The Importance of Phase to Texture Similarity

Xinghui Dong [1*], Ying Gao [2], Junyu Dong[2], Mike J. Chantler[3]

[1] Centre for Imaging Sciences, The University of Manchester, Manchester, M13 9PT, UK
[2] Department of Computer Science, Ocean University of China, Qingdao, China
[3] Texture Lab, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK
[*] xinghui.dong@manchester.ac.uk

Abstract

Although the importance of the Fourier phase to image perception has been addressed, it is unknown whether this is the case for texture similarity or not. Based on psychophysical experiments, we first show that the phase data is more important to human visual perception of texture similarity than the magnitude data. We further examine the ability of a total of 51 computational feature sets on exploiting the phase data for texture similarity estimation. However, it is found that for these feature sets the magnitude data is more important than phase. Since it has been shown in early work that there is inconsistency between the similarity data derived from human observers and the 51 feature sets, we attribute this outcome (magnitude/phase importance) to the difference between the manners in which humans and the feature sets exploit the phase data. Therefore, we are motivated to enable the 51 feature sets to exploit the phase data for effective estimation of texture similarity. This is achieved by fusing the features extracted from the original and phase-only images. It is shown that this type of fused feature sets yield better results than those derived using the 51 original feature sets. In particular, we show that this finding can also be propagated to convolutional neural network features. We believe that the improved results should be attributed to the importance of phase to texture similarity.

1. Introduction

Fourier analysis has been widely used in the studies of perception of image appearance [10], [14], [18], [27], [28], [38]. Oppenheim and Lim [28] showed that the Fourier phase (or phase for simplicity) spectrum is more important to perception of natural images than the magnitude data. Specifically, an image containing the aperiodic structure can still be identified when its power spectrum is replaced by a single-valued matrix; while this is not the case when its phase spectrum is scrambled (see Figure 1). Existing studies [29], [38], [41] also examined the effect of the phase and/or magnitude spectra on the appearance of images. To the authors' knowledge, however, none of these studies investigate the importance of the phase spectrum to texture similarity [7], [9]. Texture similarity estimation is key to texture or material recognition. The goal of texture similarity studies is to develop certain methods that can predict the degree to which pairs of textures manifest similar to what humans perceive.

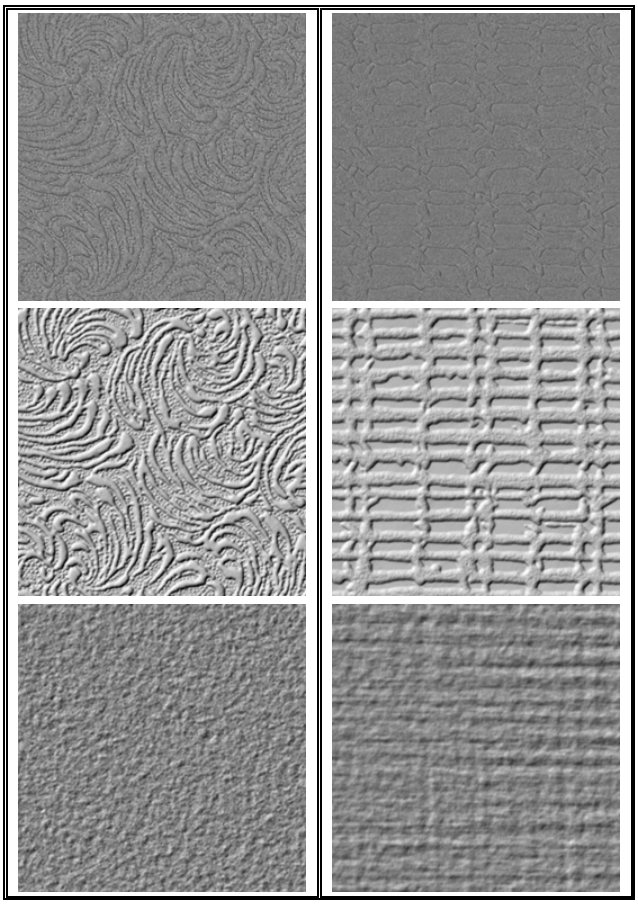

Figure 1: Each of the two columns presents three images derived from the same texture (central quarters are shown). Each of the three rows displays the phase-only, original, and magnitude-only images in turn.

Dong et al. [7], [9] examined the ability of a total of 51 existing computational feature sets to estimate perceptual texture similarity. They found that none of these feature

sets performed well compared to the performance of human observers. It was also observed that none of the 51 feature sets calculate higher order statistics (HOS) within the spatial extent larger than 19×19 pixels. Dong and Chantler [8] further compared three types of image properties, including the magnitude spectrum, local image exemplars and contours, for perception of texture. The results showed that the contour data encoding long-range HOS provides more important visual cues than the other image properties. However, contours cannot be accurately detected in some cases, for example, if the contrast between the foreground and background is low. On the contrary, the phase spectrum that also encodes long-range HOS [27] can be easily computed. It should be noted that the local phase data, which can be calculated based on the short-term Fourier transform [26] or quadrature filters [31], only exploits HOS in the relatively short-range spatial extent.

Therefore, we hypothesise that the phase spectrum is more important to texture similarity than the magnitude data. We are inspired to design and perform a psychophysical experiment using human observers. This experiment is used to investigate our hypothesis using two sets of property images: phase-only and magnitude-only (see Figure 1). The two sets of property images can be obtained from original texture images and only contain the original phase and magnitude spectra respectively. Our results show that the humans' judgements derived using the phase-only images are more consistent with those obtained using the original images than the judgements that they make when the magnitude-only images are shown. In other words, the phase spectrum is more important to human perceptual texture similarity than the magnitude data.

We further apply the 51 feature sets that Dong et al. [7], [9] examined to texture similarity estimation along with the phase-only and magnitude-only images. Surprisingly, the results show that the magnitude data is more important to these feature sets than phase. Recalling the inconsistency between the judgments obtained using humans and the 51 feature sets [7], as well as the finding derived in the psychophysical experiment that the phase data is more important to humans than the magnitude data for texture similarity, hence, our conjecture is that this inconsistency results from the difference in the ability of humans and the 51 feature sets to exploit the phase data.

In this context, it is likely that the performance of the 51 feature sets [7], [9] can be boosted if they are enabled to exploit the phase data. To this end, we propose a multi-channel feature fusion approach by jointly exploiting the features extracted from the original and corresponding phase-only images using a feature set. Since the phase-only images exclude the interference of the original magnitude data, the use of these images makes the exploitation of the phase spectrum possible. In contrast, it is not practical to directly utilise the (Fourier) phase spectrum because of the phase unwrapping issue [40]. The performance of the 51 fused feature sets is superior to that of the original feature sets and that of the fusions of the features computed from the original and magnitude-only images, on average. We also apply the feature fusion approach to convolutional neural network (CNN) features [33] and obtain the similar observation.

The contributions of this paper include that: (1) the confirmation of the importance of phase to perceptual texture similarity; (2) the finding that the 51 feature sets [7], [9] incline to utilise the magnitude data rather than the phase data; (3) the proposal of a multi-channel feature fusion method to enable the 51 feature sets use the phase data; and (4) the observation that the multi-channel feature fusion method can also be generalised to CNN features.

The rest of this paper is organised as follows. We first review the related work in the next section. Then, we examine the importance of the phase spectrum to perceptual texture similarity in Section 3. In Section 4, we investigate the ability of the 51 feature sets to exploit the phase spectrum for perceptual texture similarity estimation. The multi-channel feature fusion approach is proposed and tested in Section 5. Finally, we present our conclusions and discuss the future work in Section 6.

2. Related work

2.1. Fourier analysis in humans' perception of imagery

Research into the effect of the phase or magnitude spectra on humans' perception of imagery can be traced back to 1970s. Kermisch [18] removed the magnitude spectrum from images and analysed the effect of the phase spectrum on image reconstruction. Given an image, Oppenheim and Lim [28] derived two property images: magnitude-only and phase-only (see Figure 1). They showed that the phase-only image still retains the structure contained in the original image while it is not the case for the magnitude-only image.

On the other hand, research has been conducted by partially modifying the phase spectrum. Piotrowski and Campbell [29] found that the human visual system is able to use only a few of phase data for recognition of objects. In [38], humans' visual sensitivity to randomisation and quantisation of the phase spectrum of natural images was examined. Hansen and Hess [14] investigated the relative quantity of the spatial phase alignment that humans use to identify natural image structures at different spatial frequencies. In [41], it was observed that the thresholds required for detecting different types of changes were significantly lower when original images were used than those required when the phase-scrambled images were used. Recently, Emrith et al. [10] examined the effect of the randomness of the phase spectrum on humans' perception of the changes in the appearance of surface texture.

The existing work also focused on the joint use of the phase and magnitude spectra. Bretel et al. [3] presented that the disturbances of both magnitude and phase spectra result in perceptual distortions. In [36], it was demonstrated that not only the phase spectrum but also the magnitude spectrum are useful for perception of natural images.

Although the phase spectrum is important to humans' perception of imagery, phase unwrapping has to be used to recover original phase values from the principal value range [40]. However, it is still an open problem [40]. This problem explains the scarcity of the image features designed based on the Fourier phase spectrum. We thus managed to enable existing image feature sets to exploit the phase data by fusing the features extracted from the original and phase-only images. Compared with the original feature sets, the fused feature sets utilise more complicated image characteristics and are more discriminant.

2.2. Perceptual texture similarity estimation

Texture similarity can be divided into perceptual and computational similarity according to different acquisition sources, i.e. humans and computer algorithms, respectively. Texture similarity estimation yields a quantitative value or a qualitative judgement concerning the likeness of two textures. This task is key to texture or material recognition. Compared to other texture analysis topics [19], [24-26], [39], perceptual texture similarity estimation [1-2], [37] has received less attention. Recently, Dong and Chantler [6], [9] evaluated the ability of 51 feature sets to estimate higher resolution humans' perceptual texture similarity rather than the binary similarity data (same or different) used in texture classification [25], [37] or retrieval [19], [24]. It was found that none of the 51 feature sets produced the comparable performance to that of humans. The further analysis showed that none of these feature sets compute higher order statistics (HOS) over the spatial extent larger than 19×19 pixels. In other words, the 51 feature sets cannot exploit the long-range interactions that humans utilise for estimating texture similarity [7]. As a type of HOS, the Fourier phase spectrum encodes the long-range complicated patterns that comprise local phase alignments [14]. Therefore, it is likely that the performance of the 51 feature sets can be boosted if we enable these to use the phase spectrum.

In the past few years, deep convolutional neural network (CNN) approaches [21], [22], [33] have become prevalent in the computer vision community. However, there is still no convincing theoretical proof of the availability of these approaches [23]. In contrast, we are more interested in the understanding of the usefulness of the phase spectrum for texture similarity. Also, we do not have sufficient higher resolution perceptual texture similarity data as this type of data is expensive to derive [8]. The limited data prevents us from training a CNN from scratch as those approaches normally require a huge number of labelled data for training. As a result, we only used the pre-trained and the corresponding fine-tuned CNNs in this study.

3. The importance of phase to human perceptual texture similarity estimation

The research of texture similarity estimation using the human-derived ground-truth data has been conducted [1-2], [7], [9]. The importance of the long-range interactions [12], [30], [35] exploited by the human visual system to texture similarity was highlighted by Dong et al. [7], [9]. These interactions encode the complicated image patterns that can be modelled using global higher order statistics (HOS) [27]. As a type of global HOS, the phase spectrum is hence hypothesised to be important to human perceptual texture similarity estimation, particularly, compared to the magnitude data. In this section, we investigate this hypothesis by performing two new pair-of-pairs comparison experiments using the phase-only and magnitude-only images respectively. The humans' judgements derived in the original pair-of-pairs experiment [4] are used as the benchmark data.

3.1. Experimental design

3.1.1 Stimuli

Considering the higher resolution human perceptual similarity data provided along with the *Pertex* database [5], [13], the 334 *Pertex* textures were used.

Phase-only images The method that Oppenheim and Lim [28] introduced was modified in order to obtain these images. Given an image, the Fourier transform was used to decompose it into the phase and magnitude spectra: P_1 and M_1. The inverse Fourier transform was applied to P_1 and a single-valued magnitude matrix. The resultant image was filtered by a 3×3 Gaussian filter. We further decomposed the filtered image into the phase and magnitude spectra: P_2 and M_2. The magnitude spectrum M_2 and the original phase spectrum P_1 were used together with the inverse Fourier transform. The resultant image is referred to as the phase-only image. The reason for using a second inverse Fourier transform is that we want to utilise a more random magnitude matrix than the single-valued matrix. Figure 1 (top) presents two phase-only images derived from the two original images shown in Figure 1 (middle) respectively.

Magnitude-only images These images were generated using the approach that Oppenheim and Lim [28] proposed. To be specific, an original image was decomposed into the phase and magnitude spectra: P_1 and M_1 using the Fourier transform. P_1 was replaced by a random noise matrix P_3 ($\in [-\pi, \pi]$). The magnitude-only image was derived by performing the inverse Fourier transform on P_3 and the original magnitude spectrum M_1. Figure 1 (bottom) shows two magnitude-only images obtained from the two images presented in Figure 1 (middle) respectively.

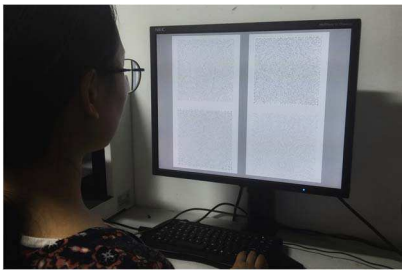

Figure 2: A trial and the setup used in the pair-of-pairs comparison experiment. Here, two pairs (left and right) of stimulus images are presented simultaneously. The task of the observer is to compare the similarity between the two pairs and decide on which pair is more similar than the other pair.

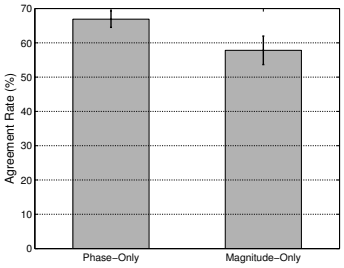

Figure 3: Means and 95% confidence intervals (error bars) of the agreement rates computed between the pair-of-pairs judgements that each observer made in the phase-only and magnitude-only experiments and those ($POPJ_{POP}$) obtained in [4].

3.1.2 Observers

In the two experiments, ten observers were used. All of them were naive to the experiment and with normal or corrected-to-normal vision.

3.1.3 Procedure

Two different experiments were performed using the phase-only and magnitude-only images respectively. The magnitude-only experiment was performed at least one week earlier than the phase-only experiment. This strategy reduces the learning effect on the observers. The procedure of each experiment is similar to that of the experiments that Clarke et al. [4] and Dong and Chantler [7] conducted. In total, 334 trials were performed in each experiment. The 334 *Pertex* [5], [13] textures were randomly used in the 334 trails respectively. In each trial, the observer was shown two pairs of images (see Figure 2). These pairs were placed at the left and right sides of the screen respectively. The task of the observer was to compare the two pairs and decide which pair is more similar than the other. If the left pair was considered as being more similar, he/she pressed the "←" key; otherwise, he/she pressed the "→" key.

3.1.4 Equipment

The stimuli were shown on an NEC LCD2090UXi monitor, which was linearly calibrated to $\gamma = 1$. The maximum luminance of the monitor was $120\text{cd}/\text{m}^2$. Thus, the stimuli appear that they are rendered using the similar lighting conditions to those in a bright room. Besides, the stimuli were resized to 512×512 pixels. In contrast, the resolution of the monitor was set to 1600×1200 pixels. The pixel dimensions of the monitor were 0.255mm×0.255mm (= 100 dpi). In this context, the stimuli were shown on the monitor with the size of 130.56mm×130.56mm.

3.1.5 Environment

The distance between the observer and the monitor was kept as around 50cm. This distance approximately produced the angular resolution of 17 cycles per degree (CPD). As a result, the stimulus subtended an angle of 14.89° in the vertical direction. We conducted the experiment in a dark room. This room has black, matte and opaque curtains and the matte walls.

3.2. Experimental results

We investigate whether the phase-only image or the magnitude-only image yields more significantly different judgements from those obtained using the original texture image. The method that Dong and Chantler [7] used was applied. The humans' pair-of-pairs judgements derived in the original pair-of-pairs experiment [4], i.e. $POPJ_{POP}$, were used as benchmarks. The judgements that the observers involved in the phase-only or magnitude-only experiments made were compared with $POPJ_{POP}$. The agreement rate (%) was used as the performance metric, which measures the percentage of the number of consisted judgements over the number of all judgements. Given an observer k ($k = 1, 2, \dots, 10$), the agreement rates computed between the judgements that this observer made in the phase-only and magnitude-only experiments and $POPJ_{POP}$ are denoted as: AR_k^{PO} and AR_k^{MO}, respectively.

Figure 3 displays the means and 95% confidence intervals of AR_k^{PO} and AR_k^{MO}. It is shown that the observers made more consistent judgements when phase-only images were presented with those derived in the original pair-of-pairs experiment [4] than the judgements that they made when magnitude-only images were shown. Figures 4 (left) and (right) show the most consistent trials obtained in the phase-only and magnitude-only experiments respectively. It is suggested that both the periodic and aperiodic patterns were available to the observers when the phase-only images were used; while only the periodic patterns were available when the magnitude-only images were used, for the pair-of-pairs comparison task.

3.3. Analysis

Furthermore, we examine the significance of the difference between AR^{PO} and AR^{MO}. We first performed the K-S test [20], [34] on the difference data in order to test its normality. As shown in Table 1, the distribution of the difference between AR^{PO} and AR^{MO} is normal. Then, a dependent *t*-test [17] was used to examine the significance

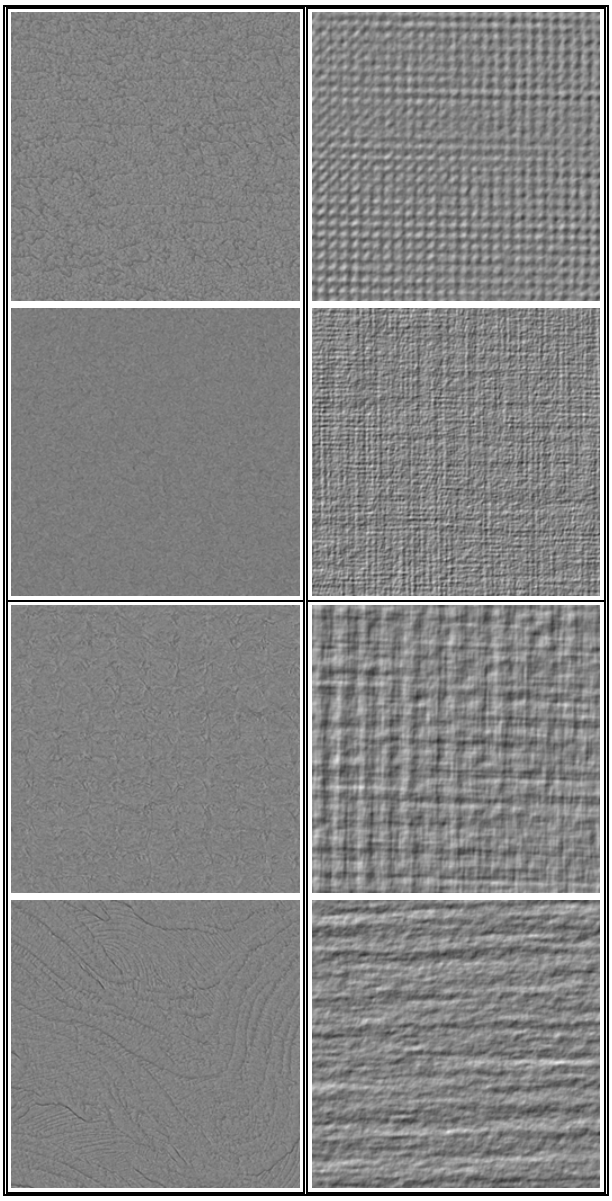

Figure 4: The most consistent trials (central quarters are shown) used in the two pair-of-pairs experiments: (left) phase-only and (right) magnitude-only, in which all ten observers decided that the top pair was more similar than the bottom pair.

K-S Test	Statistic	*df*	Sig. (*p*)	Is Normal?
$AR^{PO} - AR^{MO}$	0.250	10	0.077	Yes

Table 1: The results of the K-S test performed on the difference between AR^{PO} and AR^{MO}.

of the difference data. The results of the *t*-test are shown in Table 2. It shows that the observers agreed with those involved in the original pair-of-pairs experiment [4] when the phase-only images were used (M = 66.90, SE = 1.22) significantly more than that they agreed when the magnitude-only images were used (M = 57.82, SE = 2.12),

t-test	*t*	*p*	*r*	*df*	Is Sig.?
AR^{PO} vs. AR^{MO}	6.208	0.000	0.900	9	Yes

Table 2: The results of the dependent *t*-test (∂=0.05) performed between AR^{PO} and AR^{MO}. $r \geq 0.5$ indicates a strong effect [11].

$t(9) = 6.208$, $p < 0.05$, $r = 0.900$. These results indicate that the phase data is more important to human perceptual texture similarity estimation than the magnitude data.

4. How well do computational texture features exploit the phase spectrum for texture similarity?

In a survey of 51 feature sets, Dong and Chantler [6], [7] found that none of these exploit higher order statistics (HOS) over the spatial extent larger than 19×19 pixels. By conducting a series of evaluation experiments, they further found that those feature sets did not produce the consistent texture similarity data with humans' judgements. The inconsistency was attributed to the issue that none of the 51 feature sets use long-range interactions [12], [30], [35]. The complicated image patterns encoded in these interactions have been associated with global HOS [27], e.g. the phase spectrum. Therefore, we are inspired to investigate the ability of the 51 feature sets to exploit the phase spectrum for estimating human perceptual texture similarity. (Please refer to [6] for the details of these features sets). The pair-of-pairs comparison scheme that we used in Section 3 was used. We tested the 51 feature sets using the phase-only and magnitude-only images. The pair-of-pairs judgements derived using the two types of images were compared with those obtained using the original *Pertex* [5], [13] images, to examine the effect of the phase and magnitude spectra on texture similarity.

4.1. Experimental setup

4.1.1 Feature extraction

We performed feature extraction in the same manner as that Dong and Chantler [7] introduced. However, only the resolution of 1024×1024 pixels was used for the *Pertex* [5], [13] images. The phase-only or magnitude-only images were normalised to obtain the average intensity of 0 and standard deviation of 1. This processing eliminates the influence of 1st- and 2nd-order grey level statistics.

4.1.2 Computing pair-of-pairs judgements

The histogram-based and non-histogram-based feature vectors were first L_1 and L_2 normalised respectively. Then, the *Chi-Square* [32] and *Euclidean* distances were used to compute the distance between the two types of feature vectors, respectively. In terms of a feature set, the pair-wise distance matrix calculated between the phase-only or magnitude-only images was linearly stretched to [0, 1]. The similarity matrix was derived by subtracting this matrix from 1. The pair-of-pairs judgment was obtained according

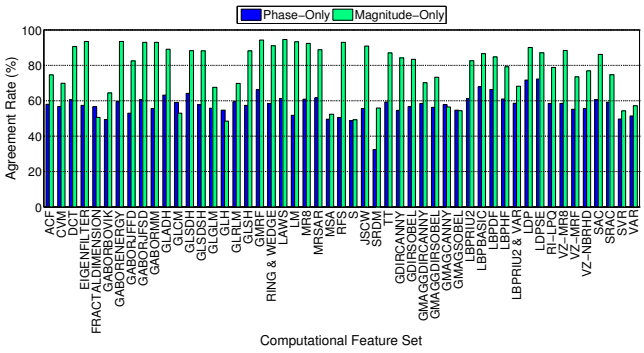

Figure 5: Each bar group shows the agreement rate computed between the pair-of-pairs judgements obtained from the original and phase-only images using a feature set (left) and that computed between the judgements obtained from the original and magnitude-only images using the same feature set (right).

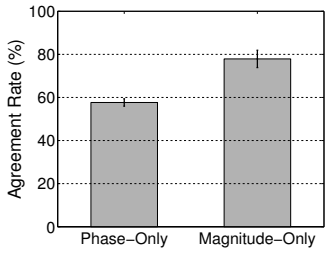

Figure 6: Means and 95% confidence intervals (error bars) of the agreement rates computed between the pair-of-pairs judgements obtained from the original and phase-only images and those computed between the judgements obtained from the original and magnitude-only images using the 51 feature sets.

to the difference between the similarities of the two pairs of images used in a trial of the pair-of-pairs experiments [7]. As a result, two judgement sets were derived using the phase-only and magnitude-only images, respectively.

4.2. Results

Given a feature set, the two sets of judgements obtained above were compared with the judgement set derived from the original *Pertex* [5], [13] images. The agreement rate (%) [7] was used to measure the consistency between two judgement sets. Using a feature set f ($f \in \{1, \cdots 51\}$), the agreement rate computed between the judgements obtained using the original and phase-only images and that calculated between the judgements derived using the original and magnitude-only images are denoted as: CAR_f^{PO} and CAR_f^{MO}. Figure 5 shows the values of CAR^{PO} and CAR^{MO} produced by the 51 feature sets. It can be seen that the agreement rate calculated between the judgements derived from the original and phase-only images is lower than that computed between the judgements obtained from the original and magnitude-only images in most cases. Furthermore, Figure 6 displays the means and 95% confidence intervals of the CAR^{PO} and CAR^{MO} values computed across the 51 feature sets. It is shown that the average agreement rate computed between the original and phase-only judgements is lower than that computed between the original and magnitude-only judgements.

K-S Test	Statistic	*df*	Sig. (*p*)	Is Normal?
$CAR^{PO} - CAR^{MO}$	0.093	51	0.200	Yes

Table 3: The results of the K-S test performed on the difference between CAR^{PO} and CAR^{MO}.

t-Test	*t*	*p*	*r*	*df*	Is Sig.?
CAR^{PO} vs. CAR^{MO}	-11.395	0.000	0.850	50	Yes

Table 4: The results of the dependent *t*-test (∂=0.05) performed between CAR^{PO} and CAR^{MO}. $r \geq 0.5$ indicates a strong effect.

4.3. Analysis

The *t*-test [17] was first used to examine the significance of the difference between CAR^{PO} and CAR^{MO}. Since *t*-test assumes that the input data follows the normal distribution, we used the K-S test [20], [34] to examine the normality of the difference data. Table 3 reports the results of the K-S test. It is shown that the distribution of the difference between CAR^{PO} and CAR^{MO} is normal. A dependent *t*-test was then applied to CAR^{PO} and CAR^{MO}. The results are presented in Table 4. As can be seen, the difference between the agreement rates computed between the judgements obtained using the original and phase-only images (M = 57.67, SE = 1.72) is significantly lower than those calculated between the judgements derived using the original and magnitude-only images (M = 77.88, SE = 4.03), $t(50)$ = -11.395, $p < 0.05$, r = 0.850. The above results demonstrate a contradict trend with those obtained using human obserbers in Section 3.

Therefore, we further compared the judgements obtained using the 51 feature sets with the humans' judgements obtained in Section 3. Using the phase-only and magnitude-only images, two sets of agreement rates were produced respectively and are shown in Figure 7. It can be seen that none of the 51 feature sets produced the higher agreement rate than 67.60% compared against the human data no matter whether the phase-only or magnitude-only images were used. With regard to the two types of images, the average agreement rates computed across the 51 feature sets are 54.44%±1.48 and 59.57%±1.33 respectively. In particular, the agreement rate calculated between humans' judgments and those obtained using the 51 feature sets went up when they could not use the phase data, compared to that calculated when they were able to use these data. Since it has been shown that humans use the phase data to estimate texture similarity, it is likely that the 51 feature sets do not exploit this type of data as well as that humans perform.

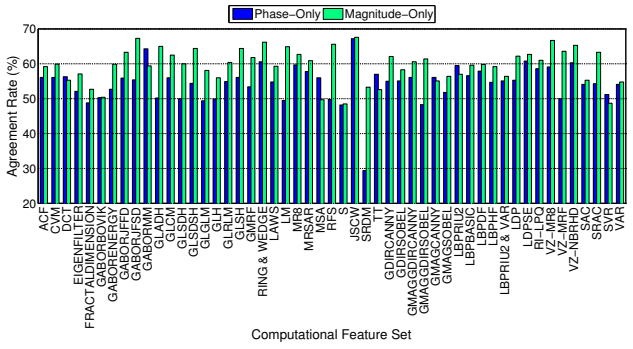

Figure 7: The agreement rates computed between the pair-of-pairs judgements of humans and those derived using the 51 feature sets when phase-only and magnitude-only images are used.

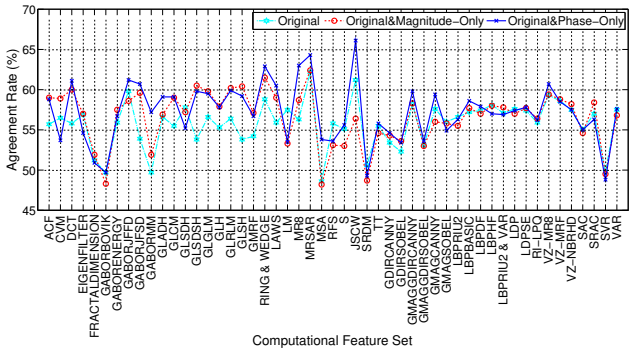

Figure 8: The agreement rates computed between the judgements of the humans used in the pair-of-pairs experiment [4] and those derived using three versions of each of the 51 feature sets.

5. The multi-channel feature fusion method exploiting the phase spectrum

In Section 3, it has been shown that the phase data is more important to humans for estimating texture similarity than the magnitude data. Nevertheless, the opposite observation was obtained for the 51 feature sets in Section 4. These results indicate that the inconsistency between the performance of humans and those obtained using the 51 feature sets is due to the issue that these feature sets do not exploit the phase data as well as that humans do.

In this context, it is likely that the performance of the 51 feature sets can be boosted if we enable these to better exploit the phase data than that they have performed. The straightforward solution can be achieved by extracting features from the phase spectrum. To our knowledge, however, rare image features are designed based on this type of data. This dilemma is due to the known phase unwrapping problem [40]. Alternatively, we propose to use the phase-only image generated from an image to exploit the phase data because this image excludes the interference of the original magnitude spectrum. Consequently, we are able to "force" the feature set to utilise the phase data by fusing the features extracted from the phase-only image with those computed from the original image. Considering the features based on convolutional neural networks (CNN) [21-22], [33] normally produce state-of-the-art results, we also propagate this method to CNN features.

5.1. Multi-channel feature sets exploiting the phase data

Given a feature set f, the feature vectors extracted from the original and phase-only images are denoted as: F_{Orig} and F_{PO} respectively. The fused feature vector $F_{Orig+PO}$ was computed according to:

$$F_{Orig+PO} = fuse(F_{Orig}, F_{PO}), \quad (1)$$

where $fuse()$ is the fusion method. We used two different fusion methods for continuous and histogram-based feature vectors respectively. For the continuous feature vectors, we used the Canonical Correlation Analysis (CCA) [15] method. On the other hand, the histogram-based feature vectors were directly concatenated because the CCA method cannot generate a histogram-based feature vector.

We also fused the feature vector F_{Orig} with that extracted from the magnitude-only image for comparison purposes. This fused feature vector is denoted as: $F_{Orig+MO}$. The scheme shown in Eqation (1) was perfomed for each of the 51 feature sets and CNN [21-22], [33] feature sets.

5.2. Perceptual texture similarity estimation using the multi-channel feature sets

We tested the two types of fused feature sets using the pair-of-pairs comparison task. The humans' judgements obtained using the original images [7] were used as the ground-truth data. The experimental setup that Dong and Chantler [7] introduced was used.

5.2.1 Using the 51 feature sets

Figure 8 shows three sets of agreement rates calculated between humans' pair-of-pairs judgements and those derived using three versions of each of the 51 feature sets: F_{Orig}, $F_{Orig+PO}$ and $F_{Orig+MO}$, respectively. It can be seen that: (1) the fusion of the features extracted from the original and phase-only images, i.e. $F_{Orig+PO}$, provided the higher agreement rate than that of the features computed from the original and magnitude-only images, i.e. $F_{Orig+MO}$, for 30 out of the 51 feature sets. In contrast, this number is 21 for the case that $F_{Orig+MO}$ outperformed $F_{Orig+PO}$; and (2) the two types of fused feature sets: $F_{Orig+PO}$ and $F_{Orig+MO}$ usually produced better results than those obtained using the corresponding original feature set: F_{Orig} (57.21%±3.62 vs. 55.79%±2.87 and 56.56%±3.36 vs. 55.79%±2.87). These results suggest that the performance of the 51 feature sets can be improved when they are enabled to utilise the phase data encoded in the phase-only images, on average. Although the joint utilisation of the

magnitude-only images can also boost the average performance of these feature sets, the margin is lower than that improved by using the phase-only images. This observation further emphasises the importance of the phase spectrum to texture similarity.

5.2.2 Using CNN features

We further applied the multi-channel feature fusion method to CNN [21-22], [33] features. Since we only have a relatively small texture dataset with the higher resolution human perceptual similarity data, it is not practical to train a CNN from scratch using this dataset. However, the features extracted from the penultimate fully-connected (FC) layer of a pre-trained CNN are normally considered as generic. Therefore, we extracted the FC features using a pre-trained CNN: VGG-VD16 [33]. In addition, we fine-tuned this CNN as this operation usually improves the performance. The FC feature sets were extracted using the pre-trained and fine-tuned VGG-VD16 [33] models. Given a CNN model, three feature sets: FC_{Orig}, FC_{PO} and FC_{MO} were extracted from the original *Pertex* [5], [13] dataset and its phase-only and magnitude-only variants, respectively.

The multi-channel feature fusion method was applied to the three feature sets. The two types of fused feature sets: $FC_{Orig+PO}$ and $FC_{Orig+MO}$ were used to estimate texture similarity. The estimated similarity data was compared against humans' data [4]. The performances obtained using the three single channel CNN feature sets and the two fused CNN feature sets are shown in Table 5. As can be observed, the fusion of the CNN features extracted from the original *Pertex* [5], [13] and phase-only images: $FC_{Orig+PO}$ outperformed the fusion of the CNN features extracted from the original and magnitude-only images: $FC_{Orig+MO}$ with a large margin. The fused features obtained from the fine-tuned CNN model generated better results than those derived using the fused features obtained from the pre-trained model. Especially, the best performance: 73.40% obtained using $FC_{Orig+PO}$ features that were extracted from the fine-tuned CNN model is very close to the agreement rate of 73.9% calculated between the judgements of two different groups of human observers [6-7].

Model	FC_{Orig}	FC_{PO}	FC_{MO}	$FC_{Orig+PO}$	$FC_{Orig+MO}$
Pre-trained	69.10	66.30	62.80	70.70	64.30
Fine-Tuned	72.80	69.20	67.50	**73.40**	65.10

Table 5: The agreement rates (%) computed between the human perceptual pair-of-pairs judgements [4] and those obtained using the features that are extracted from original *Pertex* images [5], [13], phase-only images, magnitude-only images, and two types of fused CNN features: $FC_{Orig+PO}$ and $FC_{Orig+MO}$. Here, two VGG-VD16 [33] CNN models: pre-trained and fine-tuned are used.

6. Conclusions and future work

It has been highlighted that the Fourier phase spectrum is important to perception of the appearance of natural images [14], [18], [27], [28], [38]. Particularly, the phase spectrum is more important to perception of imagery than the magnitude spectrum [28]. However, it is unknown whether or not this is the case for texture similarity. In the previous studies, Dong et al. [6], [7], [9] found that none of 51 existing texture feature sets compute higher order statistics (HOS) over spatial regions larger than 19×19 pixels. In contrast, the Fourier phase spectrum encodes long-range HOS [27].

Therefore, we were inspired to investigate the effect of the phase spectrum on texture similarity. We first performed a psychophysical experiment in order to examine which of the phase and magnitude spectra plays a more important role in this task. The phase-only and magnitude-only images obtained from original *Pertex* [5], [13] images were used as stimuli. The experimental results showed that the phase data is more important to humans for estimating texture similarity than the magnitude data. Furthermore, we tested the 51 feature sets using the perceptual texture similarity estimation task. It was found that, however, the magnitude data is more important to these feature sets than the phase spectrum.

Considering the inconsistency between the performance of human observers and those obtained using the 51 feature sets [6], [7], [9], our conjecture was that the inconsistency was resulted from the difference in the ability of humans and these feature sets to utilise the phase spectrum. In this context, it is likely that the performances of the 51 feature sets can be boosted if we enable these to better exploit the phase spectrum than they have performed. To this end, the fusion of the features computed from the original and phase-only images was conducted for each feature set. The 51 fused feature sets were applied to perceptual texture similarity estimation. Our results showed that the average performance of the 51 feature sets was improved by using the fusion scheme. Besides, the similar observation can be obtained when the fusion scheme was applied to CNN features [33]. We attribute these encouraging results to the importance of the phase data to texture similarity.

It is noteworthy that Järemo Lawin et al. [16] proposed an efficient phase unwrapping approach recently. In our future work, we intend to explore the possibility of extracting features from the phase data generated from the noisy Fourier phase spectrum using this approach. However, the key point of the current work is that we have shown the importance of the Fourier phase spectrum to texture similarity. This finding may encourage more studies in this direction.

Acknowledgement

Junyu Dong was supported by the National Natural Science Foundation of China (NSFC) (No. 61271405, 41576011) and the Ph.D. Program Foundation of Ministry of Education of China (No. 20120132110018).

References

[1] N. Abbadeni. Computational Perceptual Features for Texture Representation and Retrieval. IEEE Transactions on Image Processing, 20(1):236-246, 2011.

[2] M. Amadasun, and R. King. Textural features corresponding to textural properties. IEEE Transactions on System, Man and Cybernetics, 19(5):1264-1274, 1989.

[3] J. Bretel, T. Caelli, R. Hilz, and I. Rentschler. Modelling perceptual distortion: Amplitude and phase transmission in the human visual system. Hum. Neurobiol., 1:61-67, 1982.

[4] A.D.F. Clarke, X. Dong and M. J. Chantler. Does Free-sorting Provide a Good Estimate of Visual Similarity? In Proc. the 3rd International Conference on Appearance, 2012.

[5] A.D.F. Clarke, F. Halley, A.J. Newell, L.D. Griffin, and M.J. Chantler. Perceptual Similarity: A Texture Challenge. In Proc. BMVC, 2011.

[6] X. Dong. Perceptual Texture Similarity Estimation. PhD thesis, Heriot-Watt University, 2014.

[7] X. Dong and M. J. Chantler. The Importance of Long-Range Interactions to Texture Similarity. In Proc. CAIP, 2013.

[8] X. Dong and M. J. Chantler. Perceptually Motivated Image Features Using Contours. IEEE Transactions on Image Processing, 25(11):5050-5062, 2016.

[9] X. Dong, T. Methven, and M. J. Chantler. How Well Do Computational Features Perceptually Rank Textures? A Comparative Evaluation. In Proc. ICMR, 2014.

[10] K. Emrith, M. J. Chantler, P. R. Green, L. T. Maloney, and A. D. F. Clarke. Measuring Perceived Differences in Surface Texture due to Changes in Higher Order Statistics. Journal of Optical Society of America A, 27(5):1232-1244, 2010.

[11] A. Field. Discovering Statistics Using SPSS, 2009.

[12] D. J. Field, A. Hayes and R. F. Hess. Contour integration by the human visual system: evidence for a local 'association field'. Vision Research, 33:173-193, 1993.

[13] F. Halley. Pertex v1.0, 2011. http://www.macs.hw.ac.uk/texturelab/resources/databases/pertex/.

[14] B. C. Hansen and R. F. Hess. Structural sparseness and spatial phase alignment in natural scenes. Journal of Optical Society of America A, 24:1873-1885, 2007.

[15] D. R. Hardoon, S. R. Szedmak, and J. R. Shawe-taylor. Canonical Correlation Analysis: An Overview with Application to Learning Methods. Neural Computation, 16(12):2639-2664, 2004.

[16] F. Järemo Lawin, P. E. Forssén, and H. Ovrén. Efficient Multi-frequency Phase Unwrapping Using Kernel Density Estimation. In Proc. ECCV, 2016.

[17] F.B. Joanl. Guinness, Gosset, Fisher, and Small Samples. Statistical Science, 2(1):45-52, 1987.

[18] D. Kermisch. Image reconstruction from phase information only. Journal of Optical Society of America A, 60(1):15-17, 1970.

[19] F. Khelifi and J. Jiang. k-NN Regression to Improve Statistical Feature Extraction for Texture Retrieval. IEEE Trans. on Image Precessing, 20(1):293-298, 2011.

[20] A.N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. Giornale dell'Istituto Italiano degli Attuari, 4: 83-91, 1933.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In Proc. NIPS, 2012.

[22] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. Nature, 521, 2015.

[23] J. Malik. What led computer vision to deep learning? Communications of the ACM, 60(6):82-83, 2017.

[24] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell., 8:837-842, 1996.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution Grey-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. Pattern Anal. Mach. Intell., 24:971-987, 2002.

[26] V. Ojansivu, E. Rahtu, and J. Heikkila. Rotation Invariant Local Phase Quantisation for Blur Insensitive Texture Analysis. In Proc. ICPR, pp. 1-4, 2008.

[27] B.A. Olshausen and D.J.Field. Natural image statistics and efficient coding. Network: Computation in Neural Systems, 7(2):333-339, 1996.

[28] A. V. Oppenheim and J. S. Lim. The Importance of Phase in Signals. Proceedings of the IEEE, 69(5):529-541, 1981.

[29] L. N. Piotrowski and F. W. Campbell. A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. Perception, 11:337-346, 1982.

[30] U. Polat. Functional architecture of long-range perceptual interactions. Spatial Vision, 12:143-162, 1999.

[31] J. Portilla and E.P. Simoncelli. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. Int'l J. Computer Vision, 40(1):49-71, 2000.

[32] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. Numerical Recipes in C, 2nd ed., 1992.

[33] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Visual Recognition. In Proc. ICLR, 2015.

[34] N. Smirnov. Tables for estimating the goodness of fit of empirical distributions. Annals of Mathematical Statistics, 19:279-281, 1948.

[35] L. Spillmann and J.S. Werner. Long-range interactions in visual perception. Trends in Neurosciences, 19:428-434, 1996.

[36] Y. Tadmor and D. J. Tolhurst. Both the phase and the amplitude spectrum may determine the appearance of natural images. Vision Res., 33:141-145, 1993.

[37] H. Tamura, S. Mori, and T. Yamawaki. Textural Features Corresponding to Visual Perception. IEEE Trans. Systems, Man, and Cybernetics, 8:460-473, 1978.

[38] M. G. A. Thomson, D. H. Foster, and R. J. Summers. Human sensitivity to phase perturbations in natural images: a statistical framework. Perception, 29: 1057-1069, 2000.

[39] M. Varma and A. Zisserman. A Statistical Approach to Material Classification Using Image Patch Exemplars. IEEE Trans. Pattern Anal. Mach. Intell., 31:2032-2047, 2009.

[40] L. Ying. Phase unwrapping. Wiley Encyclopedia of Biomedical Engineering, 6:1-11, 2006.

[41] A. Yoonessi and F. A. A. Kingdom. Comparison of sensitivity to color changes in natural and phase-scrambled scenes. Vision Res., 25: 676-684, 2008.