# Texture Characterization and Compression based on Human Perception in the JPEG2000 framework

Gamal Fahmy
Lane Department of Computer Science and
Electrical Engineering
West Virginia University
Email: {fahmy@csee.wvu.edu }

John Black Jr., and Sethuraman Panchanathan
Research Center for Ubiquitous Computing (CubiC)
Department of Computer Science and Engineering
Arizona State University
Email: {john.black@asu.edu , panch@asu.edu }

**Abstract**

Over the last decade perceptually based image compression has gained significant importance. This is because it relies on Human Visual Perception (HVP) in measuring the reconstruction quality in the compression process, as humans are the end users for images. Visual data that is perceived by humans can be characterized in terms of three parameters, Magnitude, Phase and Orientation of the spatial frequency content. While existing perceptually based image compression techniques exploits the first parameter, the novel contribution of this paper is its focus on the use of phase data for perceptually based texture compression. In this paper a HVS based texture characterization approach is applied to measure the perceived (by humans) phase coherence in the image. Then images are more compressed after removing the unperceived phase redundancy. Finally subjective tests are performed to measure the reconstruction quality of the proposed compression approach. The proposed compression algorithm has been applied in the JPEG2000 framework. Simulation results that demonstrate the efficiency of the proposed approach are presented.

## Introduction

Despite rapid progress in mass-storage density, processor speeds, and digital communication, the demand for data storage and data-transmission bandwidth continues to surpass the capabilities of available technologies. Hence there have been increasing efforts to develop reliable compression techniques that provide cost effective solutions for the storage and transmission of visual data.

One goal of image or video compression is to remove *redundancy* from the data stream in a manner that allows the redundant data to be regenerated during the decompression process. This process *(called lossless compression)* uses mathematical algorithms to remove redundant information.

However, even after all of the redundancy has been eliminated from the visual data, there are still opportunities for further compression, based on the perceptual limitations of the human visual system (HVS). These limitations have been extensively studied during the last several decades, allowing researchers to distinguish between perceived and unperceived visual data. The limits of visual perception provide an opportunity to discard further (unperceivable) data. The discarding of this unperceivable data is known as *lossy* compression.

Early *lossy* image compression techniques relied on minimizing the distortion between the original image and the reconstructed image using *mathematical* measures, such as PSNR (peak to peak signal to noise ratio) or RMSE (root mean square error). However these mathematical measures where later shown to correlate rather poorly with human subjective judgments of reconstruction quality.

Early research into perceptually-based compression was done primarily by audio and speech researchers. Later, a new class of perceptually-based *image* compression algorithms was developed. These new algorithms used perceptually-based metrics (which were based on the contrast, luminance and frequency con tent of the image) and provided superior compression ratios, while maintaining the same level of perceived reconstruction quality as the earlier lossless compression algorithms. It is this perceptually-based approach that is the subject of this paper.

## Background

The Discrete Cosine Transform (DCT) was one of the first successful transforms that decomposed data into multiple spatial frequency bands. Its success was largely due to its energy concentration characteristics. Several perceptually-based compression techniques have used DCT. For example, Watson [4] presented the well known DCTune perceptual coding method, which used DCT and then used an image-dependent quantization matrix (based on the frequency sensitivity and contrast masking of the HVS at each spatial frequency) to modify the resulting DCT coefficients. Hontsch and Karam [5] modified Watson's algorithm, by specifying a computation method for the quantization matrix on a block by block basis.

Recently, the discrete wavelet transform has emerged as a powerful tool in data compression. This transform maps an image into a set of coefficients that constitutes a multi-scale representation of the image. In [2], a promising perceptual image coder (PIC) technique which employs the wavelet transform is presented, where a fixed HVS-based quantization level is used for each wavelet band (i.e. each range of spatial frequencies). In the recent JPEG2000 standard, HVS based quantization is adopted as one of the compression modes. Considering that (1) wavelet based coding has been chosen as the transform technique for JPEG2000, and (2) psycho-visual studies indicate that the HVS processes visual information in a multi-scale manner (like the wavelet transform), we have chosen to explore the concept of HVP (Human Visual Perception) based *texture* characterization and compression in the wavelet domain.

In the HVS (Human Visual System) literature, texture can be usefully characterized in terms of three main parameters: (1) Spatial frequency magnitude (SFM) which refers to the local contrast of the frequency coefficient, (2) Spatial frequency phase (SFP) which represents the spatial *position* (both up/down and right/left) of gradient edges in a local region, (3) Spatial frequency orientation (SFO), which represents the *angle* of edges within a local region.

In general, previous HVS based image compression algorithms have exploited the magnitude parameter (SFM), by quantizing the spatial amplitude changes, based on their spatial frequency – hence assigning fewer bits to the less perceptible spatial frequency coefficients. However the perceptibility of both the second and third parameters, (SFP) and (SFO), haven't been exploited for image compression.

In [1], Marr hypothesized that the content of images is analyzed by the HVS through several independent spatial frequency channels.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Marr applied a range of Laplacian-of-Gaussian filters (each representing a different spatial frequency) to images, and observed that there are often points in the image where all of these filter outputs have zero crossings at the same spatial location. These points are known as *coincidents*. Marr developed an algorithm to detect these *coincidents* in an image by processing it through a number of filters, each of which represents a different spatial frequency.

In reviewing the existing literature, it is evident that there are unexplored opportunities for perceptually-based compression, using Spatial Frequency Phase (SFP).

## Theory

We define the term SFP to be the parameter that specifies the spatial location of gradient edges in an image (i.e. the up/down and right/left location of the edge within a spatial locality). Based on Marr's work we assume that the HVS analyzes the content of the image through a number of semi-independent spatial frequency channels. There are some points in an image where the content of all (or most) of these channels correlate with each other. These points of phase coherence (i.e. *coincidents*) are perceived as visual features. When the phases of the various spatial frequency channels all have a zero value at the same location. The number of these *coincidents* is a useful a measure of the phase coherence between bands. Images that contain no *coincidents* are perceived to have no salient features.

Because uncorrelated phase information in an image is not readily perceived, we can discard all (or part) of the SFP information in each wavelet (spatial frequency) band during the compression process, and then assign an *arbitrary* phase value to each band when the image is reconstructed, without affecting the perceived content of the image. Although the original and the reconstructed images are represented by a different set of wavelet coefficients, the two images are not *perceived* to be substantially different. Examples of these images include speckle texture images, such as, Metal, Fabrics, etc….

Any sinusoidal signal $\sin(\theta)$ is defined by a *phase* component $\theta$, and a *magnitude* component that can be calculated by taking the magnitude at any two 90-degree phase shift points in that signal, fig. 1, according to the following equation:

$$Mag = \sqrt{\sin(\theta)^2 + \sin(\theta + 90)^2} \quad (1)$$

The wavelet transform is accomplished with a pair of filters (a high-pass filter and a low-pass filter). The high-pass filter can be envisioned as the product of a pure sinusoid and a Gaussian signal. The wavelet coefficient produced by this high-pass filter indicates the magnitude of a single spatial frequency in the original image within the filter's neighborhood. By using two coefficients that are 90 degrees apart, it is possible to compute the *magnitude* of the sinusoidal component, while ignoring the *phase*.

In computing the magnitude for each neighborhood (while discarding the phase information) we derive a *single* value from each original *pair* of 90-degree-phase-shifted wavelet coefficients. Thus, the amount of data used to represent the *magnitude* of the spatial frequency content is only 1/2 of the information in the wavelet representation of that band. If this magnitude information can be shown to adequately represent the perceived spatial frequency content without the phase information, we can double the compression ratio, compared to wavelet representation.

Fig. 2 shows the values used to implement a 7-element, high-pass wavelet filter kernel. The 1st element of this kernel corresponds to the *zero crossing* of the sinusoidal component of the high-pass filter. The 4th element corresponds to a *peak* value of that same sinusoidal component of the high-pass filter. Therefore, these two elements correspond to two points on that sinusoidal component that are 90 degrees apart. This single pixel offset implies that wavelet coefficients that are separated by rounding half the kernel size (3) in the transformed image correspond to samples taken 90 degrees apart in the spatial domain, as shown in Fig. 2.

In this paper we apply Marr's method (with a modification) on multiple wavelet bands to characterize *textures*. (We use a wavelet kernel instead of a Laplacian of Gaussian kernel to decompose the image into spatial frequency bands). We note here that Marr detected *coincidents* across several frequency bands using a *circularly-symmetric* Laplacian-of-Gaussian filter that doesn't pass any orientation information. However, the wavelet transform employs independent *horizontal* and *vertical* spatial frequency filters, which do pass orientation information.

The wavelet transform of a 2D image is a 2D matrix of wavelet coefficients, as shown in fig. 3. These coefficients collectively encode the spatial frequency *magnitude* and *orientation* at each locality within the original image.

## Methodology

In our methodology, we calculate the number of *coincidents* in an image, in order to characterize and classify texture. The number of *coincidents* in the image indicates the amount of perceptible phase information, and thus the amount of perceived visual information that will be lost if the phase information is discarded. Thus, the number of *coincidents* can be thought of as a metric for the anticipated image degradation if our compression method is used (i.e. more *coincidents* => greater degradation).

### Calculating the number of Coincidents

*Step 1: Process the image through 16 different spatial frequency channels*

In calculating the number of *coincidents* in the image, we follow a technique similar to Marr's. We first process the image through into 16 different spatial frequency bands, (The 5 level dyadic decomposition that is the default in JPEG2000, fig. 3) which are each represented by a 2D array of real wavelet coefficients. This is done with a 9x7 wavelet kernel [3]. This makes our proposed approach compatible with the JPEG2000 standard.

*Step 2: Measure the correlation across all bands*

We then determine the number of *coincidents* in the image by (1) counting the number of coincident zero crossings across the 16 frequency bands for every original image point in the wavelet decomposed image, and (2) thresholding the coincident count at each point. Any count that survives the thresholding operation is considered to be a coincident.

### The Proposed Compression Algorithm

In order to discard the phase information, we replace each pair of coefficients at 90-degree shift points with the square root of the sum of the squares of those coefficients, according to the following equation.

$$Mag(m) = \sqrt{W^2(m) + W^2(m - [n/2])} \quad (2)$$

Where $W(m)$ is the $m^{th}$ wavelet coefficient, and $W(m-[n/2])$ is its 90-degree phase shift wavelet coefficient.

In order to make our algorithm compatible with JPEG2000, we use the 9x7 default kernel size, (where n=7 for the high pass direction). We thus replace each pair of 90-degree phase shift coefficients with a single value. (Note: For reasons to be given later, we retain the *signs* of both the coefficients, which represent the *quadrant* of the phase.) For the 2-D wavelet domain as in fig. 3, we apply equation (2) in the LH band in the high pass kernel direction (vertical).

Since the LH band in fig, 3 contains horizontal lines, we can apply equation (2), only in the column direction, as this is the direction that contains phase data. Similarly in the HL band, equation (2) can only be applied in the row direction. The shaded blocks in fig. 4 Shows the resulting wavelet coefficients after removing the translation phase in the image.

Hence for any wavelet decomposition, the coefficient arrays for the LH band and the HL band can be shrunk to half their original size.

Sometimes the wavelet coefficients at the two 90-degree phase-shift points used in equation (2) have the same sign, and sometimes they have different signs. However, this information is lost when the magnitude is computed, using the (positive) square root of the sum of the squares. Some improvement in the reconstructed image can be obtained by retaining the signs of these original coefficients, which represent the *quadrant* of the original coefficient pair. In doing so, we retain a small portion of the spatial frequency phase that was encoded by the original wavelet coefficient pair.

Because the number of coefficients has been reduced to half, the resulting coefficient array looks like fig 4, where the shaded region represents the reduced array of coefficients.

However, in order to make the coefficient array compatible with the JPEG2000 arithmetic coder, we duplicate those coefficients and attach the two sign bits (representing the quadrant) to the two identical values. This makes the number of the coefficients the same, but with duplicated coefficient values, which will correspond to a lower entropy value for the whole image.

## Procedure

We selected 100 texture images from well known texture databases [6]. We first calculated the number of *coincidents* in each of these images, as described in the previous section. Then we applied our proposed compression technique to all of these images, followed by the JPEG2000 arithmetic coding technique. The original and reconstructed images were then shown to subjects on a Trinitron monitor with a resolution of 1152*864 pixels. The viewing distance was 60 cm and the luminance conditions were according to the CCIR recommendations [7]. Both the original and the reconstructed images were displayed simultaneously, side-by-side. Some of the subjects who viewed these image had training in an image processing while others did not. They were all asked to give an evaluation of the degradation in the reconstructed image (as compared to the original image) according to the following scale: (5) “imperceptible”, (4) “perceptible, not annoying”, (3) ”slightly annoying”, (2) “annoying”, (1) “very annoying”.

## Results

Fig. 5 shows a plot of the MOS (mean opinion score) of the subjects, versus the number of *coincidents* for all of the images. (Each point on this plot represents a single texture image.) The two curves in fig. 6 show the entropy value of each of images before and after applying our coding algorithm. Table 1 compares the size of the compressed bit stream obtained with the JPEG2000 arithmetic coder, *with* and *without* applying our compression method, where in both cases the reconstructed image have the same quality.

## Discussion

There is a hyperbolic relationship between the number of *coincidents* in an image and the reconstructed quality, as expressed by the MOS. The higher the number of *coincidents* in an image, the higher the degree of phase coherence between the spatial frequency channels of the decomposed images, and the more likely humans will be to detect differences between the original and the reconstructed images, thus producing a lower MOS.

The *entropy* of the wavelet coefficient array after removing the phase information is much lower than the entropy of the wavelet coefficient array before removing the phase information, as shown in fig.6. As a result, the JPEG2000 arithmetic coder produces a more compressed bit stream for the phase-removed coefficient array than for the original coefficient array, as shown in Table 1.

The degree of JPEG2000 bit stream compression obtained with the proposed compression algorithm (as shown in Table 1) is not as great as the reduction in the *entropy* value (as shown in fig. 6). We attribute this to the fact that the arithmetic coder of JPEG2000 doesn’t specifically exploit the redundancy in the wavelet coefficient magnitudes generated from our approach. This redundancy is seen in the *identical* absolute values of pairs of 90-degree phase shifted coefficients in the final wavelet coefficient array. This of course leaves open an opportunity for future research in designing an arithmetic coder that can exploit this redundancy to achieve higher degrees of compression.

The method proposed in this paper determines the number of coincident points in any image, and then uses that information to decide whether phase removal will be useful. However, increasing or decreasing the contrast of the original image (prior to compression) can change the MOS of the reconstructed image. This is due to the fact that distortions introduced by the compression/reconstruction process might be below the threshold of human visual perception when the original contrast is low, but elevated above that threshold as the image contrast is increased. However, our proposed preprocessing step (which counts the number of *coincident*s in the image) also includes a quantizing step (due to round-off of the real number coefficients) so that low contrast images produce fewer zero crossings than high contrast images – and thus fewer *coincidents*. Thus, if the contrast of the original image is enhanced, it will contain more *coincidents*, and will be judged unsuitable for compression using the proposed algorithm.

## Conclusions

The exploitation of spatial frequency phase (SFP) and spatial frequency orientation (SFO) is a very immature (but promising) field of research. In this paper, we have presented a texture characterization and compression approach that discards non-perceivable spatial frequency phase in images. The proposed approach has been found to be useful for image compression, and could be incorporated into the JPEG2000 standard. Extensive simulation results (some of which have been presented in this paper) demonstrate the usefulness of the proposed approach.

## Unsolved Problems

Marr demonstrated that humans perceive spatially coincident zero crossings as salient features, indicating that spatial phase is significant in images with many *coincidents*. While this paper has explored the use of coincident zero crossings to detect important phase coherence, there has not yet been much research into the question of whether other types of phase coherence (such as the coincidence of the *peaks* in the various spatial frequency components) are also perceived as salient visual features. If so, those additional *coincidents* might further constrain the discard of phase information.

The design of an arithmetic coder, that exploits the redundancy in the wavelet coefficients obtained from our proposed compression method and also is compatible with the JPEG2000 standard would provide greater compression of JPEG2000 streams. There are also the

unanswered question of whether Spatial Frequency *Orientation* (SFO) information might be discarded in images that contain imperceptible orientation content, and how best to determine when this type of phase information might be successfully discarded. It is possible that significant compression performance improvement can be achieved by exploiting the limits of orientation perception in the HVS.

Due to the voluminous growth of visual data, future image databases will likely employ highly compressed formats. Since it will be desirable to perform image retrieval based on compressed domain coefficients, it is important that these coefficients represent salient visual features. If SFO and SFP indices prove to be useful, it will be important to incorporate measures of these types of content into future classification algorithms.

**References**


[1] D. Marr, " Vision", *WH Freeman, 1982.*

[2] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image dependent quantization and post-quantization", *ICASSP, 1989.*

[3] Marc Anotonini, Michel Barlaud, Pierre Mathieu and Ingrid Duabechies, " Image Coding using Wavelet transform", *IEEE Trans. on Image Processing, vol. 1, No. 2, April 1992.*

[4] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images", *Soc. For Info. Display Digest of Technical paper, 1993.*

[5] I. Hontsch and L. Karam, "APIC: Adapative perceptual image coding based on subband decomposition with locally adaptive weighting", *ICIP 1997.*

[6] Vistex database at the Vision and Modeling group at the MIT Media Lab,*www.white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html.*

[7] "Method for subjective assessment of quality of television pictures", *Int. Telecom Union, Geneva, 1990, CCIR Recommendation 600-1.*


**Table 1**

| Bit stream in JPEG2000/Image | without phase data | with phase data |
|---|---|---|
| Brick | 18k byte | 18.1k byte |
| Fabric | 15.2 k byte | 15.4 k byte |
| Water | 16.43k byte | 16.431k byte |
| Metal | 15.6k byte | 15.68k byte |

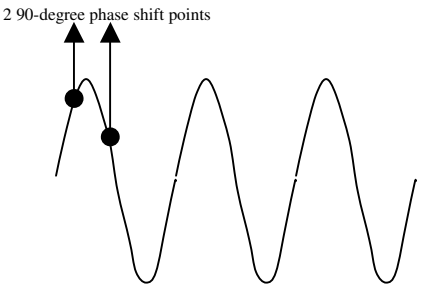


Fig.1: two 90-degree phase shift points from where you can calculate the magnitude of the signal and discard the phase

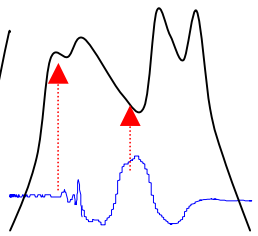

Fig. 2: 2 90-degree phase shift points from a signal that is convolved with a wavelet function

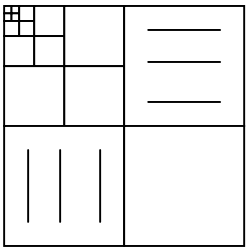

Fig. 3: 5 level dyadic wavelet decomposition, where the LH band detects horizontal lines, and the HL band detects vertical lines

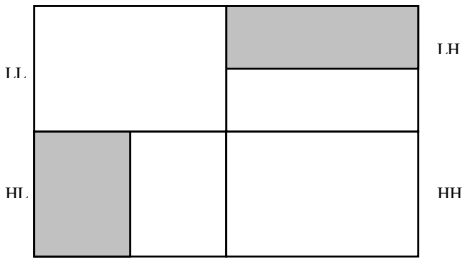


Fig.4: the four wavelet bands for 1 level of decomposition, where the shaded areas represents the results set of coefficients after applying our compression approach, this approach is only applied on the HL and LH bands and only in the high pass power spectra direction for this band

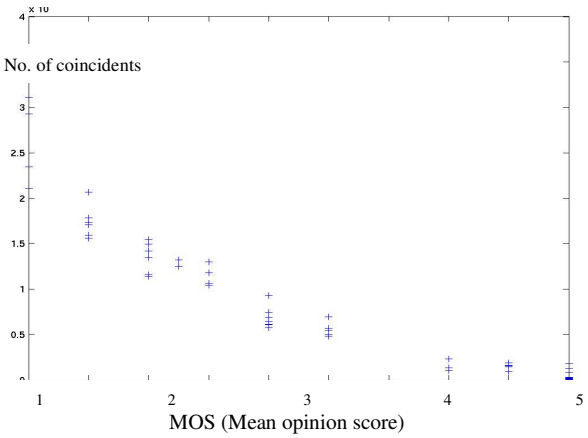


Fig.5: the relation between the number of *coincidents* indifferent images and their MOS reconstruction quality after discarding phase information

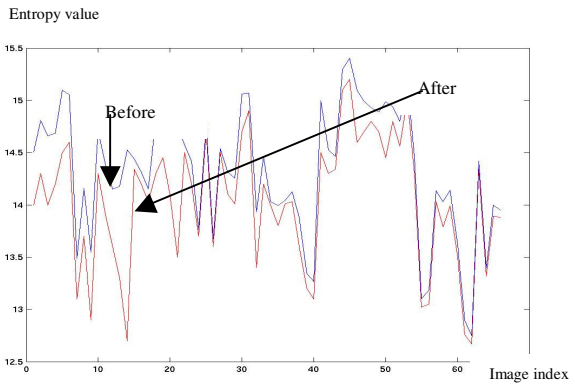


Fig.6: The entropy values before (Blue) and after (Red) applying our coding approach in JPEG2000