contrast change” is not measured appropriately here, because contrast change (i.e., amplitude difference) is simply regarded as distortions no matter if the contrast is enhanced or degraded. Most existing metrics fail to handle WIQ (as shown in Table 2), because the image distortion in WIQ, termed as wireless channel distortion, is often uneven and localized. Hence, the simulated RF of ISA bases, the distortion factor of absolute amplitude difference, and the pooling strategy of summation function are probably too oversimplified, since subjective assessment for such distortions may involve a more complex process in high-level vision.

Although the proposed computational model is an advance in that it can mimic the nonlinear properties of neurons, it is still oversimplified; it focuses on V1 and ignores mechanisms beyond V1. Despite a good match between the model predictions and behavior judgment, our model cannot be regarded as the uniquely correct computational approximation of the underlying physiological process. Using reasonable set of localized filters, we can devise more image quality metrics, e.g., a PCA (principal component analysis)-based metric (Zhang et. al., 2011). The essential issue is to carefully design the weights for each filter. Using over-complete filters is equivalent to weight the filters, so could promise a good result too. However, a key point here is that the ISA bases at identical image scales are assigned with *equal* weights (i.e. the same  $\alpha$ ,  $\beta$  and  $\gamma$ ), since there is no evidence that any set of V1 neurons have priority or account for the majority.

## Conclusions

We have studied a simple and accurate image quality metric, where all parameters can be trained by a converged algorithm and where no parameter is set empirically. The metric simulates the RFs of V1 using ISA bases, simple cells using phase detectors, complex cells using amplitude detectors, their co-actions using the product, and the later stages using a log-logistic mapping. We make the least possible priori assumptions for this metric.

In the comparative study, a metric based on various ISA bases and amplitude-phase combinations stands out, and suggests the following views about V1.

i) Both phase and amplitude are indispensable for IQA, and thus both simple cells and complex cells contribute to IQA. Besides the amplitude detection, the phase difference provides another potential way of information reduction.

ii) Not all the phase information is helpful for IQA; only the phase which discriminates coarse orientations is essential.

iii) Product of phase and amplitude can capture these combinations and thus the co-actions between simple and complex cells, rather than summation or other nonlinear operators, and thereby the human visual system tolerates the amplitude-invariant and phase-invariant distortions.

iv) The amplitude-phase combination occurs prior to the pooling, which implies the linkages among simple and complex cells precede the aggregation of the neurons that represents various locations of the visual field.

## Appendix A: independent subspace analysis

ISA has two layers; the first-layer unit simulates the RF of simple cells, while each of the second-layer units pools over a small neighborhood of adjacent first-layer units to mimic complex cells. To be precise, given the input image patch  $\mathbf{x} \in \mathbb{R}^I$ , the activation of each first-layer unit is:

$$s_j^2 = \left( \sum_{i=1}^I \mathbf{W}_{ji} x_i \right)^2$$

and the activation of each second-layer unit is

$$S_t = \sqrt{\sum_{j=1}^J s_{(t-1) \cdot J + j}^2} = \sqrt{\sum_{j=1}^J \left( \sum_{i=1}^I \mathbf{W}_{(t-1) \cdot J + j, i} x_i \right)^2}$$

$\mathbf{W} \in \mathbb{R}^{(J \cdot T) \times I}$  is the weight matrix of the first layer and also the ISA transform matrix;  $I$ ,  $J$  and  $T$  are the input dimension (number of pixels in a patch), subspace size (number of the first-layer units to be pooled for a second-layer unit), and number of the subspaces (number of the second-layer units), respectively. The row vectors of  $\mathbf{W}$ , as ISA bases, support a linear-transformed space and are grouped into  $T$   $J$ -D subspaces. ISA trains  $\mathbf{W}$  via sparse representations in the second-layer, by equivalently solving:

$$\min_{\mathbf{W}} \sum_{k=1}^K \sum_{t=1}^T S_t(\mathbf{x}^k; \mathbf{W}), \quad \text{s. t. } \mathbf{W}\mathbf{W}^T = \mathbf{I}$$

where the training set  $\{\mathbf{x}^k\}_{k=1}^K$  are whitened to have zero mean and identity covariance, and also have the dimension reduced to  $J \cdot T$  by PCA. The orthonormal constraint guarantees that transform  $\mathbf{W}$  is invertible.

## Appendix B: proof of relation to SSIM

The Structure SIMilarity (SSIM) (Wang, et. al., 2004, 2008) uses a product-based combination of three comparisons, as

$SSIM(\mathbf{x}_r, \mathbf{x}_d) = [s(\mathbf{x}_r, \mathbf{x}_d)]^\beta [c(\mathbf{x}_r, \mathbf{x}_d)]^\alpha [l(\mathbf{x}_r, \mathbf{x}_d)]^\gamma$   
where the structure comparison function is

$$s(\mathbf{x}_r, \mathbf{x}_d) = \frac{\langle \mathbf{x}_r - \mu_r, \mathbf{x}_d - \mu_d \rangle}{|\mathbf{x}_r - \mu_r| |\mathbf{x}_d - \mu_d|}$$

the contrast comparison function is

$$c(\mathbf{x}_r, \mathbf{x}_d) = \frac{2|\mathbf{x}_r - \mu_r| |\mathbf{x}_d - \mu_d|}{|\mathbf{x}_r - \mu_r|^2 + |\mathbf{x}_d - \mu_d|^2}$$

the luminance comparison function is

$$l(\mathbf{x}_r, \mathbf{x}_d) = \frac{\mu_r \mu_d}{\mu_r^2 + \mu_d^2}$$

and column vectors  $\mathbf{x}_r$  and  $\mathbf{x}_d$  consist of pixels in the reference and distorted  $8 \times 8$  patch from the same location, with the means of  $\mu_r$  and  $\mu_d$ , respectively.

If and only if the ISA converges to a single subspace, we have

$$\mathbf{s} = \mathbf{W} \cdot \mathbf{x} = \mathbf{W} \cdot (\mathbf{x} - \mu)$$

for both  $\mathbf{s}_r \sim \mathbf{x}_r$  and  $\mathbf{s}_d \sim \mathbf{x}_d$ , where the ISA transform  $\mathbf{W}$  is trained on zero-mean data and remain orthogonal. Hence, we have

$$\begin{aligned} \langle \mathbf{s}_r, \mathbf{s}_d \rangle &= (\mathbf{x}_r - \mu_r)^T \mathbf{W}^T \mathbf{W} (\mathbf{x}_d - \mu_d) \\ &= (\mathbf{x}_r - \mu_r)^T (\mathbf{x}_d - \mu_d) \\ &= \langle \mathbf{x}_r - \mu_r, \mathbf{x}_d - \mu_d \rangle \end{aligned}$$

and  $|\mathbf{s}| = |\mathbf{x} - \mu|$ . Then, obvious is the equivalence between the phase difference and the structure comparison of SSIM:

$$\theta = \arccos[s(\mathbf{x}_r, \mathbf{x}_d)]$$

and the relation between the amplitude difference and the contrast comparison:

$$|\rho| = \sqrt{[1 - c(\mathbf{x}_r, \mathbf{x}_d)] \cdot [|\mathbf{x}_r - \mu_r|^2 + |\mathbf{x}_d - \mu_d|^2]}$$

Consequently, Equation (4) can be rewritten as:

$$d = C \cdot [1 - c(\mathbf{x}_r, \mathbf{x}_d)]^{\frac{\alpha}{2}} [\arccos[s(\mathbf{x}_r, \mathbf{x}_d)]]^\beta$$

where  $C = [|\mathbf{x}_r - \mu_r|^2 + |\mathbf{x}_d - \mu_d|^2]^{\frac{\alpha}{2}}$

That is, Equation (4) combines two essential components of SSIM in a slightly different manner.

## Appendix C: additive log-logistic model

The function of the perceptual quality  $q$  with respect to the distortions  $d$  is defined as:

$$q = \frac{1}{1 + a(\sum d)^b}, \quad (a > 0) \quad (6)$$

where parameters  $a$  and  $b$  control the shape of the log-logistic curve. We call it the additive log-logistic model, since it has a link form as:

$$\left(\frac{1-q}{aq}\right)^{1/b} = \sum d$$

that is, the distortions sum up and yield the monotonically transformed quality. Here, we use three levels of summations as:

$$\sum d = \sum_l \left( \frac{\gamma_l}{K_l} \sum_{k=1}^{K_l} \sum_{t=1}^T d_{l,k,t} \right)$$

The outermost sum on the right-hand side is over the image scales (indexed by  $l$  and linearly weighted by parameter  $\gamma_l$  at each scale), the middle sum is over the totally  $K_l$  locations of patches all around image and normalized by  $K_l$ , and the innermost sum is over the totally  $T$  subspaces; they approximate the spatial-frequency pooling, the spatial pooling, and the subspace pooling, respectively. The local distortion  $d$  combines the phase difference and amplitude difference, for instance but not limited to

$$d_{l,k,t} = \theta_{l,k,t}^{\beta_l} |\rho_{l,k,t}|^{\alpha_l} \quad (3)$$

The goodness-of-fit between the predicted  $\{q\}$  and the subjectively rated  $\{\hat{q}\}$  is evaluated by the likelihood of  $\{q\}$  given  $\{\hat{q}\}$ . We assume binomial distribution as the priori distribution of  $\{\hat{q}\}$ , for the non-Gaussianity of opinion scores as well as the computational simplicity. Given totally  $M$  independent databases where the  $m$ -th database contain totally  $N_m$  samples, the total log-likelihood is:

$$\mathcal{L} = \sum_{m=1}^M \sum_{n=1}^{N_m} \left[ \hat{q}_{m,n} \log \left( \frac{q_{m,n}}{1 - q_{m,n}} \right) + \log(1 - q_{m,n}) \right] \quad (7)$$

The metric accuracy is evaluated by the maximal total log-likelihood, with the estimated parameter sets  $\{\alpha, \beta, \gamma\}$  and  $\{a, b\}$ . Note that  $\{\alpha, \beta, \gamma\}$  remain constant for different databases, while  $(a_m, b_m)$  is adaptive to the  $m$ -th database but does not affect the ordinal prediction. By the gradient-descent method, the parameter estimation based on maximum likelihood has a solution below.

$$\frac{\partial \mathcal{L}}{\partial \alpha_l} = \sum_{m,n} \left[ \frac{(q_{m,n} - \hat{q}_{m,n}) b_m}{\sum d} \sum_{k,t} \frac{\gamma_l d_{n,l,k,t} \log |\rho_{n,l,k,t}|}{K_l} \right]$$

$$\frac{\partial \mathcal{L}}{\partial \beta_l} = \sum_{m,n} \left[ \frac{(q_{m,n} - \hat{q}_{m,n}) b_m}{\sum d} \sum_{k,t} \frac{\gamma_l d_{n,l,k,t} \log \theta_{n,l,k,t}}{K_l} \right]$$

$$\frac{\partial \mathcal{L}}{\partial \log \gamma_l} = \sum_{m,n} \frac{(q_{m,n} - \hat{q}_{m,n}) b_m \sum_{k,t} d_{n,l,k,t}}{\sum d}$$

$$\frac{\partial \mathcal{L}}{\partial \log a_m} = \sum_{n=1}^{N_m} (q_{m,n} - \hat{q}_{m,n})$$

$$\frac{\partial \mathcal{L}}{\partial b_m} = \sum_{n=1}^{N_m} [(q_{m,n} - \hat{q}_{m,n}) \log(\sum d)]$$

Note that the logarithmic  $\gamma_l$  and  $a_m$  are solved, so as to guarantee  $\gamma_l$  and  $a_m$  always positive.

## Acknowledgments

The authors thank Imants Svalbe, Jason Wang, Longin Jan Latecki, and Zhibo Chen for insightful discussions.

Commercial relationships: the Technicolor Company has filed a patent about image quality metrics on which F. Zhang and W. Jiang are listed as inventors. No other author has a proprietary interest in any material or method mentioned.

Corresponding author: Fan Zhang.

Email: [fan.zhang@ieee.org](mailto:fan.zhang@ieee.org)

Address: 21/F, Lincoln House, Taikoo Place, 979 King's Road, Quarry Bay, Hong Kong.

## References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2, 284 – 299.
- Aron, A., Aron, E. N., & Coups, E. (2006). *Statistics for psychology*. Prentice Hall
- Burce, N. D., & Tsotsos J. K. (2006). A statistical basis for visual field anisotropies. *Neurocomputing* 69, 1301 – 1304.
- Clarke, A. D. F., Green, P. R., & Chandler, M. J. (2012). The effects of display time and eccentricity on the detection of amplitude and phase degradations in textured stimuli. *J. Vis.* 12, 1 – 11.
- Essock, E. A., Haun, A. M., & Kim, Y. J. (2009). An anisotropy of orientation-tuned suppression that matches the anisotropy of typical natural scenes. *J. Vis.* 9, 1 – 15.
- Field, D. J., & Chandler, D. M. (2012). Method for estimating the relative contribution of phase and power spectra to the total information in natural-scene patches. *J. Opt. Soc. Am. A* 29, 55 – 67.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience* 14, 1195 – 1201.
- Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in Macaque visual cortex. *Science* 259, 100 – 103
- Gallant, J. L. (2000). The neural representation of shape. In Karen K. De Valois (Ed.), *Seeing* (2<sup>nd</sup> Ed. 311 – 324), San Diego: Academic Press.
- Geisler, W. S. Motion streaks provide a spatial code for motion direction. *Nature* 400, 65 – 69.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models*, Taylor and Francis.
- Hsiao, W. H., & Millane, R. P. (2004). Effects of spectral amplitude and phase errors on image reconstruction. *Proc. SPIE* 5562, 27 – 37.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154.
- Hyvriäinen, A., & Hoyer, P. (2000). Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspace. *Neural Computation* 12, 1705 – 1720.
- Hyvriäinen, A., & Hoyer, P. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vis. Res.* 41, 2413 – 2423.
- Issa, N. P., Rosenberg, A., & Husson, T. R. (2008). Models and measurement of functional maps in V1. *J. Neurophysiol* 99, 2745 – 2754.
- Joubert, O. R., Rousset, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene context with equalized amplitude spectrum and increasing phase noise. *J. Vis.* 9(1): 2, 1 – 16.
- Jutten, C. & Héroult, J. (1991). Blind separation of source, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Process.* 24: 1 – 10.
- Kamitani, Y. & Tong, F. (2005) Decoding the visual and subjective contents of the human brain. *Nat. Neurosci* 8, 679 – 685.
- Karklin, Y., & Lewicki, M. S. (2008). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 457, 83 – 86.
- Mannion, D. J., McDonald, J. S., & Clifford, C. W. G. (2010) Orientation anisotropies in human visual cortex. *J. Neurophysiol* 103, 3465 – 3471.
- Narwaria, M., Lin, W., McLoughlin, I., Emmanue, S., & Chia, L. T. (2012) Fourier transform based scalable image quality measure. *IEEE Trans. Image Process.* 21, 3364 – 3377.
- O'Carroll, D. C., Bidwell N. J., Laughlin, S. B., & Warrant E. J. (1996) Insect motion detectors matched to visual ecology. *Nature* 382, 63 – 66.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607 – 609.

- Oppenheim, A. V., & Lim, J. S. (1981) The importance of phase in signals. *Proc. IEEE* 69, 529 – 541.
- Piotrowski, L. N., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception* 11, 337 – 346
- Ponomarenko, N., Lukin, V., Zelensky, A., et. al., (2009). TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics, *Advances of Modern Radioelectronics* 10, 30-45.
- Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., & Markey, M. (2009) Complex wavelet structural similarity: a new image similarity index. *IEEE Trans. Image Process.* 18, 2385 – 2401
- Sasaki, Y., Rajimehr, R., Kim, B. W., et. al., (2006). The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron* 51, 661 – 670.
- Scherer, K. R., Shorr, A., & Johnstone, T., (2001). Appraisal processes in emotion: theory, methods, research. Canary, NC: Oxford University Press.
- Tusa, R. J., & Palmer, L. A., & Rosenquist, A. C. (1978). The retinotopic organization of area 17 (striate cortex) in the cat. *J. Comp. Neurol.* 177, 213–235
- Ungerleider, L. G., & Haxby J.V. (1994). ‘What’ and ‘where’ in the human brain. *Curr. Opin. Neurobiol.* 4, 157 – 165
- Wang, Z., Bovik, A. C., Sheikh, H.R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600 – 612.
- Wang, Z. & Simoncelli, E. P. (2008). Maximum differentiation (MAD) competition: a methodology for comparing computational models of perceptual quantities. *J. Vision* 8, 1 – 13.
- Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vis. Res.* 46, 1520 – 1529.
- Winkler, S. (2012) Analysis of public image and video databases for quality assessment. *IEEE J. Sel. Topics Signal Process.* 6, 616 – 625.
- Zetsche, C., & Krieger, G. (1999). The atoms of vision: Cartesian or polar? *J. Opt. Soc. Am. A* 16, 1554 – 1565.
- Zhang, F., Liu, W., Lin, W., & Ngan, K. N. (2011). Spread spectrum image watermarking based on perceptual quality metric. *IEEE Trans. Image Process.* 20, 3207 – 3218.
- Zhang, F., Lin, W., Chen, Z., & Ngan, K. N. (2013). Additive log-logistic model for modeling networked video quality. *IEEE Trans. Image Process.* 22, 1536 – 1547.
- Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011) FSIM: a feature similarity index for image quality assessment. *IEEE Trans. Image Process.* 20, 2378 – 2386