Improved Visible Differences Predictor
Using a Complex Cortex Transform

Alexey Lukin

Laboratory of Mathematical Methods of Image Processing,
Department of Computational Mathematics and Cybernetics,
Moscow Lomonosov State University, Russia
lukin@graphics.cs.msu.ru

Abstract

Prediction of visible differences involves modeling of the human visual response to distortions in the image data. Following the approach of Daly [1], this paper introduces several algorithm improvements allowing for more accurate calculation of threshold elevation and modeling of the facilitation effect during phase-coherent masking. This is achieved by introducing a complex-valued cortex transform that separates the response magnitude from the instantaneous phase within each band of the cortex transform. The magnitude component is used for calculation of the mask contrast. The phase component takes part in modeling of the facilitation effect.

Keywords: *Visible Differences Predictor, VDP, Cortex Transform, Image Quality Assessment, Complex Cortex Transform, Human Visual System, HVS, Masking.*

1. INTRODUCTION

Prediction of visible differences means estimation of subjective visibility of distortions in the image data. Algorithms for prediction of such differences are important in automated quality assessment of imaging systems, including lossy compression of video signals, assessment of transmission channel distortions, optimization of realistic image synthesis algorithms, etc.

Many image quality metrics have been proposed in the literature. The most successful objective metrics include models of the *Human Visual System* (HVS) for prediction of such effects as non-uniform sensitivity to spatial frequencies and visual masking, like the *Visual Differences Predictor* (VDP) proposed by Daly [1].

Daly's VDP uses a (modified) *cortex transform* [2] to decompose the image into subbands of different spatial frequency and orientation. It allows modeling of frequency-dependent and orientation-dependent masking in the human visual system. For each cortex transform band, the contrast of the difference signal and the contrast of the masking signal are evaluated. Threshold elevations are calculated from the contrast of the mask signal. They are used to calculate the probability of detection of the difference signal, subject to visual masking. The detection probabilities are summed over all cortex transform subbands.

The work of Mantiuk et. al. [3] proposes several improvements to the model of Daly, including evaluation of contrast in JND (just noticeable difference) units, and varying CSF (contrast sensitivity function) depending on the local luminance adaptation level.

A shortcoming of the "traditional" cortex transform is the inability to accurately model phase-invariant masking (explained in the next section). For example, a chirp image signal in Fig. 1a would produce an oscillating signal in each cortex band, as in Fig. 1b. This, in turn, would produce an oscillating mask contrast signal and oscillating threshold elevation, as in Fig. 1c.

In this paper, a modification of the cortex transform is introduced to obtain phase-independent estimates of the masking contrast, as in Fig. 1d. Section 2 describes the cortex transform and its phase-variance. Section 3 introduces the Complex Cortex Transform (CCT) and its computation algorithm. Section 4 illustrates the use of CCT for evaluation of masking thresholds in VDP. Section 5 presents the computational results of threshold elevations.

(a)   (b)   (c)   (d)
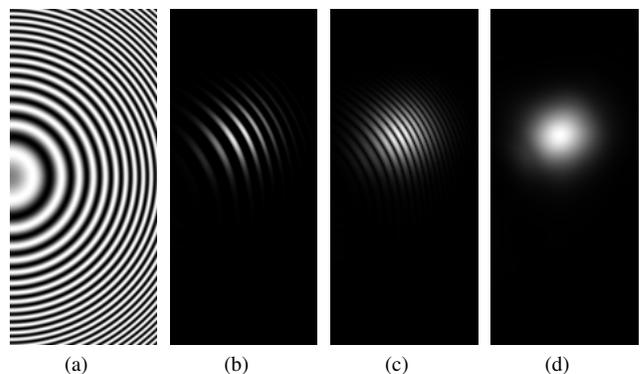
Figure 1: (a) A chirp image; (b) Response of a "traditional" cortex filter (only positive signal part is shown); (c) Threshold elevation image (or mask contrast) produced using a "traditional" cortex filter; (d) Threshold elevation image produced using a "complex" cortex filter proposed in this paper.

2. CORTEX TRANSFORM IN VDP

The cortex transform is first described by Watson in [2] as an efficient means of modeling the neural response of retinal cells to visual stimuli. The cortex filter in the frequency domain is produced as a product of 2 filters: the 'dom' filter providing frequency selectivity and the 'fan' filter providing orientation selectivity:

$$\text{cortex}_{k,l}(\rho,\theta) \equiv \text{dom}_k(\rho) \cdot \text{fan}_l(\theta), \qquad (1)$$

where k is the index of the frequency band, l is the index of orientation, and (ρ,θ) are polar coordinates in the frequency space (corresponding Cartesian coordinates will later be denoted as (ω_1,ω_2)). Fig. 2 illustrates frequency responses of several cortex filters, and Fig. 3a shows the example *impulse response* (point spread function) of the cortex filter.

The cortex transform decomposes the input image $I(i,j)$ into a set of subband images $B_{k,l}(i,j)$ (*cortex bands*) as follows

$$B_{k,l}(i,j) \equiv \mathcal{F}^{-1}\left\{\text{cortex}_{k,l}(\rho,\theta) \cdot \mathcal{F}\left\{I(i,j)\right\}\right\}, \qquad (2)$$

where $\mathcal{F}$ is the 2D discrete Fourier transform.

(a)	(b)
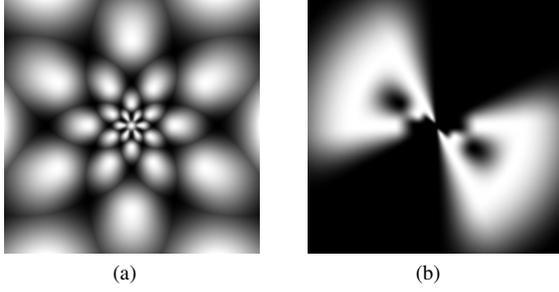
Figure 2: Frequency responses of several cortex filters (brightness represents gain for the given spatial frequency).
(a) Responses plotted separately for cortex$_{0,0}$; $_{1,0}$; $_{1,2}$; $_{1,4}$; $_{2,1}$; $_{2,3}$; $_{2,5}$; $_{3,0}$; $_{3,2}$; $_{3,4}$; $_{4,1}$; $_{4,3}$; $_{4,5}$; $_{5,0}$; $_{5,2}$; $_{5,4}$ — each filter produces 2 symmetrical blobs on a complex 2D frequency plane;
(b) Responses of 8 (out of 31) cortex filters are added together towards a constant gain.

It can be noted (Fig. 2) that cortex filters are linear-phase bandpass filters. Their frequency responses are designed to sum up to 1, which means that the sum of cortex bands is equal to the input image: $\sum_{k,l} B_{k,l}(i,j) = I(i,j)$.

One of the problems with the cortex transform is that real-valued cortex bands lack separation between magnitude and phase components of the neural response. According to formula (14.32) in [1], the strength of visual masking $T_e^{k,l}(i,j)$ (also known as *Threshold Elevation*) depends on the absolute value of the normalized mask contrast $\left| m_n^{k,l}(i,j) \right|$

$$T_e^{k,l}(i,j) = \left(1 + k_1 \left(k_2 \left| m_n^{k,l}(i,j) \right|^s \right)^b \right)^{1/b}, \qquad (3)$$

where k_1, k_2, s, and b are psychophysically derived constants.

The normalized mask contrast $m_n^{k,l}(i,j)$ is calculated as the cortex transform of the CSF-filtered input image in the perceptually linearized luminance scale:

$$m_n^{k,l}(i,j) = \mathcal{F}^{-1} \left\{ \text{cortex}_{k,l}\left(\rho, \theta\right) \cdot \text{csf}(\rho, \theta) \cdot \mathcal{F}\left\{ I(i,j) \right\} \right\} \quad (4)$$

It is easy to see that for sinusoidal mask signals, the mask contrast $m_n^{k,l}(i,j)$ exhibits oscillations between 0 and the magnitude of the masker (Fig. 6a,b). These oscillations are also present in the threshold elevation map. Their cause lies in insufficient separation of magnitude and phase of neural response by the modulus operation $\left| m_n^{k,l}(i,j) \right|$ in Eq. (3).

According to this simplified model, sinusoidal gratings produce maximal threshold elevation at positive and negative peaks of the masker waveform and absolutely no masking at zero crossings of the waveform. This contradicts with psychophysical data [4] assuming that sinusoidal gratings produce spatially uniform (or nearly-uniform) threshold elevation.

3. THE COMPLEX CORTEX TRANSFORM

3.1 Explanation of Goals

To eliminate this mismatch between calculated masking maps and psychophysical experiments, we suggest a more sophisticated model for separation of magnitude and phase information in the neural response. The importance of such separation has been acknowledged in [1] and [3], but the separation algorithm has not been elaborated.

Our proposed algorithm stems from the publication of Pollen and Ronner [5] which investigates phase relationships between adjacent cells in the visual cortex. It has been found that adjacent simple neural cells are often tuned to the same orientation and spatial frequency, but their responses differ by the phase angle that is often approximately $\pi/2$. In other words, receptive fields of corresponding neural cells comprise *quadrature* (phase-complementary) filters.

A recently published *Berkeley Wavelet Transform* [6] is the orthogonal wavelet transform using phase-complementary wavelets of 4 different orientations. Its filters are localized in space, frequency and orientation, so the transform is suitable for modeling of visual receptive fields.

However our method is based on the cortex transform [2],[1] because it allows for more flexible tiling of the 2D frequency plane and better frequency/orientation selectivity. Our goal is to modify the cortex transform in order to enable efficient magnitude/phase separation and provide the shift invariance of magnitude estimates for sinusoidal gratings.

3.2 A 2D Hilbert Transform

A well studied method for extracting magnitude and phase information of narrow-band 1D signals is the *Hilbert transform* [7]. The Hilbert transform can be considered as a filter that rotates every frequency component of the signal by $\pi/2$, for example $\mathcal{H}\left(\cos t\right) = \sin t$. The Hilbert transform can be used to convert a real-valued signal $f(t)$ into an analytic complex-valued signal $f(t) + i\mathcal{H}f(t)$, whose instantaneous magnitude and phase are defined as

$$A(t) = \sqrt{f^2(t) + (\mathcal{H}f(t))^2}, \qquad (5)$$

$$\phi(t) = \arctan \frac{\mathcal{H}f(t)}{f(t)}. \qquad (6)$$

An efficient computational algorithm for the Hilbert transform employs a direct $\pi/2$ phase rotation of the complex-valued Fourier spectrum $(\mathcal{F}f)(\omega)$ of signal $f(t)$:

$$(\mathcal{F}\mathcal{H}f)(\omega) = i \cdot \text{sgn}(\omega) \cdot (\mathcal{F}f)(\omega)$$

One problem with the Hilbert transform is that it does not have a trivial extension to the 2D case. One possible 2D extension called "skewed Hilbert transform" [8] applies a 1D Hilbert transform along only vertical or only horizontal direction in a 2D image. This is equivalent to multiplying the 2D Fourier spectrum of the image $(\mathcal{F}I)(\omega_1, \omega_2)$ by either $i\,\text{sgn}\omega_1$ or $i\,\text{sgn}\omega_2$ (for vertical or horizontal direction).

In order to build phase-complementary filters of arbitrary orientation, we are proposing a modification of the skewed 2D Hilbert transform that multiplies a 2D Fourier spectrum of the image by the similar filter

$$H_l(\omega_1, \omega_2) = i\,\text{sgn}(p\,\omega_1 + q\,\omega_2), \qquad (7)$$

where the line equation $p\,\omega_1 + q\,\omega_2 = 0$ specifies the desired "direction" of the modified 2D Hilbert transform, and l is the orientation index from Eq. (1).

3.3 Design of Quadrature Cortex Filters

For each cortex filter $\text{cortex}_{k,l}(\omega_1, \omega_2)$, we are designing a phase-complementary filter $\widehat{\text{cortex}}_{k,l}(\omega_1, \omega_2)$ with the same passband frequency range and orientation using the modified 2D Hilbert transform of the impulse response of the filter. Their frequency responses are related as

$$\widehat{\text{cortex}}_{k,l}(\omega_1, \omega_2) \equiv H_l(\omega_1, \omega_2) \cdot \text{cortex}_{k,l}(\omega_1, \omega_2),$$

where p and q in Eq. (7) are linked with the cortex filter orientation angle θ_c (depending on l) as

$$\tan \theta_c = \frac{p}{q}.$$

Fig. 3 shows impulse responses of 2 phase-complementary cortex filters for $k = 2$, $l = 1$, and $\theta_c = \pi/6$. It can be noted that $\text{cortex}_{k,1}$ is a linear-phase filter and its impulse response is centrally symmetrical (Fig. 3a).

(a) (b)
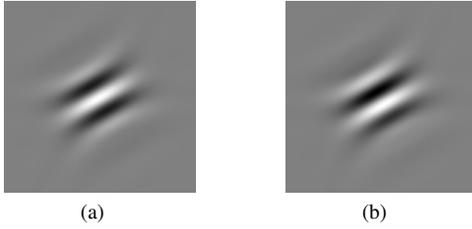
Figure 3: Impulse responses of a pair of phase-complementary cortex filters: (a) $\text{cortex}_{k,1}$; (b) $\widehat{\text{cortex}}_{k,1}$.

This pair of phase-complementary cortex filters can be combined into a single complex-valued quadrature filter:

$$\text{cc}_{k,l}(\omega_1, \omega_2) = \text{cortex}_{k,l}(\omega_1, \omega_2) + i\, \widehat{\text{cortex}}_{k,l}(\omega_1, \omega_2) \quad (8)$$

The decomposition of the input image into complex-valued subbands using filters $\text{cc}_{k,l}$ will be called a Complex Cortex Transform (CCT).

$$\text{CCT}_{k,l}(i,j) \equiv \mathcal{F}^{-1}\left\{\text{cc}_{k,l}(\omega_1, \omega_2) \cdot \mathcal{F}\{I(i,j)\}\right\} \quad (9)$$

3.4 Properties of the Complex Cortex Transform

1. Since both $\text{cortex}_{k,l}$ and $\widehat{\text{cortex}}_{k,l}$ are real filters, from Eqs. (2), (8), and (9) we obtain that cortex bands $B_{k,l}(i,j)$ are equal to the real part of corresponding CCT bands.

$$B_{k,l}(i,j) = \text{Re}\left\{\text{CCT}_{k,l}(i,j)\right\}$$

Substituting this into Eq. (4), we obtain

$$m_n^{k,l}(i,j) = \text{Re}\left\{\mathcal{F}^{-1}\left\{\text{cc}_{k,l} \cdot \text{csf} \cdot \mathcal{F}\{I(i,j)\}\right\}\right\}$$

2. The real and imaginary CCT bands contain phase-complementary responses, similarly to those described in [5]:

$$\text{Im}\left\{\text{CCT}_{k,l}(i,j)\right\} = \mathcal{F}^{-1}\left\{H_l \cdot \mathcal{F}\left\{\text{CCT}_{k,l}(i,j)\right\}\right\}$$

3. Similarly to the 1D case (Eqs. (5) and (6)), instantaneous magnitude and phase within each CCT band can be approximated as

$$A_{k,l}(i,j) = |\text{CCT}_{k,l}(i,j)|,$$

$$\phi(t) = \arg\left\{\text{CCT}_{k,l}(i,j)\right\}.$$

Here the separation of magnitude and phase information happens along the orientation θ_c of the corresponding cortex filter $\text{cc}_{k,l}$.

3.5 Summary of the CCT Algorithm

The Complex Cortex Transform defined by Eq. (9) is calculated similarly to the regular cortex transform (Eq. (2)), except the fact that the cortex filter is now complex-valued, and the resulting subband images $\text{CCT}_{k,l}(i,j)$ are complex-valued too. The algorithm looks like follows:

1. Complex-valued spectra $\text{cc}_{k,l}(\omega_1, \omega_2)$ of CCT filters are pre-calculated using Eqs. (1) and (8);

2. A 2D Fourier transform $(\mathcal{F}\{I\})(\omega_1, \omega_2)$ of the real-valued input image $I(i,j)$ is calculated;

3. The complex-valued spectrum from step 2 is multiplied by the complex-valued filter $\text{cc}_{k,l}$ from step 1, for each frequency range k and orientation l;

4. An inverse complex-valued 2D Fourier transform is calculated for each spectrum obtained at step 3 — this is the resulting band $\text{CCT}_{k,l}(i,j)$ of the Complex Cortex Transform.

In the discrete case, all Fourier transforms can be replaced by Discrete Fourier Transforms (DFT) and computed via FFT. Boundary effects need to be taken into consideration when using DFT filtering, because multiplication of discrete spectra leads to a *circular* convolution in the spatial domain. This is less of an issue with quickly decaying impulse responses of cortex filters, but may still require explicit extension of the image data beyond its support area.

4. USING CCT FOR MODELING OF MASKING

4.1 Modeling Threshold Elevation Using CCT Magnitude

Now since the Complex Cortex Transform is available for efficient separation of magnitude and phase information in the neural response, it can be incorporated into Eq. (4) to yield the new model of phase-invariant masking:

$$\begin{aligned} \widehat{m}_n^{k,l}(i,j) &= \left|\mathcal{F}^{-1}\left\{\text{cc}_{k,l} \cdot \text{csf} \cdot \mathcal{F}\{I(i,j)\}\right\}\right| \\ &= \left|\text{CCT}_{k,l}\{I_{\text{csf}}\}\right| \end{aligned} \quad (10)$$

where I_{csf} is the input image $I(i,j)$ pre-filtered by the CSF filter.

As illustrated in Fig. 6 and Section 5, this brings the calculated Threshold Elevation map in agreement with psychophysical data suggesting that sinusoidal masks produce approximately uniform masking.

4.2 Modeling the Facilitation Effect

The publication of Daly on VDP [1] discusses the *facilitation* or the *pedestal effect* — lowering of masking thresholds when the signal and the mask have the same frequency and phase. A more extensive study of the effect is available in [4]. It shows that facilitation is quickly diminishing when frequency or phase of the signal are departing from the frequency and phase of the mask. However the effect is strong enough to be included into the HVS model proposed here. It lowers masking thresholds by up to 2.5 times in the approximate range of 0.4–1.0 JND units.

To model the facilitation effect, we are proposing an additive term to the threshold elevation formula in Eq. (3):

$$\widehat{T}_e^{k,l}(i,j) = \left(1 + k_1\left(k_2\left|\widehat{m}_n^{k,l}(i,j)\right|^s\right)^b\right)^{1/b}$$
$$- \ s_{k,l}(i,j) \cdot f\left(\left|\widehat{m}_n^{k,l}(i,j)\right|\right),$$

where $s_{k,l}(i,j) \in [0,1]$ is the strength of facilitation effect at each point and $f(t)$ is the negative Gaussian-shaped term added to the threshold elevation around the mask contrast of 0.7 (Fig. 4). To comply with psychophysical data from [4] (Fig. 2, 4), it can be modeled as

$$f(t) = 0.6\exp\frac{(t-0.7)^2}{-2(0.25)^2}.$$
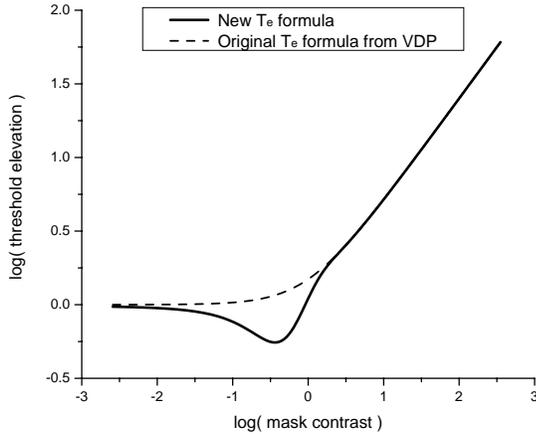
Figure 4: Threshold elevation curve $\widehat{T}_e^{k,l}(i,j)$ modeling the facilitation effect of strength $s_{k,l} = 0.5$.

According to [4], the strength of the facilitation effect depends on the proximity of frequency and phase estimates of the signal and the masker. In order to model this, we suggest using the phase information provided by CCTs of the mask and the signal:

$$s_{k,l}(i,j) = \exp\frac{\left(\phi_m^{k,l}(i,j) - \phi_s^{k,l}(i,j)\right)^2}{-2\sigma^2},$$

where $\phi_m^{k,l}(i,j) = \arg\mathrm{CCT}_{k,l}\left\{I_\mathrm{csf}\right\}(i,j)$ is the instantaneous phase angle of the mask CCT band, $\phi_m^{k,l}(i,j)$ is the similarly defined instantaneous phase angle of the signal CCT band, and $\sigma = 0.25$ provides about 2 times reduction in facilitation when the phase difference is $\pi/10$.

This model assumes that since both signal and mask are in the same CCT band, their frequencies are close. This may not always be true because bands of the cortex transform (and CCT) span one-octave frequency intervals. A more accurate approach to detection of same-frequency and same-phase signals may involve spatial averaging of absolute phase angle differences by averaging of $s_{k,l}(i,j)$:

$$\widehat{s}_{k,l}(i,j) = \mathcal{G}\left(s_{k,l}(i,j)\right),$$

where $\mathcal{G}$ is the Gaussian filtering operator. The radius of averaging is subject to additional psychophysical research, but our initial recommendation is to set it to 2 periods of the central frequency ρ_c of the cortex filter $\mathrm{cc}_{k,l}$.

4.3 Additional Modifications of the VDP

4.3.1 Luminance Adaptation

In [1] and [3] two luminance adaptation models are presented: the *global model* that averages the baseband luminance and the *local model* that uses pixel-wise values of luminance. In [3] the local model is used for adaptation of CSF filtering: different CSF filters are used depending on the adaptation luminance to model the psychophysical CSF data more accurately.

It can be argued that pixel-wise adaptation cannot produce stable results as the image resolution increases: the resulting CSF-filtered image may contain artifacts due to frequent switching of CSF filters. We suggest using a *spatially-smoothed* luminance for calculation of luminance adaptation. The radius of such smoothing is subject to additional research, but the initial estimate of $1°$ of the visual field is suggested.

4.3.2 Luminance Nonlinearity

Several referenced works describe ways for nonlinear transformation from the color model of the input image to the perceptually uniform luminance scale. An interesting observation made during our experiments may be helpful in the related research. Most images stored on personal computers do not have embedded color profiles and assume the sRGB color model. A typical way to display such images is to directly put their RGB (or luminance) values into the frame buffer of a video display adapter.

On the other hand, the most popular way of calibrating displays of personal computers (offered by many photographic web-sites) includes a black and white gradation chart with equal steps in RGB luminance, as the one in Fig. 5. The user is suggested to adjust brightness, contrast and gamma of the display until all the luminance steps in the chart become discernible and produce approximately equal subjective brightness increments.
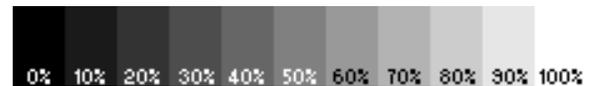
Figure 5: A typical monitor calibration chart.

It means that the user is suggested to adjust the display until *sRGB space becomes perceptually uniform* with respect to luminance. On the other hand, such a variant of "brightness" uniformity is different from uniformity in terms of JND steps, which are arguably more relevant for modeling of masking than brightness uniformity. This requires us to reconsider the nonlinear luminance transformation happening in the front end of VDP algorithms.

5. RESULTS OF MODELING

To evaluate the new model of masking, maps of threshold elevation have been computed for several test images. The source images have been generated in a perceptually uniform color space to eliminate the need for a nonlinear luminance mapping. Here we present threshold elevation maps for 2 test images that are important in psychophysical experiments.

The first mask image is a sinusoidal grating across the gradient background (Fig. 6a). Since the gradient is linear in a perceptually uniform luminance space, it is expected that the threshold elevation resulting from the grating will be spatially uniform.

Fig. 6 plots the resulting threshold elevation maps in the cortex band which contains the strongest response to the frequency of the grating. The luminance scale of the maps is stretched for easier evaluation.

As can be seen in Fig. 6b, the resulting T_e map generated by the original VDP by Daly exhibits oscillations due to inability of the cortex transform to separate phase and magnitude components of the masking stimulus.

Fig. 6c shows the modified T_e map produced by the "phase uncertainty" algorithm described by Mantiuk et. al. [3]. The algorithm smoothes variations in the threshold elevation map, but lowers the overall masking level.

Fig. 6d shows the $\widehat{T}_e$ map generated by the proposed method. It shows almost uniform elevation of the thresholds.
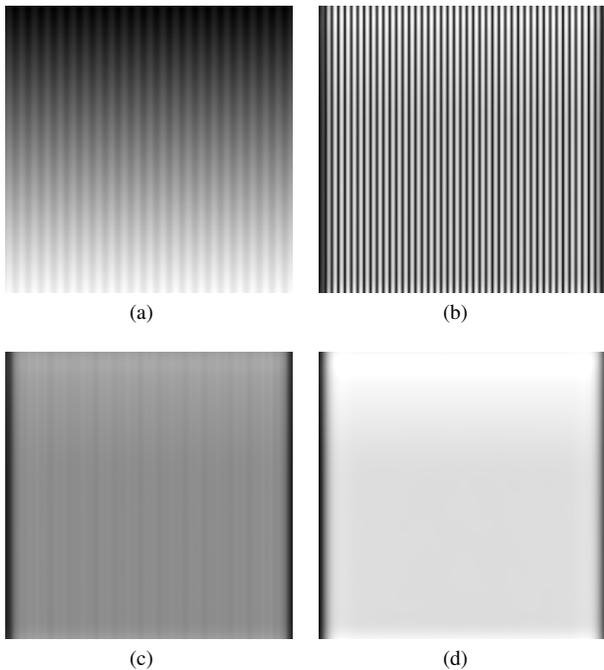
Figure 6: (a) Grating mask; (b) T_e using original VDP [1]; (c) T_e using "phase uncertainty" method from [3]; (d) $\widehat{T}_e$ using the proposed method.

The second mask image is the idealized edge image (Fig. 7a). The expected pattern of threshold elevation peaks at the edge and drops off away from the edge [9]. In Fig. 7b,c,d, the resulting T_e maps are plotted for one of the lower-frequency cortex bands with orientation corresponding to the edge orientation. Boundary effects resulting from the use of the Discrete Fourier Transform are visible around left and right boundaries of the map due to circular extension of the image.

It can be seen that the original VDP has produced the elevation map that actually has a notch at the location of the edge, flanked by a few peaks. The proposed approach for calculation of $\widehat{T}_e$ has shown the most uniform result again.
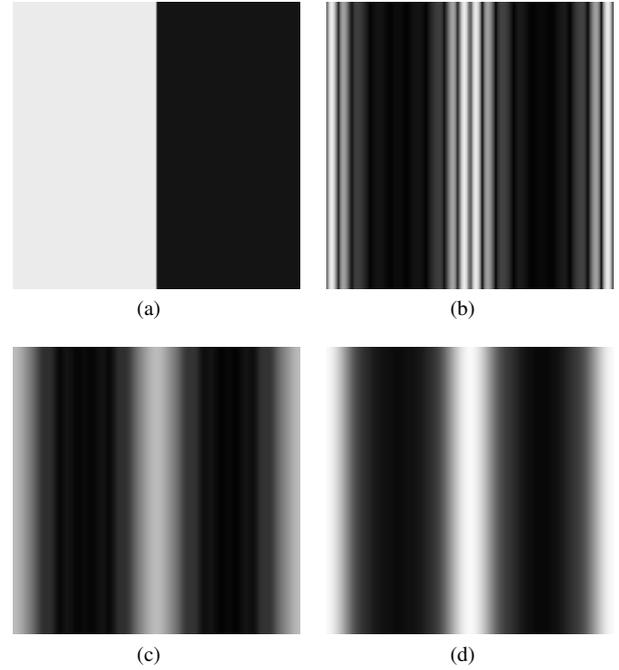
Figure 7: (a) Edge masker; (b) T_e using original VDP [1]; (c) T_e using "phase uncertainty" method from [3]; (d) T_e using the proposed method.

6. CONCLUSION

We have introduced a modification of the cortex transform using pairs of phase-complementary cortex filters. This new transform, called a Complex Cortex Transform, allows separation of magnitude and phase components of the neural response in visual cortex.

It is proposed to use the CCT for modeling of visual masking in VDP algorithms. In this work, CCT has been shown to improve consistency of threshold elevation estimates in the Visual Differences Predictor by Mantiuk/Daly. The improved VDP is also able to model the facilitation effect happening when mask and target signals are of the same frequency and phase.

In this paper, the advantages of the proposed masking model have been shown on simple, artificially generated test images. More thorough experiments need to be performed on natural images in order to assess correlation of the improved distortion measure with subjective quality data.

Acknowledgement

7. REFERENCES

[1] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," *Digital Image and Human Vision*, pp. 179–206, 1993.

[2] A. Watson, "The cortex transform: Rapid computation of simulated neural images," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 311–327, 1987.

[3] R. Mantiuk, S. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting visible differences in high dynamic range images - model and its calibration," in *Human Vision and Electronic Imaging X*, S. Daly B. Rogowitz, T. Pappas, Ed., 2005, vol. 5666, pp. 204–214.

[4] G. Legge and J. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America*, vol. 7, no. 12, pp. 1458–1471, 1980.

[5] D. Pollen and S. Ronner, "Phase relationships between adjacent simple cells in the visual cortex," *Science*, vol. 212, no. 4501, pp. 1409–1411, 1981.

[6] B. Willmore, R. Prenger, M. Wu, and J. Gallant, "The berkeley wavelet transform: A biologically inspired orthogonal wavelet transform," *Neural Computation*, vol. 20, no. 6, pp. 1537–1564, 2008.

[7] M. Johansson, "The hilbert transform," M.S. thesis, Vaxjo University, 1999.

[8] J.P. Havlicek, J.W. Havlicek, N. Mamuya, and A. Bovik, "Skewed 2d hilbert transforms and computed am-fm models," in *IEEE Int. Conf. Image Processing*, 1998, vol. 1, pp. 602–606.

[9] S. Macknik, "Visual masking approaches to visual awareness," *Progress in Brain Research*, , no. 155, pp. 177–215, 2006.

ABOUT THE AUTHOR

Alexey Lukin has graduated from a Moscow State University in 2003 and received a Ph.D. degree in 2006 for his work on perceptually motivated audio signal and image processing. He is now a member of scientific staff at the Lab of Mathematical Methods of Image Processing. Alexey's interests include image processing, audio signal processing, multiresolution filter banks and adaptive methods of spectral analysis. His publications can be found at http://imaging.cs.msu.ru.

COMMENTS OF REVIEWERS

1. *Reviewer*: End of section 2: it's not clear what do "oscillations of sinusoidal mask signal" mean. Sinusoidal signal originally oscillates between zero and magnitude. But this is correct in space domain. As I remember, elevation formula should be applied to phase-frequency domain — not directly. So, for fixed $\sin(ax + n)$ there will not be any oscillations in T_e.

   Author replies: Eqs. (4) and (3) for T_e are actually adapted from [1], and later they are also used in [3]. They are based on "instantaneous" space-domain values — a shortcoming that is addressed in this paper.

2. *Reviewer*: A model of threshold elevation presented originally by Daly and referred in this paper as [1] is not the only model. More sophisticated model is presented in *Peter G.J. Barten "Contrast sensitivity of the human eye and its effects on image quality" (PIE International Society for Optical Eng., 1999)*. It reflects masking decreasing in small masking contrast areas. And mismatch of theoretical and practical data could be caused by usage of non-optimal model, not by specific issues with separation of magnitude/phase.

   Author replies: This publication of Barten describes a masking model for simple stimuli: sinusoidal gratings and noise. However it does not discuss how a complex image should be decomposed into sinusoidal components for calculation of masking at every spatial frequency. A Fourier analysis is mentioned in that publication, but it is not an acceptable means of such separation because of loss of space locality. The subject of this paper is improvement of the cortex transform — a means of separating the image into frequency subbands. Barten's masking model can be applied to the subbands produced by the proposed method.

3. *Reviewer*: A modified formula for T_e is another way for representing the facilitation effect at low masking contrasts. It seems that this is rather important to compare it with the model presented by Barten.

   Author replies: The model of Barten calculates the facilitation effect for simple stimuli. The attempt of this paper is to extend the model to complex stimuli. I agree that the developed model lacks experimental validation.

4. *Reviewer*: Section 4.2: Facilitation is accounted for in newer masking models, e.g. *A.B. Watson and J.A. Solomon "Model of visual contrast gain control and pattern masking" (Journal of the Optical Society of America, 1997)*. Facilitation is present only for stimuli of very similar spatial frequency and orientation. This rarely happens in complex images and it was the reason why Daly did not want to have it in the VDP.

5. *Reviewer*: Subsection 4.3.1: I would not agree with blurring by $1°$ Gaussian. The spatial adaptation mechanism probably cannot be explained by a linear filter because it originates from several mechanisms with different spatial extent and time constants. A huge part of adaptation happens in cones, so it is restricted to very small spots, definitely smaller than $1°$. But I agree that the linear interpolation of several CSF-filtered images, as done in [3], is not very good either. These filters do not interpolate very well linearly.

   Author replies: I agree that mechanisms of luminance adaptation for masking are more complex than Gaussian smoothing. However Gaussian blur with a certain radius is probably better than adaptation to a single pixel of unknown size.

6. *Reviewer*: Subsection 4.3.2: Usually Gamma correction means that we estimate non-linearity of a monitor and try to compensate for it. As a result, the monitor will have approximately linear transfer function. However human perception has a non-linear feeling of physical luminance — it is proportional to the cubic root of the input value. Soww, this is actually a good question: what we should take into account and how we should do it.