Presented at: IS&T/SPIE Annual Symposium on Electronic Imaging, San Jose, CA, Jan. 19-22, 2009
Published in: Human Vision and Electronic Imaging XIV, Proc. SPIE, vol. 7240. @SPIE

Statistics of Natural Image Sequences: Temporal Motion Smoothness by Local Phase Correlations

Zhou Wang and Qiang Li

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Dept. of Electrical Engineering, University of Texas at Arlington, Arlington, TX, USA
zhouwang@ieee.org, qiang.li@mavs.uta.edu

ABSTRACT

Statistical modeling of natural image sequences is of fundamental importance to both the understanding of biological visual systems and the development of Bayesian approaches for solving a wide variety of machine vision and image processing problems. Previous methods are based on measuring spatiotemporal power spectra and by optimizing the best linear filters to achieve independent or sparse representations of the time-varying image signals. Here we propose a different approach, in which we investigate the temporal variations of local phase structures in the complex wavelet transform domain. We observe that natural image sequences exhibit strong prior of temporal motion smoothness, by which local phases of wavelet coefficients can be well predicted from their temporal neighbors. We study how such a statistical regularity is interfered with "unnatural" image distortions and demonstrate the potentials of using temporal motion smoothness measures for reduced-reference video quality assessment.

Keywords: natural image statistics, temporal motion smoothness, image sequence statistics, complex wavelet transform, local phase correlation, image quality assessment, reduced-reference video quality assessment

1. INTRODUCTION

One approach that has recently attracted wide interests in biological vision research is to study the natural visual environment.[1,2] The general belief is that the biological visual systems are highly adapted to processing natural images, which constitute a tiny cluster in the space of all possible images. Studying the statistics of natural images thus provides an indirect but effective means to understand the biological visual information processing systems. Furthermore, statistical prior knowledge about images also plays an essential role in the design of Bayesian approaches[3,4]for solving many machine vision and image processing problems.

While great effort has been made to study the statistical regularities of static natural images,[1,2] much less has been done for natural image sequences. One approach is to compute the autocorrelation function of the image sequence along both spatial and temporal directions. Assuming spatial and temporal stationality, such an autocorrelation function can be studied more conveniently in the Fourier transform domain as a spatiotemporal power spectrum.[5] It has been found that the spatiotemporal power spectrum of natural image sequences demonstrate interdependence between spatial and temporal frequencies, and the interdependence may be accounted for by assuming a $1/f^p$ static power spectrum and a rotationally invariant distribution of velocities.[5] Independent component analysis has also been applied to local 3-D blocks extracted from natural image sequences.[6] It was shown that the components obtained by optimizing independence are filters localized in space and time, spatially oriented, and directionally selective. Similar shapes of linear components were also obtained by optimizing sparseness via a matching pursuit algorithm.[7] Other prior models about natural image sequences have also been assumed, though not directly measured. For example, in the literature of optical flow estimation, it is often assumed that image motion or optical flow is spatially smooth.[8] As a result, the motion or optical flow vectors measured locally should vary smoothly across space. Explicit prior models in favor of lower speed of motion has also been assumed[5,9,10] and applied to Bayesian optical flow estimation.[9] In a recent study,[11] the shape of the "biological" speed prior was inferred directly from psychophysical speed perception experiment under the existence of noise. The inferred prior verifies the strong preference of slower motion and shows significantly heavier tails than a Gaussian.

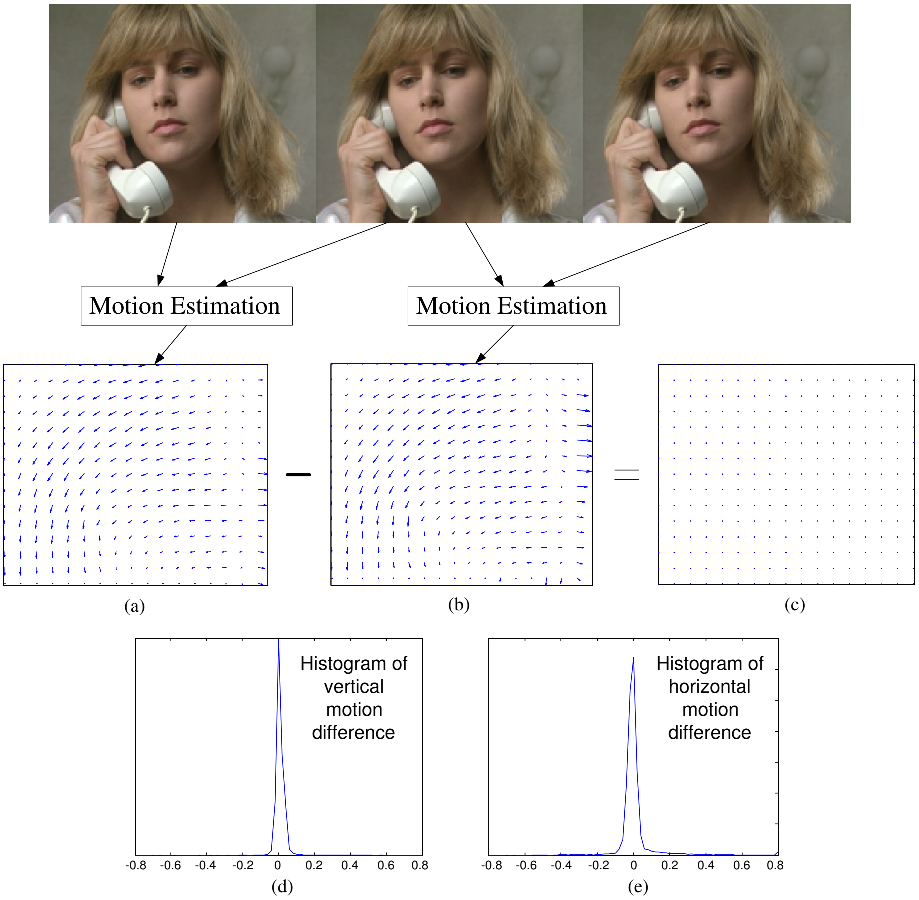


Figure 1. Illustration of motion smoothness of natural image sequences. The motion vector fields estimated for consecutive video frames are slowly varying over both space and time.

Besides the preference for lower-speed and spatially-smooth motion, here we are interested in another type of statistical regularity of natural image sequences − the smoothness of motion along temporal direction. Figure 1 gives an illustration, where (a) and (b) are motion vector fields estimated from three consecutive frames of the "Susie" sequence. It can be observed that the motion vectors are slowly-varying not only over space, but also over time, which is confirmed by the difference motion vector fields shown in (c). The histograms of the vertical and horizontal components of (c) are plotted in (d) and (e), respectively, where the high peaks at 0 indicate the statistical preference of temporal motion smoothness.

Figure 1 suggests a direct method to capture temporal motion smoothness, i.e., estimating the motion vector fields of consecutive video frames and then measuring the variations of the motion vectors along temporal direction. However, motion estimation is an computationally expensive task, which often involves a complicated search procedure (e.g., in block-matching motion estimation algorithms[12]) or requires solving adaptive equations at each spatial location (e.g., in optical flow-based motion estimation methods[8]).

In this paper, we propose to investigate temporal motion smoothness in the complex wavelet transform domain, where the magnitudes of complex wavelet coefficients exhibit translation invariance properties,[13] and the relative phase patterns between the coefficients have found to be the most informative in describing local image structures.[14,15] In previous work, global (Fourier) and local (wavelet) phases have been found to carry important information about image structures.[14,16–18] The local phase structure of static natural images demonstrates clear statistical regularities and has intriguing perceptual implications.[14] In the computer vision literature, local phase has been used in a number of applications such as estimation of image disparity[19] and motion,[20,21] description of image textures,[22] and recognition of persons using iris patterns.[23] However, the behaviors of local phase variations over time, whether such behaviors can be used to characterize "natural" image sequences, and how "unnatural" image distortions interfere with such behaviors have not been deeply investigated.

In the next section, we derive the local phase relationships for the ideal cases of temporal motion smoothness. Such phase relationships allow us to predict the phase structures of complex wavelet coefficients along temporal direction. The accuracy of these ideal phase predictions for real natural images is studied empirically in Section 3 and a probability model that can account for the predictions is proposed. In Section 4, we investigate how "unnatural" image distortions disturb such temporal statistical regularities of local phase. Section 5 demonstrates the potential applications of temporal motion smoothness measurement in reduced-reference video quality assessment. Finally, Section 6 draws conclusions and discuss potential extensions of the work.

2. TEMPORAL MOTION SMOOTHNESS BY LOCAL PHASE CORRELATIONS

Let $f(x)$ be a given real static signal, where x is the index of spatial position. When $f(x)$ represents an image, x is a 2-D vector. For simplicity, in the derivations below, we assume x to be one dimensional. However, the results can be easily generalized to two and higher dimensions. A time varying image sequence can be created from the static image $f(x)$ with rigid motion and constant variations of average intensity:

$$h(x,t) = f(x+u(t)) + b(t) \,. \tag{1}$$

Here $u(t)$ indicates how the image positions move spatially as a function of time. $b(t)$ is real and accounts for the time-varying background luminance changes. This formulation can be viewed as a generalization of the brightness constancy assumption[8, 24] (in which $b(t) \equiv 0$), but the inclusion of the luminance change improves flexibility and stability of the representation. For example, when the lighting condition of a fixed scene changes over time, the brightness constancy assumption would not hold, but the situation would be better described with this formulation.

Now consider a family of symmetric complex wavelets whose "mother wavelets" can be written as a modulation of a low-pass filter $w(x)= g(x)\, e^{j\omega_c x}$, where ω_c is the center frequency of the modulated band-pass filter, and $g(x)$ is a slowly varying and symmetric function. The family of wavelets are dilated/contracted and translated versions of the mother wavelet:

$$w_{s,p}(x) = \frac{1}{\sqrt{s}}\, w\left(\frac{x-p}{s}\right) = \frac{1}{\sqrt{s}}\, g\left(\frac{x-p}{s}\right)\, e^{j\omega_c (x-p)/s} \,, \tag{2}$$

where $s \in R^+$ is the scale factor, and $p \in R$ is the translation factor. Considering the fact that $g(-x) = g(x)$, and using the convolution theorem and the scaling and modulation properties of the Fourier transform, we can compute the complex wavelet transform of a given signal $f(x)$ as

$$\begin{aligned} F(s,p) &= \int_{-\infty}^{\infty} f(x)\, w^*_{s,p}(x)\, dx = \left[f(x) * \frac{1}{\sqrt{s}}\, g\left(\frac{x}{s}\right)\, e^{j\omega_c x/s} \right]_{x=p} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \sqrt{s}\, G(s\,\omega - \omega_c)\, e^{j\omega p}\, d\omega \,, \end{aligned} \tag{3}$$

where $F(\omega)$ and $G(\omega)$ are the Fourier transforms of $f(x)$ and $g(x)$, respectively. Applying such a complex wavelet transform to both sides of Eq. (1) at a given time instance t, we have

$$\begin{aligned} H(s,p,t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \sqrt{s}\, G(s\,\omega - \omega_c)\, e^{j\omega(p+u(t))}\, d\omega \\ &= \frac{e^{j(\omega_c/s)u(t)}}{2\pi} \int_{-\infty}^{\infty} F(\omega) \sqrt{s}\, G(s\omega - \omega_c) e^{j\omega p} e^{j(\omega - \omega_c/s)u(t)} d\omega \\ &\approx F(s,p)\, e^{j(\omega_c/s)u(t)} \,. \end{aligned} \tag{4}$$

Here $b(t)$ is eliminated because of the bandpass nature of the wavelet filters. The approximation is valid when the envelope window $g(t)$ is slowly varying and the motion $u(t)$ is small. In the extreme case, the approximation becomes exact when $g(x) \equiv 1$, i.e., $G(\omega) = \delta(\omega)$, or when there is no motion, i.e., $u(t) = 0$. A more convenient way to understand Eq. (4) is to take a logarithm on both sides, which gives

$$\log H(s,p,t) \approx \log F(s,p) + j(\omega_c/s)u(t)\,. \tag{5}$$

Note that the first term of the right-hand-side does not change over time. The key property of Eq. (5) is that at a given scale s and a given spatial position p, the imaginary part of the logarithm of the complex wavelet coefficient changes linearly with $u(t)$. In other words, the local phase structures over time can be fully characterized by the movement function $u(t)$. Taylor series expansion of $u(t)$ at a specific time instance t_0 yields

$$u(t) = u(t_0) + u'(t_0)(t-t_0) + \frac{u''(t_0)}{2}(t-t_0)^2 + \cdots + \frac{u^{(n)}(t_0)}{n!}(t-t_0)^n + \cdots\,. \tag{6}$$

We call $u(t)$ N-th order smooth if its (N+1)-th and higher order derivatives with respect to t are all zeros. For instance, zero-order smooth motion implies no motion [$u(t)$ is a constant over time], first-order smooth motion corresponds to constant speed [$u'(t)$ is a constant], and second-order smooth motion leads to constant acceleration [$u''(t)$ is a constant], and so on. Notice that here the definition of motion smoothness is different from the notion of motion smoothness typically used in optical flow estimation,[8] where motion smoothness refers to the slow variations of motion vectors over space. We believe that *temporal motion smoothness* is a better term to describe the concept we are discussing here.

In order to relate temporal motion smoothness with the time-varying complex wavelet transform relationship of Eq. (5), we must examine the complex wavelet coefficients at multiple time instances. A convenient choice is to start from a time instance t_0 and sample the sequence at consecutive time steps $t_0 + n\Delta t$ for $n = 0, 1, ..., N$. The N-th order derivatives of $u(t)$ at t_0 can be approximated by the following N-th order differentiator:

$$u^{(N)}(t_0) = \frac{1}{(\Delta t)^N}\sum_{n=0}^{N}(-1)^{n+N}\binom{N}{n}u(t_0+n\Delta t)\,. \tag{7}$$

Now we define the N-th order *temporal correlation function* as follows:

$$L_N(s,p) = \sum_{n=0}^{N}(-1)^{n+N}\binom{N}{n}\log H(s,p,t_0+n\Delta t)\,. \tag{8}$$

By Eq. (5), we have

$$\begin{aligned}
L_N(s,p) &\approx \sum_{n=0}^{N}(-1)^{n+N}\binom{N}{n}\left[\log F(s,p) + j(\omega_c/s)u(t_0+n\Delta t)\right] \\
&= (-1)^N \log F(s,p)\left[\sum_{n=0}^{N}(-1)^n\binom{N}{n}\right] + j\frac{w_c}{s}\left[\sum_{n=0}^{N}(-1)^{n+N}\binom{N}{n}u(t_0+n\Delta t)\right] \\
&= j\frac{w_c(\Delta t)^N}{s}u^{(N)}(t_0)\,,
\end{aligned} \tag{9}$$

where we have used Eq. (7) and the fact that $\sum_{n=0}^{N}(-1)^n\binom{N}{n} = 0$. Now suppose that the motion is (N-1)-th order smooth, then $u^{(N)}(t_0) = 0$, and therefore

$$L_N(s,p) \approx 0\,. \tag{10}$$

It needs to be kept in mind that this approximation is achieved based on the ideal formulation of Eq. (1) and the ideal assumption of (N-1)-th order temporal motion smoothness. Real natural image sequences are expected to deviate from these assumptions. However, by looking at the statistics of $L_N(s,p)$ (especially its imaginary part, which is a measure of temporal local phase correlation), one may be able to quantify such deviation and use it as an indication of the strength of temporal motion smoothness.

In addition, we define the following temporal weighted averaging function in the log-complex wavelet domain:

$$M_N(s,p) = \sum_{n=0}^{N} \binom{N}{n} \log H(s,p,t_0 + n\Delta t)\,. \tag{11}$$

We find it also helpful in characterizing the statistical properties of natural image sequences and will demonstrate its usefulness in the next section.

3. IMAGE SEQUENCE STATISTICS

For a given image sequence, we decompose each frame using the complex version[22] of the steerable pyramid,[25] a multi-scale wavelet decomposition whose basis functions are spatially localized, oriented, and roughly one octave in bandwidth. Specifically, a 3-scale 2-orientation pyramid is computed, resulting in six oriented subbands, a highpass residual band, and a lowpass residual band. By aligning the oriented subbands at the same orientation and scale but across different frames, we obtain a discrete (in both space and time) version the function $H(s,p,t)$ for a particular scale and orientation. We then compute $L_N(s,p)$ and $M_N(s,p)$ for $N = 1,2,3,4$ for all the coefficients within the subband.

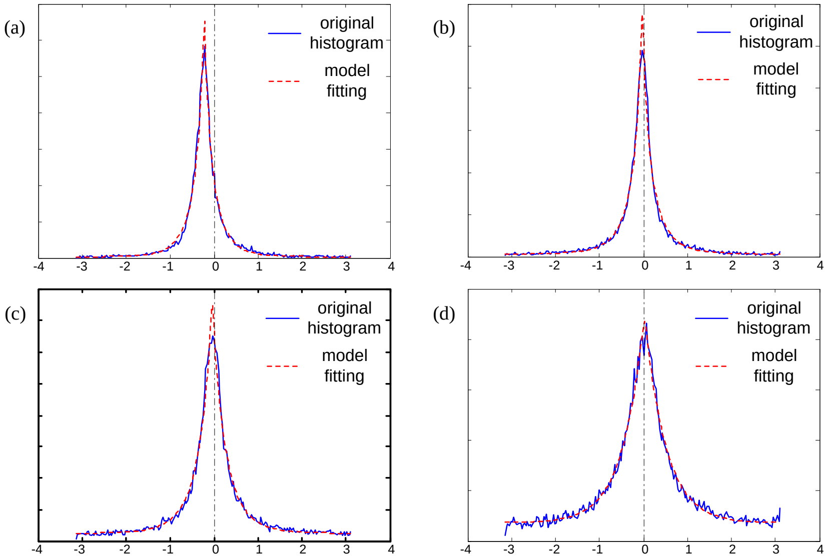


Figure 2. Marginal statistics of the imaginary parts of the first-order (a), second-order (b), third-order (c), and fourth-order (d) temporal correlation functions $L_N(s,p)$. The image sequence demonstrates strong temporal motion smoothness.

To study temporal motion smoothness, we first examine the marginal distribution of the imaginary part of the temporal correlation coefficient $imag\{L_N(s,p)\}$. The histograms of $imag\{L_N(s,p)\}$ for $N = 1,2,3,4$ of the "Susie" sequence are shown in Fig. 2. It can be observed that all the histograms peak near zero, and the peaks move toward zero with the increasing order of the temporal correlation function. Although Fig. 2 only shows the statistical results from a single image sequence, similar results were obtained for most of the other sequences we tested*. This demonstrates strong prior of temporal motion smoothness of natural image sequences. Another important observation is that the histograms are quite peaky, much more than the von Mises distribution widely used in describing statistics of circular data.[26] We empirically found that a four-parameter function that can

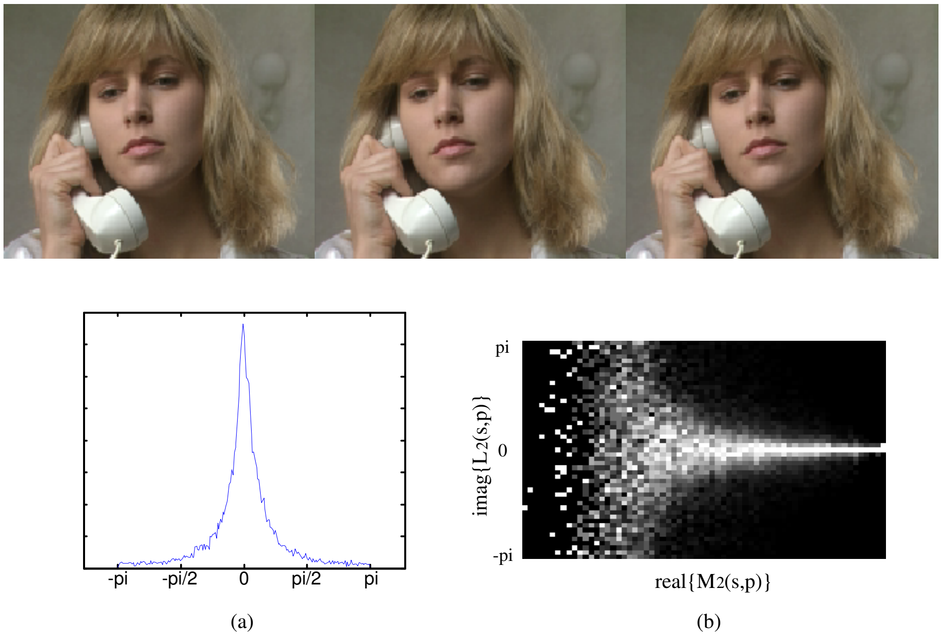
*Exceptions were observed for the image frames across scene changes and for the image frames with very large motion (where the distances of moving objects between frames are beyond the coverage of the wavelet filter envelops).

almost always well describe the data is given by

$$p_m(\theta) = \frac{1}{Z} \left\{ \exp \left[- \left(\frac{|\sin[(\theta - \theta_0)/2]|}{\alpha} \right)^{\beta} \right] + C \right\} \tag{12}$$

where θ is the phase variable, Z is a normalization constant, and the four parameters θ_0, α, β and C controls the center position, width, peakedness and the baseline of the function, respectively. We numerically fit the histograms with the model by minimizing the Kullback-Leibler distance[27] (KLD) between the observed and the model distributions. Some fitting results are demonstrated in Figure 2. We have used this fitting model for reduced-reference image quality assessment, which will be detailed in Section 5.

We have also studied the relationship between temporal motion smoothness and the strength of the underlying local signal. In particular, we generate the conditional histogram of the imaginary part of $L_N(s,p)$ versus the real part of $M_N(s,p)$, which provides a useful measure of local signal strength. The result is demonstrated in Figure 3(b), where each column in the 2-D histogram is normalized to one. Again, the histogram shows strong temporal motion smoothness, and such a statistical regularity becomes stronger with the increase of local signal strength. This is not surprising because small magnitude coefficients typically come from the smooth background regions in an image and are easily disturbed by background noise.



(a)

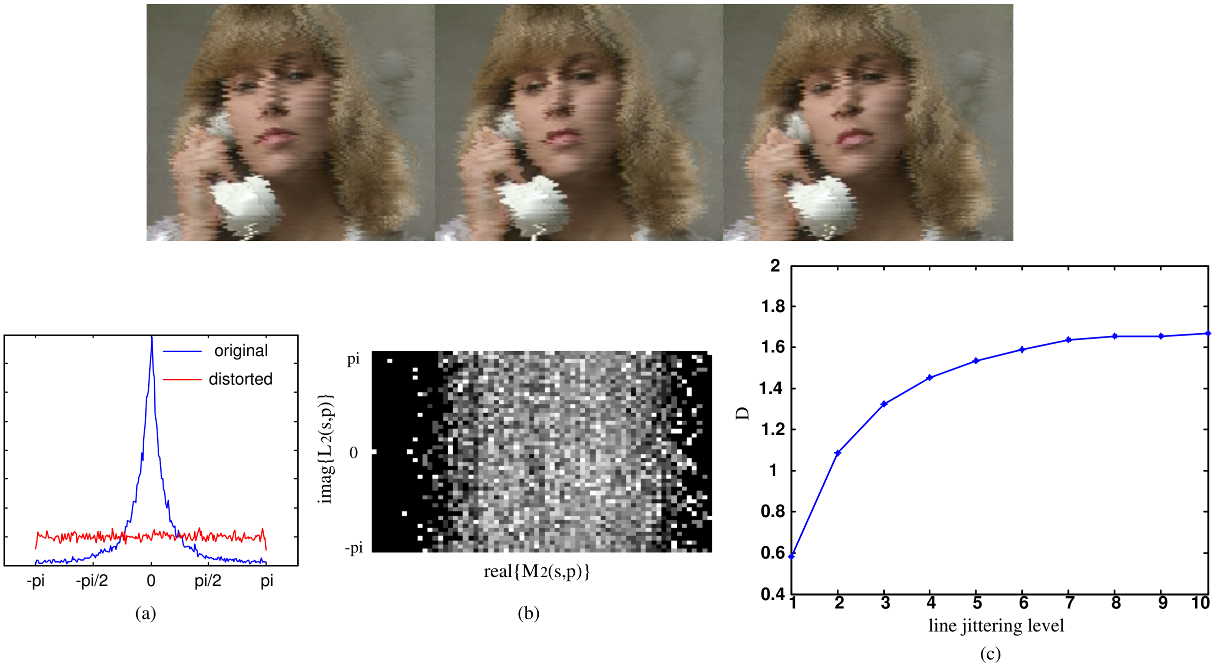
(b)

Figure 3. Three consecutive frames of the image sequence "Susie" and statistics of the second-order temporal correlation function $L_2(s,p)$. (a) Marginal histogram of the imaginary part; (b) Histogram of the imaginary part of $L_2(s,p)$ conditioned on the real part of $M_2(s,p)$.

4. INTERFERENCE WITH "UNNATURAL" DISTORTIONS

The merit of natural image prior models should be evaluated by their capabilities of distinguishing natural and unnatural images. Here we simulate a set of "unnatural" image distortions that often occur in real-world applications and examine how these distortions interfere with the temporal motion smoothness prior.

The distortions being tested are divided into two categories. The first category of distortions do not change individual pixel values but directly disturb temporal motion smoothness by shifting the positions of pixels. Specifically, we investigated the effects of line jittering, frame jittering and frame dropping distortions, each



(a)

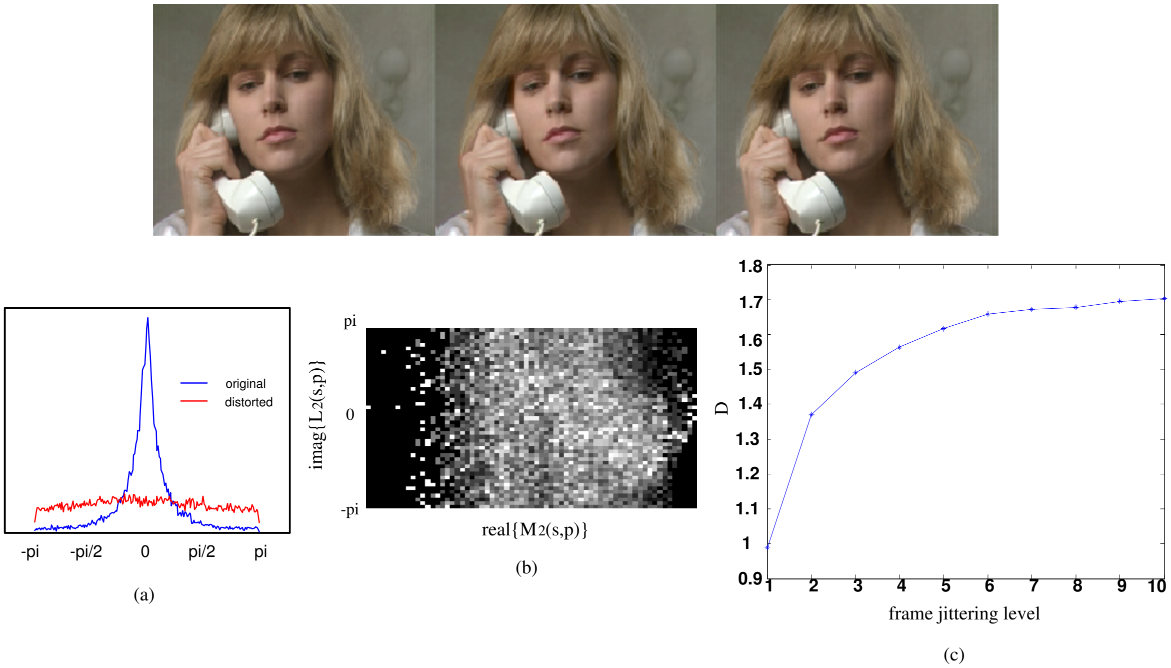
(b)

(c)

Figure 4. Three consecutive frames of the image sequence “Susie” distorted with line jittering and statistics of the second-order temporal correlation function $L_2(s, p)$. (a) Marginal histogram of the imaginary part; (b) Histogram of the imaginary part of $L_2(s, p)$ conditioned on the real part of $M_2(s, p)$; (c) Objective RRVQA score D as a function of line jittering level.

of which is associated with certain real-world scenario. In particular, line jittering occurs when two fields of interlaced video signals are not synchronized, frame jittering is often caused by irregular camera movement such as hand shaking, and frame dropping usually happens when the bandwidth of a real-time communication channel drops and some video frames have to be discarded to reduce the bit rate of the video signal being transmitted. To simulate line jittering, we shift each line in a video frame horizontally by a random amount uniformly distributed between a range of $[-S, S]$, where S defines the level of jittering distortion. Figure 4 shows the results of liner jittering. Comparing the marginal and conditional histograms (Fig. 4(a) and (b)) with those in Fig. 3, we observe that the distributions of temporal phase correlation coefficients become almost flat, which implies that the prior structure of temporal motion smoothness shown in Fig. 3 is severely disturbed. Frame jittering is simulated in a similar way, only that the entire frame (rather than each line in the frame) is shifted together. Again, the statistical regularity of temporal motion smoothness has been destroyed, as demonstrated in Fig. 5. To simulate frame dropping, we discard N out of every $N + 1$ frames and use N to define the level of frame dropping. The dropped frames will then be filled by repeating their previous frames. Figure 6 shows the effect of frame dropping. It can be seen that the sharpness of the marginal and conditional histograms is significantly reduced and the centers of the peaks in the distributions are shifted away from 0, demonstrating a clear disruption of temporal motion smoothness.

The second category of distortions directly alter the values of individual image pixels. In particular, we studied the effects of additive white Gaussian noise contamination and Gaussian blur distortion. Although they do not directly change the motion information contained in the video, they reduce the sharpness of local image structures, and thus affect the local phase correlations across frames. In Fig. 7, white Gaussian noise is added to each frame of the video sequence, where the noise level is defined as the standard deviation of the Gaussian distribution. In Figure 8, each video frames is blurred spatially by convolving with a linear filter of Gaussian shape, where the standard deviation of the Gaussian filter defines the blur level. It can be observed that in both cases, the strong prior of temporal motion smoothness is significantly reduced.



(a)

(b)

(c)

Figure 5. Three consecutive frames of the image sequence "Susie" distorted with frame jittering and statistics of the second-order temporal correlation function $L_2(s,p)$. (a) Marginal histogram of the imaginary part; (b) Histogram of the imaginary part of $L_2(s,p)$ conditioned on the real part of $M_2(s,p)$; (c) Objective RRVQA score D as a function of frame jittering level.

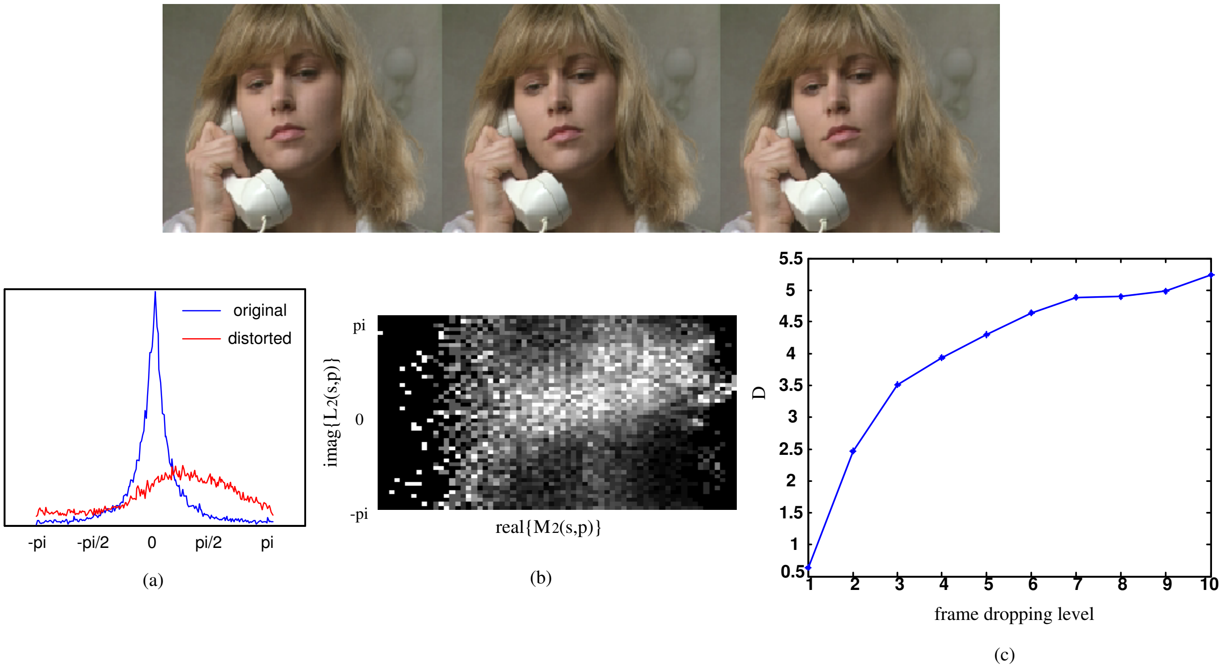
5. APPLICATION TO REDUCED-REFERENCE VIDEO QUALITY ASSESSMENT

From the study in previous sections, we observe that temporal motion smoothness is a common feature of natural image sequences but disrupted by various types of "unnatural" image distortions. One direct application of such a feature is to use it for reduced-reference video quality assessment (RRVQA), which aims to estimate video quality degradations with only partial information about the "perfect-quality" reference video (This is different from full-reference video quality measures such as peak signal-to-noise ratio and the structural similarity index[28] that require full access to the original video). The idea is to use temporal motion smoothness measures extracted from the reference video signal as the RR features and then quantify video quality degradations based on the variations of these RR features in the distorted video signal.

For a given image sequence, we first divide it into groups of pictures (GOPs), each containing 3 consecutive frames. For each GOP, we apply a complex steerable pyramid decomposition to all 3 frames and compute the second order temporal correlation function $L_2(s,p)$ for each oriented subband. The observations of the marginal histograms shown in Figs. 4 to 8 suggest that the variations in the marginal distributions of $imag\{L_2(s,p)\}$ between the original and distorted image sequences can be used as a measure of image distortions. A convenient way to quantify such variations is to compute the KLD[27] between them:

$$d(p\|q) = \int p(\theta) \log \frac{p(\theta)}{q(\theta)} d\theta \,, \tag{13}$$

where $p(\theta)$ and $q(\theta)$ are the probability density functions of $imag\{L_2(s,p)\}$ of the original and distorted signals, respectively. To accomplish this, the histograms of both the original and distorted signals must be available. The latter can be easily computed from the distorted signal, which is always available. The difficulty is in obtaining the histogram of the reference signal. Using all the histogram bins as RR features would result in either a heavy RR data rate (when the bin size is fine) or a poor approximation accuracy (when the bin size is coarse). To overcome this problem, we make use of the fitting model of Eq. (12), such that only four parameter (θ_0, α, β and



(a) (b) (c)

Figure 6. Three consecutive frames of the image sequence "Susie" with frame dropping distortion and statistics of the second-order temporal correlation function $L_2(s,p)$. (a) Marginal histogram of the imaginary part; (b) Histogram of the imaginary part of $L_2(s,p)$ conditioned on the real part of $M_2(s,p)$; (c) Objective RRVQA score D as a function of frame dropping level.

C) are needed to describe it (as opposed to all the histogram bins). Furthermore, to account for the variations between the model and the true distribution, we compute the KLD between $p_m(\theta)$ and $p(\theta)$ as

$$d(p_m\|p) = \int p_m(\theta) \log \frac{p_m(\theta)}{p(\theta)} d\theta \tag{14}$$

In summary, a total of 5 RR features (4 features to describe $p_m(\theta)$ together with $d(p_m\|p)$) are extracted from each subband of the original signal.

To evaluate the quality of the distorted image sequence, we first estimate the KLD between the probability density function $q(\theta)$ of the $imag\{L_2(s,p)\}$ coefficients computed from the distorted signal and the model $p_m(\theta)$ estimated from the original signal:

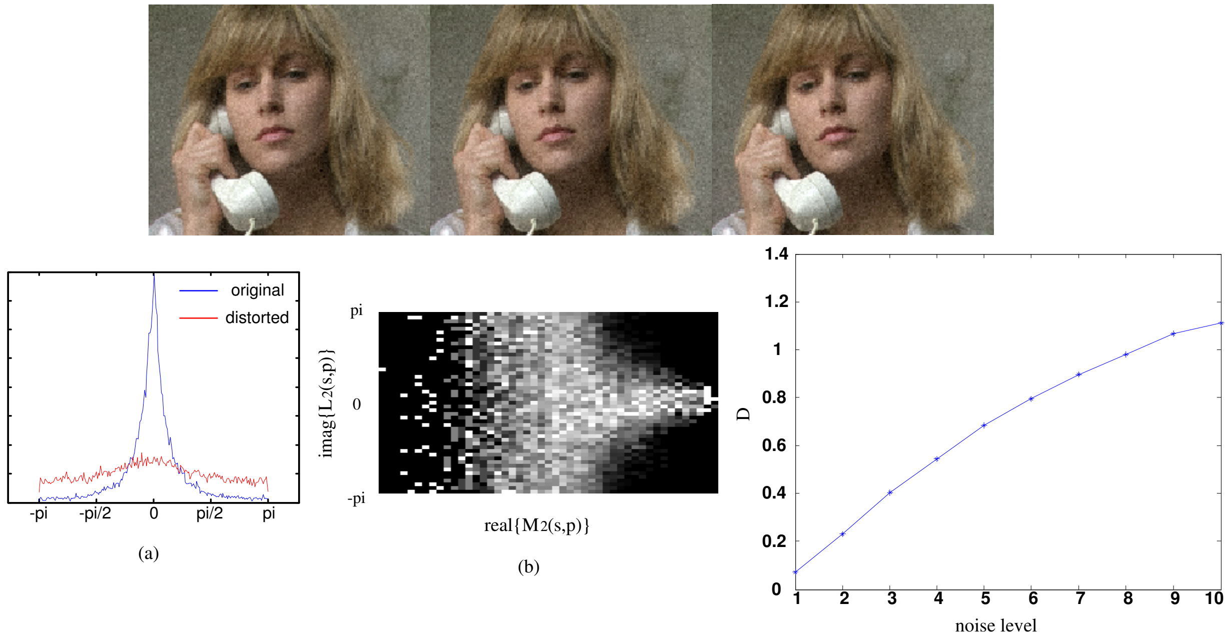
$$d(p_m||q) = \int p_m(\theta) \log \frac{p_m(\theta)}{q(\theta)} d\theta \,. \tag{15}$$

Combining this with the available RR feature $d(p_m||p)$, we obtain an estimate of the KLD between $p(\theta)$ and $q(\theta)$:

$$\hat{d}(p||q) = d(p_m||q) - d(p_m||p) = \int p_m(\theta) \log \frac{p(\theta)}{q(\theta)} d\theta \,. \tag{16}$$

With the additional cost of adding one more RR parameter $d(p_m||p)$, Eq. (16) not only delivers a more accurate estimate of $d(p\|q)$ than Eq. (15), but also provides a useful feature that when there is no distortion between the original and distorted signals (which implies that $p(\theta) = q(\theta)$ for all θ), both the targeted distortion measure $d(p\|q)$ and estimated distortion measure $\hat{d}(p\|q)$ are exactly zero. Finally, the overall quality degradation of the distorted image sequence is computed as

$$D = \frac{1}{K} \sum_{GOPs} \sum_{subbands} \hat{d}(p\|q) \,, \tag{17}$$



(a)

(b)

(c)

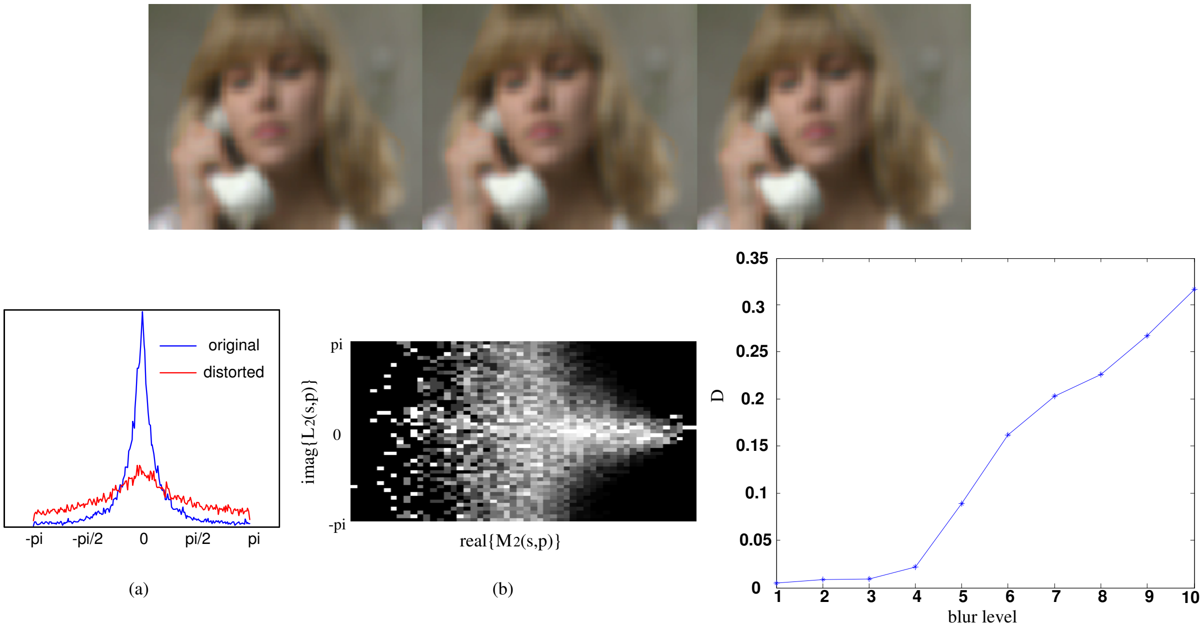
Figure 7. Three consecutive frames of the image sequence "Susie" contaminated with different levels of white Gaussian noise and statistics of the second-order temporal correlation function $L_2(s,p)$. (a) Marginal histogram of the imaginary part; (b) Histogram of the imaginary part of $L_2(s,p)$ conditioned on the real part of $M_2(s,p)$; (c) Objective RRVQA score D as a function of noise level.

where K is the number of GOPs in the image sequence.

We test the proposed algorithm using five types of distortions, including line jittering, frame jittering, frame dropping, additive white Gausssian noise contamination and Gaussian blur, as described in Section 4. The results for the "Susie" image sequence of the five distortion types are shown in Figs. 4 to 8 (c), respectively, where the horizontal axes indicate the distortion levels and the vertical axes show the distortion measure computed using Eq. (17). It can be observed that the same objective distortion measure D is consistently increasing with the strength of each individual type of distortion. Similar results were obtained for other image sequences we tested. This demonstrates the potential of the proposed method for general-purpose RRVQA, which is different from most VQA approaches in the literature where ad-hoc features tuned to specific distortion types (such as blocking[29] and ringing[30] artifacts) are often used, and thus limit their application scope. Another interesting observation is regarding the frame jittering and frame dropping distortions. Notice that with these two types of distortions, the quality of each individual frame remains high quality, and thus frame-by-frame quality assessment approaches would give high quality scores to the image sequences undergoing these distortions, but the proposed method can capture them quite effectively without any specific change of the algorithm.

6. CONCLUSION AND DISCUSSION

We propose a new method to capture the statistical regularities of natural image sequences. In particular, we investigated the local phase structures of images along the temporal direction. We developed a temporal correlation function, which is a useful tool to measure the temporal motion smoothness of image sequences. We observed that natural image sequences exhibit strong prior of temporal motion smoothness. A probability model is proposed to describe the marginal statistics of temporal phase correlation coefficients. We demonstrated how typical "unnatural" image distortions interfere with the temporal motion smoothness prior. The distortions between the marginal distributions of the temporal motion smoothness of the original and distorted image



(a)

(b)

(c)

Figure 8. Three consecutive frames of the image sequence "Susie" distorted with different levels of Gaussian blur and statistics of the second-order temporal correlation function $L_2(s, p)$. (a) Marginal histogram of the imaginary part; (b) Histogram of the imaginary part of $L_2(s, p)$ conditioned on the real part of $M_2(s, p)$; (c) Objective RRVQA score D as a function of blur level.

sequences are used to predict video quality degradations. The advantage of the proposed local phase correlation-based approach is that temporal motion smoothness is measured without an explicit motion estimation process, which is often computationally expensive. In addition, different orders of temporal motion smoothness can be modelled under a unified framework (Specifically, zero-, first- and second-order temporal motion smoothness correspond to no motion, constant speed, and constant acceleration, respectively).

The temporal motion smoothness prior has useful implications for biological vision. One of the major tasks of biological visual systems is to track moving objects, where the observer must be able to efficiently distinguish self-motion and the motion created by the moving objects in the visual world. Prior knowledge about the motions in the natural visual environment would be useful information for this purpose.

The temporal motion smoothness prior may also be used for solving a number of computer vision and image processing problems. Besides RRVQA being studied in this paper, the prior model has the potential to be employed for no-reference video quality assessment (where no information about the reference video is available). It can also be applied to video compression, which aims to remove the statistical redundancies within image sequences, and prior knowledge about natural image sequences will certainly be helpful. Furthermore, the prior may also be used for other applications such as motion estimation, target tracking, and video filtering, denoising and restoration.

REFERENCES

[1] Ruderman, D. L., "The statistics of natural images," *Network: Computation in Neural Systems* **5**, 517–548 (1996).

[2] Simoncelli, E. P. and Olshausen, B., "Natural image statistics and neural representation," *Annual Review of Neuroscience* **24**, 1193–1216 (May 2001).

[3] Portilla, J., Strela, V., Wainwright, M., and Simoncelli, E. P., "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans Image Processing* **12**, 1338–1351 (November 2003).

[4] Simoncelli, E. P., "Bayesian multi-scale differential optical flow," in [*Handbook of Computer Vision and Applications*], Jähne, B., Haussecker, H., and Geissler, P., eds., **2**, ch. 14, 397–422, Academic Press, San Diego (April 1999).

[5] Dong, D. W. and Atick, J. J., "Statistics of natural time-varying images," *Network: Computation in Neural Systems* **6**, 345–358 (1995).

[6] van Hateren, J. H. and Ruderman, D. L., "Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex," *Proc R Soc Lond B* **265**, 2315–2320 (1998).

[7] Olshausen, B. A., "Learning sparse, overcomplete representations of time-varying natural images," in [*Proc. IEEE Int. Conf. Image Proc.*], **1**, 41–44 (Sept. 2003).

[8] Beauchemin, S. S. and Barron, J. L., "The computation of optical flow," *ACM Computing Surveys* **27**, 433–467 (Sept. 1995).

[9] Simoncelli, E. P., Adelson, E. H., and Heeger, D. J., "Probability distributions of optical flow," in [*Proc Conf on Computer Vision and Pattern Recognition*], 310–315, IEEE Computer Society, Mauii, Hawaii (June 3-6 1991).

[10] Weiss, Y., Simoncelli, E. P., and Adelson, E. H., "Motion illusions as optimal percepts," *Nature Neuroscience* **5**, 598–604 (June 2002).

[11] Stocker, A. A. and Simoncelli, E. P., "Noise characteristics and prior expectations in human visual speed perception," *Nature Neuroscience* **9**, 578–585 (April 2006).

[12] Dufaux, F. and Moscheni, F., "Motion estimation techniques for digital TV: a review and a newcontribution," *Proceedings of the IEEE* **83**, 858–876 (June 1995).

[13] Selesnick, I., Baraniuk, R., and Kingsbury, N., "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine* **22** (Nov. 2005).

[14] Wang, Z. and Simoncelli, E. P., "Local phase coherence and the perception of blur," in [*Nerual Information Processing Systems*], (Dec. 2003).

[15] Wang, Z. and Simoncelli, E. P., "Translation insensitive image similarity in complex wavelet domain," in [*Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*], (Mar. 2005).

[16] Oppenheim, A. V. and Lim, J. S., "The importance of phase in signals," *Proc. of the IEEE* **69**, 529–541 (1981).

[17] Morrone, M. C. and Burr, D. C., "Feature detection in human vision: A phase-dependent energy model," *Proc. R. Soc. Lond. B* **235**, 221–245 (1988).

[18] Kovesi, P., "Phase congruency: A low-level image invariant," *Psych. Research* **64**, 136–148 (2000).

[19] Fleet, D. J., "Phase-based disparity measurement," *CVGIP: Image Understanding* **53**(2), 198–210 (1991).

[20] Fleet, D. J. and Jepson, A. D., "Computation of component image velocity from local phase information," *Int'l J Computer Vision* **5**(1), 77–104 (1990).

[21] Magarey, J. F. A. and Kingsbury, N. G., "Motion estimation using a complex-valued wavelet transform," *IEEE Trans. Signal Proc.* **46**, 1069–1084 (Apr. 1998).

[22] Portilla, J. and Simoncelli, E. P., "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int'l Journal of Computer Vision* **40**, 49–71 (December 2000).

[23] Daugman, J., "Statistical richness of visual phase information: update on recognizing persons by iris patterns," *Int'l J Computer Vision* **45**(1), 25–38 (2001).

[24] Horn, B. K. P. and Schunck, B. G., "Determining optical flow," *Artificial Intelligence* **17**, 185–203 (1981).

[25] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J., "Shiftable multi-scale transforms," *IEEE Trans Information Theory* **38**, 587–607 (March 1992). Special Issue on Wavelets.

[26] Fisher, N. I., [*Statistical analysis of circular data*], Cambridge University Press, New York (2000).

[27] Cover, T. M. and Thomas, J. A., [*Elements of Information Theory*], Wiley-Interscience, New York (1991).

[28] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing* **13**, 600–612 (Apr. 2004).

[29] Wang, Z., Bovik, A. C., and Evans, B. L., "Blind measurement of blocking artifacts in images," in [*Proc. IEEE Int. Conf. Image Proc.*], **3**, 981–984 (Sept. 2000).

[30] Marziliano, P., Dufaux, F., Winkler, S., and Ebrahimi, T., "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Processing: Image Communication* **19**, 163–172 (Feb. 2004).